

University of Bath



PHD

Structure-function relationships of alpha lactalbumin

Pike, Ashley C . W.

*Award date:*  
1995

*Awarding institution:*  
University of Bath

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 22. May. 2019

# **STRUCTURE – FUNCTION RELATIONSHIPS OF ALPHA LACTALBUMIN**

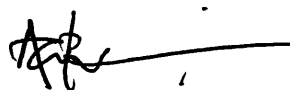
submitted by Ashley C.W. Pike  
for the degree of PhD  
of the University of Bath  
1995

## **COPYRIGHT**

Attention is drawn to the fact that copyright of this thesis rests with its author.

This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.



UMI Number: U539261

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U539261

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



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

UNIVERSITY OF BATH	
LIBRARY	
26	18 DEC 1995
PHD	

S095930



# Abstract

The functional role of two aromatic substructures and the ‘flexible loop’ region of bovine  $\alpha$ -lactalbumin (LA) have been investigated using protein engineering and X-ray crystallography. An efficient non-fusion protein expression system has been developed to produce LA in *E. coli*. Twenty mutant LAs, containing single-site substitutions in a cluster of four invariant residues (aromatic cluster I), the hydrophobic box and a flexible loop adjacent to the cleft, have been expressed, purified and kinetically characterised. Changes in aromatic cluster I and the flexible loop have specific effects on LA’s ability to modulate the substrate specificity of bovine galactosyltransferase (GT) in the lactose synthase complex. Substitutions at His-107, Gln-117 and Trp-118 reduce affinity for GT, whereas changes at Phe-31 and His-32 predominantly affect LA’s ability to promote glucose binding and have lesser effects on affinity. In contrast, substitutions at the hydrophobic box affect the folding and integrity of the native conformation of LA.

The crystal structure of the parent recombinant LA (mLA) has been determined using X-ray crystallography and refined at 2.3Å resolution. Several of the variant LAs have been crystallised but their structures have not been elucidated. In parallel studies, the three-dimensional structures of LA from goat and guinea-pig milk have been determined at resolutions of 1.9Å and 2.3Å respectively. Although the three structures are very similar, significant differences are observed in conformation of residues 105-110. The LA crystal structures demonstrate that this ‘flexible loop’ region is particularly mobile and that it can adopt a number of different conformations depending on the

conditions used to obtain crystals. The general observation that LA's interaction site for GT is located in the flexible C-terminal portion of the molecule implies that conformational adjustments may be important for the formation and function of the lactose synthase complex.

# Acknowledgements

I would like to begin by thanking my supervisor, Ravi Acharya, for his support and encouragement during the past three years. His general enthusiasm and devotion to the LA project carried me through the good and bad times. I am also deeply grateful to Dr. Keith Brew who made it possible for me, both practically and financially, to become involved in the mutagenesis side of the project in his laboratory at the University of Miami School of Medicine. On this note I would also like to thank Jay Grobler, one of Keith's postgraduate students, with whom I worked closely in constructing the non-fusion expression vector and the variant LAs. There are a number of people in Miami who made my stay enjoyable and I would particularly like to thank Jay, Kurt and Ayce for, amongst other things, weekend fun on SoBe.

There are many people who have helped me during my three (it seems longer!) years in Bath — thank you all! Particular thanks must go to Demetres and Tassos, my fellow Daresbury insomniacs, for helping with the monoclinic mLA data collection; Rupert and Andrew, my occasional UBSA drinking partners; Simon for his encouragement during the 'dark times'; and Peter for his help with all manner of computational problems. The generous financial support of The Wellcome Trust is also gratefully acknowledged. Finally I would like to say a special thank you to Mum and Dad for their love and understanding.

October 1995

A.C.W.P.

# Abbreviations

AD-LAB	In-house Siemens area detector
AU	Crystallographic asymmetric unit
BBLA	Baboon milk alpha-lactalbumin
CD	Circular dichroism
ChB	N, N' diacetylchitobiose
CNBr	Cyanogen bromide
EDTA	Ethylenediaminetetraacetic acid
$F(hkl)$ or $F_{obs}$ , $F(hkl)$	Observed structure factor and its amplitude
$F_{calc}$	Calculated structure factor
FFT	Fast Fourier transform
GOLA	Goat milk alpha-lactalbumin
glc	glucose
GlcNAc	N-acetyl glucosamine
GPLA	Guinea-pig milk alpha-lactalbumin
GT	UDP-galactose- $\beta$ -N-acetylglucosaminide- $\beta$ 1,4 galactosyltransferase (E.C. 2.4.1.38)
HEWL	Hen egg-white lysozyme
HLA	Human milk alpha-lactalbumin
HPLC	High performance liquid chromatography
LA	Alpha-lactalbumin
LS	Lactose synthase (E.C. 2.4.1.22)
LYZ	C-type lysozyme
MAD	Multiwavelength anomalous diffraction
MIR	Multiple isomorphous replacement
mLA	methionine alpha-lactalbumin (recombinant bovine LA)
MR	Molecular replacement
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis

PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RBR	Rigid-body refinement
rms(d)	Root mean square (difference)
SDS	Sodium dodecyl sulphate
Tris	Tris-(hydroxymethyl)aminomethane
UDP	Uridine diphosphate
X30Y	Denotes amino acid at position 30 was changed from X to Y

**Symbols:**

$\alpha$	Phase angle
$\alpha, \beta, \gamma$	Eulerian angles
$B$	Isotropic temperature factor
$E$	Normalised structure factor
$h, k, l$ or $h$	Miller indices
$I$	Diffraction intensity
$K_m$	Michaelis constant
$\lambda$	Wavelength
$R$	Crystallographic residual
$\sigma$	Standard deviation
$x, y, z$	Cartesian coordinates

**One and Three Letter Amino Acid Codes:**

A	Alanine (Ala)	M	Methionine (Met)
C	Cysteine (Cys)	N	Asparagine (Asn)
D	Aspartic acid (Asp)	P	Proline (Pro)
E	Glutamic acid (Glu)	Q	Glutamine (Gln)
F	Phenylalanine (Phe)	R	Arginine (Arg)
G	Glycine (Gly)	S	Serine (Ser)
H	Histidine (His)	T	Threonine (Thr)
I	Isoleucine (Ile)	V	Valine (Val)
K	Lysine (Lys)	W	Tryptophan (Trp)
L	Leucine (Leu)	Y	Tyrosine (Tyr)

# Contents

Abstract .....	i
Acknowledgements.....	iii
Abbreviations .....	iv

## **Chapter 1 Introduction .....1-28**

1.1 Previous Studies on LA.....	3-8
1.1.1 Role in the Lactose Synthase Complex.....	3
1.1.2 Regulation of Lactose Biosynthesis .....	5
1.1.3 Role of Metal Binding .....	7
1.2 Structural Characteristics of LA .....	9-19
1.2.1 Primary Sequence .....	9
1.2.2 Relationship with C-type Lysozymes .....	13
1.2.3 Three-dimensional Structure .....	15
1.2.4 The N and A States .....	18
1.3 Structural Basis of LA Function .....	19-27
1.3.1 Chemical Modification.....	19
1.3.2 Functional Sites in LA.....	20
1.3.3 Functional Sites in GT.....	24
1.3.4 Monosaccharide Bridge Model of LA Action .....	25
1.4 Present Work .....	27-28

## **Chapter 2 Recombinant Bovine LA ..... 29-65**

2.1 Previous Expression Systems.....	29-30
2.2 Construction of a Non-fusion Expression System.....	31-34
2.2.1 Introduction.....	31
2.2.2 Construction of the Expression Vector .....	31
2.3 Generation and Purification of Recombinant LA .....	34-38
2.3.1 Expression and Initial Extraction.....	34
2.3.2 Folding .....	35
2.3.3 Final Purification .....	37
2.4 Generation and Expression of Mutant mLAs.....	39-47
2.4.1 Mutant Design.....	39
2.4.2 PCR Mutagenesis .....	41
2.4.3 Expression, Folding and Purification .....	43
2.4.4 Summary .....	46
2.5 Kinetic Characterisation of mLA and the mLA Mutants .....	47-65

2.5.1 Kinetic Mechanism of GT and The Lactose Synthase Complex .....	47
2.5.2 Interpretation of Steady State Kinetic Measurements of LA Activity ..	49
2.5.3 Functional Properties of mLA and the Mutants .....	51
2.5.4 Significance of Kinetic Parameters and Assessment of Errors .....	59
2.5.5 Summary .....	62
<b>Chapter 3 Crystallisation .....</b>	<b>66–84</b>
3.1 Introduction .....	66-70
3.1.1 Screening Methods .....	67
3.1.2 Sample Purity .....	68
3.1.3 Crystalline Forms of LA .....	69
3.2 Initial Crystallisation Screens .....	70-82
3.2.1 Experimental Design .....	70
3.2.2 Bovine Milk LA (BOLA) .....	73
3.2.3 Recombinant LA (mLA) .....	76
3.2.4 mLA Mutants .....	81
3.2.5 Goat LA (GOLA) .....	82
3.3 Summary .....	82-84
<b>Chapter 4 Data Collection and Processing .....</b>	<b>85–105</b>
4.1 Diffraction Theory .....	85-90
4.1.1 Laue Conditions .....	86
4.1.2 Bragg's Law .....	86
4.1.3 Reciprocal Lattice and Ewald Sphere .....	87
4.1.4 Electron Density Equation .....	89
4.2 In-House Data Collection .....	90-99
4.2.1 Introduction .....	90
4.2.2 Experimental Methods .....	92
4.2.3 Data Processing and Analysis .....	93
4.2.4 Diffraction Data .....	95
4.2.4.1 BOLA Trigonal ( <i>P</i> 321) .....	95
4.2.4.2 mLA Monoclinic III ( $\text{Na}^+$ Form) .....	97
4.2.4.3 mLA Hexagonal ( <i>P</i> 622) .....	98
4.2.4.4 GOLA Monoclinic ( <i>P</i> 2 <sub>1</sub> ) .....	99
4.3 Synchrotron Data Collection .....	100-105
4.3.1 Introduction .....	100
4.3.2 mLA Monoclinic ( $\text{Mg}^{2+}$ Form) .....	101
4.3.3 Summary .....	105
<b>Chapter 5 Molecular Replacement .....</b>	<b>106–135</b>
5.1 The Molecular Replacement Method .....	106-112
5.1.1 Patterson Function .....	107
5.1.2 Rotation Function .....	108

5.1.3 Translation Function .....	110
5.1.3.1 Patterson Searches .....	110
5.1.3.2 Correlation Searches .....	111
5.2 Software .....	112-116
5.2.1 Automatic Molecular Replacement ( <i>AMoRe</i> ) .....	112
5.2.2 <i>X-PLOR</i> Molecular Replacement .....	115
5.3 Results .....	117-134
5.3.1 Recombinant LA (mLA) .....	117
5.3.1.1 Monoclinic III Form .....	117
5.3.1.1.1 Self-Rotation Function .....	118
5.3.1.1.2 Cross-Rotation and Translation Functions .....	119
5.3.1.1.3 The Solution .....	120
5.3.1.2 Hexagonal Form (A109P-mLA).....	122
5.3.2 Bovine Milk LA.....	125
5.3.2.1 <i>AMoRe</i> .....	127
5.3.2.2 <i>X-PLOR</i> .....	129
5.3.3 Goat Milk LA.....	131
5.3.3.1 Self-Rotation Function .....	131
5.3.3.2 Cross-Rotation and Translation Functions.....	132
5.3.3.3 The Solution .....	133
5.4 Summary.....	134-135
<b>Chapter 6 The Quest for Derivatives.....</b>	<b>136-144</b>
6.1 Introduction .....	136-137
6.2 Experimental Methods.....	138-140
6.3 Results .....	141-142
6.4 Summary.....	142-144
<b>Chapter 7 Refinement and Model Building .....</b>	<b>145-180</b>
7.1 Refinement Methods.....	145-155
7.1.1 Introduction.....	145
7.1.2 <i>X-PLOR</i> .....	147
7.1.2.1 The Energy Function.....	147
7.1.2.2 Monitoring the Refinement.....	148
7.1.2.3 Generalised Refinement Protocol.....	150
7.1.2.4 Map Calculation.....	154
7.1.2.5 Manual Rebuilding .....	154
7.1.2.6 Assessing Model Quality .....	155
7.2 Monoclinic mLA .....	155-167
7.2.1 Refinement.....	155
7.2.2 Assessment of the Quality of the Final Model .....	160
A. Agreement with Electron Density.....	161
B. Ramachandran Plot and General Stereochemistry .....	162
C. Temperature Factors.....	163



D. Similarity of NCS-Related Molecules .....	164
7.3 Goat mL A .....	167-175
7.3.1 Refinement .....	167
7.3.2 Assessment of the Quality of the Final Model .....	170
A. Agreement with Electron Density .....	170
B. Ramachandran Plot and General Stereochemistry .....	171
C. Temperature Factors .....	172
D. Similarity of NCS-Related Molecules .....	173
7.4 Guinea-pig LA .....	175-180
7.4.1 Refinement .....	175
7.4.2 Assessment of the Quality of the Final Model .....	177
A. Agreement with Electron Density .....	177
B. Ramachandran Plot and General Stereochemistry .....	178
C. Temperature Factors .....	179
<b>Chapter 8 Description of the Structures .....</b>	<b>181-207</b>
8.1 Introduction .....	181-182
8.2 Similarity of the Structures .....	182-195
8.2.1 Overall Similarity .....	182
8.2.1.1 Secondary Structure .....	182
8.2.1.2 Tertiary Structure .....	184
8.2.2 Conformation of the Calcium Binding Site .....	191
8.2.3 Solvent Structure .....	194
8.3 Functional Regions in LA .....	196-203
8.3.1 Aromatic Clusters I and II .....	196
8.3.2 Flexible Loop Region .....	198
8.3.3 The Cleft Region .....	200
8.4 General Discussion .....	204-207
<b>Appendix .....</b>	<b>208-218</b>
<b>References .....</b>	<b>219-236</b>
<b>Reprints .....</b>	<b>237</b>

# Chapter 1

## Introduction

Enzyme activity is regulated by a multitude of different mechanisms. Control is typically maintained either via covalent mechanisms, such as protein phosphorylation or proteolytic cleavage, or via non-covalent mechanisms. Non-covalent regulation of enzyme activity is often mediated through reversible heterologous protein-protein interactions. This type of regulation is commonly mediated either directly or indirectly by conformational changes. In some cases, such as the activation of cyclin-dependent kinases by cyclin, the extent of molecular adjustment in the enzyme's structure can be quite startling (Jeffrey *et al.*, 1995). However, detailed understanding of the structural basis that underlies these modulations is, in most cases, quite limited. Clearly, a general feature of this type of regulation will involve a certain degree of conformational flexibility in the relevant molecules but it is not apparent whether this generalisation applies in all cases.

The enzyme lactose synthase (LS; E.C. 2.4.1.22) provides an ideal, albeit unusual, system in which to study heterologous protein-protein interactions. The catalytic activity of this enzyme, and the concomitant rate of milk lactose biosynthesis, is

regulated by a unique mechanism. In this enzyme complex, the catalytic potential of a glycosyltransferase, UDP-galactose- $\beta$ -N-acetylglucosaminide- $\beta$ 1,4 galactosyltransferase, is temporarily 'hijacked' and its substrate specificity is altered through an interaction with a 'specifier' protein ( $\alpha$ -lactalbumin). As a result, the transferase's affinity for glucose increases dramatically and this facilitates the biosynthesis of lactose. Such 'affinity control' of pre-existing enzyme activity is unusual as the majority of regulation mechanisms, mediated by modulatory proteins, involve unequivocal changes in activity. In the LS complex, the role of the specifier protein, LA, is akin to that played by metal ions in the activation of certain enzymes both in terms of its kinetic effects and functional nature (Khatra *et al.*, 1974).

LA, the 'specifier' protein in the LS complex and the subject of this work, is a globular, low molecular weight (~14KDa) protein component of the whey (non-casein) fraction of milk. It is expressed exclusively in the lactating mammary gland and, after fulfilling its modulatory role, is secreted into the milk. LA possesses a number of other interesting biological features in addition to its role in the LS complex. It is homologous to the C-type lysozymes, despite a radical change in biological activity, and represents an extreme example of functional divergence (Brew *et al.*, 1967). The structure-function relationships of lysozymes and LA are of particular interest both in terms of protein evolution and the way in which new biochemical functions arise. LA also readily forms a number of relatively stable, molten globule intermediates that are relevant to protein folding. The molten globule state of LA is extensively studied and is considered to be an archetypal folding intermediate (Kuwajima, 1989; Radford and Dobson, 1995). Finally LA is a calcium-binding metalloprotein (Hiroaka *et al.*, 1980) and, unlike lysozymes,

binds a number of different metal ion species that affect its conformational properties (Kronman, 1989).

The remaining sections in this chapter give a brief overview of the function and molecular properties of LA with particular emphasis on its role in the LS complex. Further details can be found in a number of recent reviews (Kronman, 1989; McKenzie and White, 1991; Brew and Grobler, 1992).

## **1.1 Previous Studies on LA**

### **1.1.1 Role in the Lactose Synthase Complex**

Milk is a complex mixture of proteins, carbohydrates, lipids and inorganic components and its composition varies significantly between various extant mammals. Lactose is the major free carbohydrate of most milks and accounts for about 3 to 7% of the total milk weight (Jenness, 1970). In addition to its nutritional role, lactose is also the principal osmotically active substance in milk and is responsible for generating most of the aqueous phase (Linzell and Peaker, 1971). Early studies showed that, although the majority of the carbon in lactose is derived from blood glucose, the precursors of the glucose and galactose moieties of lactose are not the same. Wood and co-workers suggested that lactose is enzymatically synthesised from UDP-galactose and glucose (see Brew and Hill, 1975). Subsequent studies demonstrated that the enzyme responsible for lactose biosynthesis, lactose synthase (LS), exists as a microsomal enzyme in the mammary glands of lactating cows and guinea-pigs (Watkins and Hassid, 1962) and in a soluble form in unpasteurised cow milk (Babad and Hassid, 1964; 1966). Brodbeck and Ebner (1966) demonstrated that LS is composed of two protein components (designated

‘A’ and ‘B’). In isolation, neither the ‘A’ or ‘B’ proteins has synthase activity, but when these two components are recombined full activity is regained.

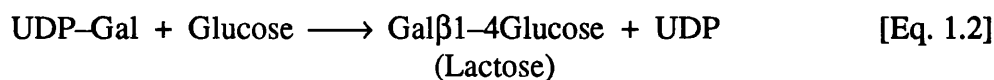
Further studies identified LA as the ‘B’ protein of the LS complex (Ebner *et al.*, 1966; Brodbeck *et al.*, 1967). The so-called ‘A’ protein is a galactosyltransferase that catalyses the following manganese-dependent reaction in the absence of LA (Brew *et al.*, 1968):



In this reaction, galactose is transferred from UDP-galactose into a  $\beta$ -linkage with the 4-hydroxy group of N-acetylglucosamine (GlcNAc). This transferase, UDP-galactose- $\beta$ -N-acetylglucosaminide- $\beta$ 1,4 galactosyltransferase (GT; EC 2.4.1.38), is a 55KDa type II membrane glycoprotein of the *trans*-golgi membranes of the mammary gland and other tissues (Strous, 1986). In the golgi apparatus, GT, along with other glycosyltransferases, participates in the biosynthesis of oligosaccharide chains of secretory and membrane-bound glycoconjugates (Paulson and Colley, 1989). GTs are also found on the cell surface where they function as recognition molecules during a variety of cell-cell and cell-matrix interactions (Pierce *et al.*, 1980; Shur, 1984; Shaper and Shaper, 1992).

*In vivo*, golgi GT’s main substrate is  $\beta$ -linked N-acetylglucosaminyl moieties of N-linked oligosaccharides of glycoproteins and glycolipids. This terminal glycosylation represents one of the final stages in glycoprotein processing and accounts for the majority of GT’s workload in the golgi apparatus of secretory cells. GT does not catalyse

the biosynthesis of lactose (Eq. 1.2) in the absence of LA due to a poor affinity for glucose ( $K_m^{\text{glc}}$  of about 2M).



In the LS complex, LA promotes the binding of glucose to GT, by reducing the  $K_m^{\text{glc}}$  by about three orders of magnitude (to the low millimolar range), and enables lactose to be synthesised at physiological concentrations of glucose. LA also acts as a competitive inhibitor of galactosyl transfer to glycoprotein and extended oligosaccharide substrates (Powell and Brew, 1976b; Bell *et al.*, 1976). This inhibition arises due to mutually exclusive binding of LA and oligosaccharides to GT. In the case of free GlcNAc as substrate, LA binding to the complex reduces the  $K_m$  and  $V_{\text{max}}$  so that activation is obtained at low substrate concentrations and inhibition at higher concentrations due to the decrease in product release (Brew and Grobler, 1992). The transfer of galactose to glycoproteins is not completely abolished in lactating mammary gland epithelial cells as LA dissociates from the LS complex at the end of each catalytic cycle. Therefore, the degree of glycoprotein terminal glycosylation and lactose biosynthesis catalysed by GT is determined by the relative concentrations of both GT and LA present in the golgi apparatus. Current understanding of the kinetic mechanism of the LS complex is discussed in more detail in Chapter 2.

### 1.1.2 Regulation of Lactose Biosynthesis

The regulatory role of LA is exerted through its presence or absence in the LS complex and this is reflected in the rate of lactose synthesis. LA is expressed specifically

in the rough endoplasmic reticulum (RER) of the lactating mammary gland epithelial cells as a pre-protein with a hydrophobic leader sequence (Hall and Campbell, 1986). The cellular expression of LA increases dramatically at parturition due to the hormonal influences of prolactin, insulin and hydrocortisone (Turkington *et al.*, 1968). Studies on mammary gland explants demonstrate that the initial expression and subsequent accumulation of LA is primarily under prolactin control (Dodd *et al.*, 1994). Induction of LA is critical for both the production of milk lactose and normal milk. Mice who lack the LA gene, due to gene targeting, cannot sustain their offspring due to the production of reduced amounts of thickened milk that contains no detectable amounts of lactose (Stinnakre *et al.*, 1994; Stacey *et al.*, 1995). Normal milk production can be restored by introducing a functional LA gene at the mouse LA locus (Stacey *et al.*, 1995).

In addition to hormonal control, the regulation of milk lactose production is controlled by the nature of the subcellular organisation of the LS complex (Brew, 1969). LA is translocated from the site of translation across the ER and transported to the golgi apparatus. Here it interacts with the catalytic domain of GT, which projects into the lumen, to form the LS complex. LA's participation in the complex is temporary and on dissociation, it is secreted into the milk, along with lactose and the other milk proteins, by exocytosis. Additional regulation of lactose synthesis may be exerted through a  $\text{Ca}^{2+}$ -dependent release of LA from the RER membranes (Berliner and Koga, 1987; Brew and Grobler, 1992). The amount of LA in milk varies from trace levels (some monotremes and marine mammals) to concentrations of about  $5\text{ g l}^{-1}$  (McKenzie and White, 1991). In humans, LA plays an important role in infant nutrition and accounts for about 30% of the total protein in milk (Heine *et al.*, 1991). Despite the inherent regulation of lactose

biosynthesis resulting from the organisation of the LS complex, other external factors, such as the availability of glucose and UDP-galactose in the golgi, will also influence the rate of milk lactose production *in vivo* (Brew and Grobler, 1992).

### 1.1.3 Role of Metal Binding

Twelve years after its function had been realised, Hiroaka and co-workers demonstrated that LA was a calcium binding metalloprotein (Hiroaka *et al.*, 1980). LA derived from milk contains a single, tightly bound calcium ion that has an association equilibrium binding constant ( $K_a^{app}$ ) of the order  $10^7 M^{-1}$  (Kronman *et al.*, 1981; Permyakov *et al.*, 1981a). The bound ion stabilises LA against denaturation and its removal has profound effects on the protein's physico-chemical properties (Kuwajima *et al.*, 1986; Kronman, 1989).

The observation that the calcium coordinating ligands are highly conserved in all LA sequences (Shewale *et al.*, 1984; Stuart *et al.*, 1986) suggests that calcium binding is important for biochemical function of LA. Calcium is required for the efficient folding and correct disulphide bond formation of LA *in vitro* (Rao and Brew, 1989). However, correctly-folded  $Ca^{2+}$ -free LA (apo-LA) is just as active in lactose biosynthesis as the holo protein (Kronman *et al.*, 1981; Murakami and Berliner, 1983; Musci and Berliner, 1985). Although this effect is unlikely to have any physiological relevance, as all LA will be in the holo  $Ca^{2+}$ -bound state when it interacts with GT in the lumen of the golgi apparatus, it demonstrates that the modulatory properties of LA are not directly dependent on calcium binding. LA's calcium binding properties are also unlikely to be important for infant nutrition as only between 0.15% and 1% of the calcium content of



milk is associated with LA (McKenzie and White, 1991). Therefore, it appears that the role of calcium binding *in vivo* is to mediate the release and folding of nascent LA at the site of translation (Brew and Grobler, 1992). Similar calcium-dependent regulation mechanisms mediate the transport of a number of secretory proteins in the endoplasmic reticulum (Sambrook, 1990).

LA also binds more than fifteen different metal ion species in addition to calcium. These metals perturb both the conformational state and the physico-chemical properties of the protein. These effects are exceedingly complex and beyond the scope of the present work. For further details, the reader is referred to a review by Kronman (1989 and references therein). Some aspects of LA's additional metal binding properties may have physiological significance. GT, the catalytic component of the LS complex, contains binding sites for both  $Mn^{2+}$  (high affinity site I) and  $Ca^{2+}$  (low affinity site II) (Powell and Brew, 1976a). Although endogenous non-metal factors probably activate GT at the low affinity site *in vivo* (Kronman, 1989; Kuhn *et al.*, 1992), metal binding at the high affinity site is critical for GT catalysis. This cation dependence of GT raises the possibility that metal-liganded states of LA may have some functional relevance. In addition to manganese, zinc can act as a primary cation activator of GT (O'Keefe *et al.*, 1980a). It is noteworthy that both these metals bind to LA at sites that are distinct from the high affinity calcium binding site (Murakami and Berliner, 1983; Gerken, 1984; Musci and Berliner, 1985; Ren *et al.*, 1993). Both metals also have significant effects on the conformation of LA and may therefore play a role in either the formation of the LS complex or the subsequent catalytic events.

## 1.2 Structural Characteristics of LA

### 1.2.1 Primary Sequence

A number of complete amino acid sequences have been determined for LAs from the milk of different species. Representative sequences are available for most of the subclasses of the extant mammals including the eutherians (11), marsupials (2) and the monotremes (1). An alignment of all the available LA sequences is shown in Figure 1.1 along with examples of conventional and calcium-binding C-type lysozymes (LYZ).

The length of the protein varies at the C-terminus where LA's from most species have three residues following Cys-120; the rabbit protein has two (Hopp and Woods, 1979), the marsupial and monotreme LAs have none (Shewale *et al.*, 1984; Shaw *et al.*, 1993) and the rat protein has an additional 20 residues (Prasad *et al.*, 1982). There are a few other notable differences between the sequences. Platypus LA contains six additional residues not present in the eutherian sequences; a two residue insertion occurs between 14 and 15 (bovine LA numbering) and a three residue insertion occurs between 63 and 64 after optimal alignment. The sixth additional residue in platypus LA, which is also present in the marsupial LA sequences, occurs between residues 96 and 97 after optimal alignment. The marsupial LAs also differ from the eutherian LAs in the location of the amino terminal gap (between residues 13 and 14) that is required for optimal alignment of the sequences (see Figure 1.1). These deletions must have occurred since the divergence of the marsupial, monotreme and eutherian lines (Shewale *et al.*, 1984). LA from pig's milk is unusual among the eutherian sequences in that proline 67 is deleted (Godovac-Zimmermann *et al.*, 1990). Nonetheless, it is noteworthy that all

insertions/deletions in the LA sequences are located in loop regions in the three-dimensional structure and are unlikely to affect the overall conformation of the molecule.

All the LA sequences, apart from the guinea-pig (Brew, 1972), wallaby and those of Perissodactyl origin (horse, donkey and camel), contain a characteristic N-glycosylation Asn-X-Ser sequence motif (residues 45-47). Rabbit, rat and platypus LAs are uniformly glycosylated at this site (Hopp and Woods, 1979; Prasad *et al.*, 1982; Shaw *et al.*, 1993). In contrast, glycosylated forms of bovine LA account for only 10% of the total protein. The remaining LAs do not exhibit significant post-translational modification.

The overall sequence identity in the LA family ranges from 40% (e.g. platypus vs. wallaby) to 94% (e.g. bovine vs. goat). This relatively wide variation in amino acid sequence suggests that LA's modulatory properties are dependent on a relatively small portion of the protein chain. There are 30 invariant residues in all the LA sequences (those boxed in Figure 1.1) of which 12 are conserved across the whole LA/LYZ superfamily (boxed and shaded). A further group of residues (indicated by hashes in Figure 1.1), involved in the ligation of calcium, are strictly conserved in all LAs and moderately conserved in the calcium-binding LYZs. The functional significance of these invariant residues is discussed further in Section 1.3.2.

**Figure 1.1 (overleaf) — Alignment of all LA sequences with some selected LYZ sequences**

The alignment shows groups of a few select C-type LYZs (top group: sequences 1-9), Ca<sup>2+</sup>-binding LYZs (middle group: 10-12) and all LA sequences (bottom group: 13-26). The numbering on the top of each aligned block of sequences is that of hen egg-white LYZ and the numbering on the bottom corresponds to that of bovine LA. Areas of sequence identity are boxed in both the LA and LYZ sequences. The boxed and shaded areas correspond to residues that are invariant across the entire LA/LYZ superfamily. Residues in LA that form the calcium binding site are indicated by hashes (#). All the sequences are derived from the PIR and SWISS-PROT databases and were aligned using the *GCG* package of programs (Devereux *et al.*, 1984). A minimum number of gaps (.) were introduced to achieve an optimal alignment. This figure was produced by the program *ALSCRIPT* (Barton, 1993)



### 1.2.2 Relationship with C-type Lysozymes

The evolutionary relationship between LAs and chicken (C-) type lysozymes (LYZ) was first demonstrated by sequence studies of bovine LA and hen egg-white lysozyme (Brew *et al.*, 1967). This relationship has been unequivocally confirmed by protein sequences of other LAs, gene sequences and structures of LAs and LYZs (Hall *et al.*, 1982; Qasba and Safaya, 1984; Hall *et al.*, 1987) and by the three-dimensional structures of LAs and LYZs (Acharya *et al.*, 1989).

The C-type LYZs form a subgroup of a much larger family of anti-microbial endo-*N*-acetylmuramidases which have a wide distribution in the animal kingdom. In addition to the C-type LYZs, typified by hen egg-white LYZ (HEWL) but also found in mammalian secretions, the LYZ family includes the goose-type LYZs, found in the eggs of the swan, goose and ostrich, and the phage-type LYZs, found in the lysate of bacterial viruses (typified by bacteriophage T4 lysozyme). Although there is no sequence homology between these three classes of LYZs, they share similar tertiary structures and active site architectures and are believed to have diverged from a common ancestor (Grütter *et al.*, 1983).

LAs share about 40% sequence identity with both mammalian and avian C-type LYZs (Stuart *et al.*, 1986). However, their functional properties are quite different; LA has no discernible catalytic activity whereas LYZs have no detectable effect on the activity of GT. LA has lost both residues involved in LYZ's catalytic activity during evolution. Glu-35 (residue 33 in LA) has been replaced by a number of different residue types while Asp-52 (residue 49 in LA) has been exclusively replaced by a glutamate in all

LA sequences apart from that of the platypus (Figure 1.1). The trace levels of catalytic activity reported for LAs from a number of sources (McKenzie and White, 1987) are surprising considering the essential role of Glu-35 in LYZ catalysis (Malcolm, 1989). Minimal catalytic activity can however be endowed on goat LA by replacement of exon 2 (residues 28 to 86) with its homologue from hen egg-white LYZ; but this chimeric protein is unable to act as a specifier in the LS complex (Kumagai *et al.*, 1992).

As LAs are confined entirely to mammalian species, it is reasonable to assume that the progenitor of the LA/LYZ superfamily was an ancestral lysozyme gene. The common view is that the LAs diverged from the LYZs, via gene duplication, prior to the divergence of the tetrapods and the fishes (400 million years ago) but probably more recently than the divergence of the insects from the line that gave rise to the vertebrates (Dautigny *et al.*, 1991; Prager and Wilson, 1988; Grobler *et al.*, 1994). However, their exact evolutionary relationship is the subject of much debate (see Brew and Grobler, 1992). The point of contention (Figure 1.2) is whether the ability of some LYZs and the LAs to bind calcium is a recently acquired trait or an ancient feature of the superfamily which has been lost from the conventional LYZ line (Nitta and Sugai, 1989; Grobler *et al.*, 1994). The latter view is supported by the observation that only two amino acid substitutions are required to create a high affinity  $\text{Ca}^{2+}$  site in human LYZ that is closely similar in structure to that of LA (Kuroki *et al.*, 1989; Inaka *et al.*, 1991).

Clearly, the development of mammalian lactose biosynthesis was dependent on the acquisition of LS activity. More specifically, the emergence of this new biochemical activity resulted from the evolution of the specifier protein (LA) as GTs are found in a number of non-mammalian species (Brew and Grobler, 1992). However, as described

above, recent studies suggest the LA gene line originated considerably earlier in evolution than either the origins of lactation or the development of lactose synthesis (Grobler *et al.*, 1994). Hopper and McKenzie (1974) reported that a protein fraction from echidna milk was found to have both a LYZ activity and a low level of LA activity in the LS complex. Although this surprising activity requires further investigation (McKenzie and White, 1991), the existence of a 'dual activity' protein, representing an intermediate in LA evolution, remains a possibility.

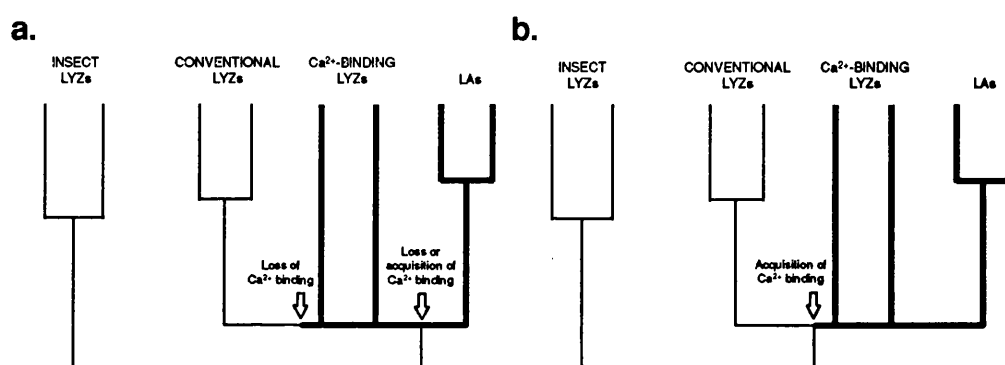


Figure 1.2 — Evolutionary development of the LA/LYZ superfamily

Two general rooted trees that schematically illustrate the possible development of the superfamily. **a.** Scheme in which the ability to bind calcium was an ancient feature of the superfamily but has been lost in some evolutionary lines (Grobler *et al.*, 1994). **b.** More conventional scheme in which the progenitor of the LAs and Ca<sup>2+</sup>-binding LYZs acquired the ability to bind calcium (Nitta and Sugai, 1989). Thick lines indicate groups that bind calcium. Figure reproduced from Grobler *et al.* (1994)

### 1.2.3 Three-dimensional Structure

Initial attempts to solve the structure of LA were hindered by a number of unforeseen practical problems (Phillips *et al.*, 1987). In the meantime, two homology models, based on the observed sequence similarity between LA and LYZ, were constructed for LA (Browne *et al.*, 1969; Warne *et al.*, 1974). These studies



demonstrated that the differences between LA (bovine) and LYZ (HEWL) sequences were compatible with their having similar conformations. These conclusions were confirmed when, nearly a decade after the crystallographic studies on LA had been initiated, the structure of baboon milk LA was reported first at 4.5Å resolution (Smith *et al.*, 1987) and subsequently at high resolution (Acharya *et al.*, 1989; 1990). More recently, the structure of human LA has been determined to a resolution of 1.7Å (Acharya *et al.*, 1991).

LA is a globular protein and ellipsoidal in shape with dimensions of 23Å x 26Å x 40Å (Acharya *et al.*, 1989). Like LYZ, the structure belongs to the  $\alpha + \beta$  class of protein folds. The molecule is divided into two 'domains' by a deep cleft, that is analogous to the active site of LYZ. Residues 1-34 and 86-123 form an all helical domain ( $\alpha$ -domain) while residues 35-85 form a smaller irregular  $\beta$ -domain (Figure 1.3). The structure is stabilised by the bound calcium ion and four disulphide bonds. Two of the disulphides (6-120 and 28-111) are in the  $\alpha$ -domain, one is in the  $\beta$ -domain (61-77) and one (73-91) links the two domains together. The calcium binding site is located in a helix-turn-helix motif that bridges the two domains. The calcium ion is coordinated by five protein and two solvent ligands (Stuart *et al.*, 1986). The topology of the binding loop ('elbow') is distinct from the E-F hand motif that binds calcium with comparable affinity in proteins such as calmodulin and parvalbumin (Stuart *et al.*, 1986; Strynadka and James, 1989).

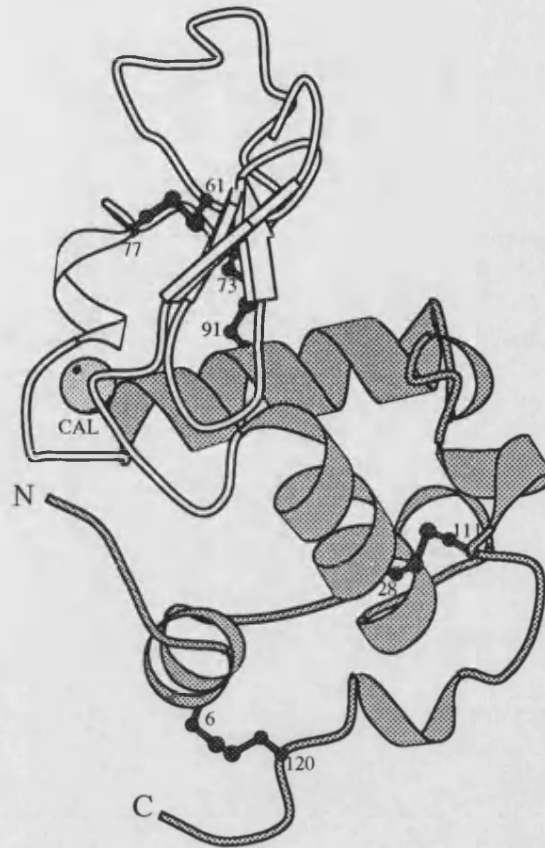


Figure 1.3 — Overall topology of the LA structure.

The three-dimensional structure of human LA is schematically represented. The all-helical  $\alpha$ -domain (grey) and the irregular  $\beta$ -domain (white) are shown along with the four disulphide bridges, calcium binding site (light grey sphere) and the N and C termini. The figure was produced with *MOLSCRIPT* (Kraulis, 1991).

Identification of the calcium ligands in LA allowed Stuart and coworkers to propose that some LYZs may bind calcium (Stuart *et al.*, 1986). Subsequently, horse, donkey, dog, pigeon egg-white and echidna LYZs, which all have similar residues to LA's conserved calcium ligands, have been shown to bind calcium with a comparable affinity (Nitta *et al.*, 1988; Godovac-Zimmermann *et al.*, 1988; Grobler *et al.*, 1994). The discovery of this subgroup of conventional LYZs has had a significant impact on comparative studies of the evolutionary relationships within the LA/LYZ superfamily (Brew and Grobler, 1992).

The conformation of LA is very similar to HEWL and, as with other homologous protein families, all the amino acid insertions/deletions between the sequences occur loop regions of the structure. The root mean square deviation between the C<sup>α</sup> atoms of equivalent residues in LA and LYZ ranges from about 1.4Å to 2.0Å (Acharya *et al.*, 1989; Acharya *et al.*, 1994). One unique feature of LA's structure, in comparison to HEWL, is the blockage of the upper cleft region by the sidechain of Tyr-103; and the resultant formation of a new hydrophobic 'box' that includes Tyr-103, Trp-104, Trp-60 and Ile-95 (Browne *et al.*, 1969; Koga and Berliner, 1985; Acharya *et al.*, 1989).

NMR studies on LA have predominantly focused on the A-state (see below) and information about the conformation of native state of LA in solution is limited. A complete assignment of the spectra from LA in solution has not yet been accomplished. Nevertheless, NMR spectroscopy has been successfully used to study the local environments of aromatic and histidyl residues in LA and its metal ion binding properties (Bradbury and Norton, 1975; Gerken, 1984; Harushima and Sugai, 1989; Alexandrescu *et al.*, 1992).

#### 1.2.4 The N and A States

In spite of LA's well defined conformation in the crystalline state, the protein has extremely complex conformational properties and exhibits a multiplicity of different conformers in solution. Exposure of the native protein (N state) to acidic conditions (pH<4.0) results in a reversible, conformational change to the acid (A-) state. A similar transconformation can be induced by elevated temperature, addition of certain metal ions (e.g. Zn<sup>2+</sup>), low concentrations of denaturants, alkaline pHs and removal of the bound

calcium by chelators. Both the N- and A-states are characterised by a multitude of conformers with distinct properties depending on the protein's metal ligation state (for a review see Kronman, 1989).

The A-state (and related denatured forms) is of considerable interest in terms of protein folding as it represents a stable intermediate in the (un-)folding of LA. This so-called 'molten globule state' is characterised by a compact, native-like secondary structure with minimal tertiary interactions (Kuwajima, 1989). Similar conformational states have been observed for a number of other proteins. Due to the expanded nature of LA's molten globule, it has an increased propensity for self-association and ability to bind hydrophobic dyes. Even in the absence of extensive sidechain packing, the molten globule state of LA contains a significant amount of 'native-like' tertiary structure in the  $\alpha$ -domain (Baum *et al.*, 1989; Alexandrescu *et al.*, 1993; Peng *et al.*, 1995; Wu *et al.*, 1995).

## 1.3 Structural Basis of LA Function

### 1.3.1 Chemical Modification

Classical chemical modification studies have provided equivocal information about the functional involvement of specific amino acid residues of LA in the LS complex (reviewed in Brew and Hill, 1975; McKenzie and White, 1991). Nitration of the two tyrosine residues in human LA (18 and 103) has no effect on LA activity but more extensive treatment results in both loss of activity and perturbations of the three-dimensional structure. This loss of activity is specifically associated with the modification of Trp-104 (Prieels *et al.*, 1975). In bovine LA, modification of either Trp-60 or Trp-118

results in a significant reduction in affinity for GT or complete inactivation (Schechter *et al.*, 1974; Bell *et al.*, 1975). More informative results have been obtained from the modification of histidine residues in LA. Specific ethoxyformylation of His-32 with diethyl pyrocarbonate in human and bovine LA results in a complete, but reversible, elimination of activity (Schindler *et al.*, 1976). In contrast, specific carboxymethylation of His-32 produces only a partial loss of activity. These results indicate that the N1 (N<sup>δ</sup>1) atom of His-32 is directly involved in LA's function in the LS complex. The partial loss of activity that accompanies the carboxymethylation of the N3 (N<sup>ε</sup>2) atom of His-32 probably arises due to either steric or conformational effects resulting from the modification (Prieels *et al.*, 1979).

The role of certain amino acid residues in complex formation with GT has been investigated by trace-labelling studies. Of the thirteen amino groups studied, the reactivities of only two of these, Lys-5 and Lys-114, were significantly perturbed on formation of the LS complex (Richardson and Brew, 1980). Further studies identified Lys-108, a residue close to the Lys-5 and Lys-114 in the three-dimensional structure of LA, as the major site of cross-linking between LA and GT in the LS complex (Sinha and Brew, 1981).

### 1.3.2 Functional Sites in LA

Comparative sequence analysis has identified a series of invariant residues that are, by their conserved nature, probably critical for LA function (see Figure 1.1). Shewale and coworkers (1984) divided these conserved residues into four functional groups. More recently, Kronman (1989) and Brew and Grobler (1992) have re-evaluated

this classification in terms of the three-dimensional structure of baboon LA and the more diverse database of LA sequences (Table 1.1).

Table 1.1 — Residues conserved in LAs and / or LYZs

Group <sup>a</sup>	Residues <sup>b</sup>	Role
<b>IA</b>	Cys 6, 28, 61, 73, 77, 91, 111, 120 Ser-34 Gly-51 Trp-104	Essential for maintaining the characteristic LA / LYZ tertiary fold
<b>IB</b>	Gln-54 Ile-55	Directly or indirectly involved in substrate binding in LYZ
<b>II</b>	Leu-81, Asp-82 Asp-83, Asp-87 Asp-88	Components of the Ca <sup>2+</sup> - binding site
<b>IIIA</b>	Phe-53 Ile-95 Gly-100 Ala-106	Possible structural roles
<b>IIIB</b>	Phe-31, His-32 Leu-115, Gln-117 Trp-118	Components of the GT binding site
<b>IIIC</b>	Glu-25 Gly-35 Lys-94	Residues with unknown function

<sup>a</sup> Groups are different to those defined by Shewale *et al.*, 1984

<sup>b</sup> Residue numbers refer to bovine LA numbering

Figure adapted from Table 3 in Brew and Grobler (1992).

Briefly, group I comprises residues, conserved across the entire LA/LYZ superfamily, that are either responsible for maintaining the integrity of the tertiary structure (IA) or involved in substrate binding in the LYZs (IB) (Warme *et al.*, 1974; Acharya *et al.*, 1989). Residues in group II form the calcium binding site in LAs and the Ca<sup>2+</sup>-binding LYZs. These residues are strictly conserved in the LA sequences but are more variable in the Ca<sup>2+</sup>-binding LYZs. Nevertheless, the residues that are responsible for calcium ligation (Asp-82, 87, 88) are all invariant. Finally, group III comprises of residues that are conserved only in LAs and are therefore likely to be important in LA's

function. These group III residues can be subdivided into three classes: those that perform a unique structural role in LA (IIIA), components of the GT binding site (IIIB) and those with unknown function (IIIC).

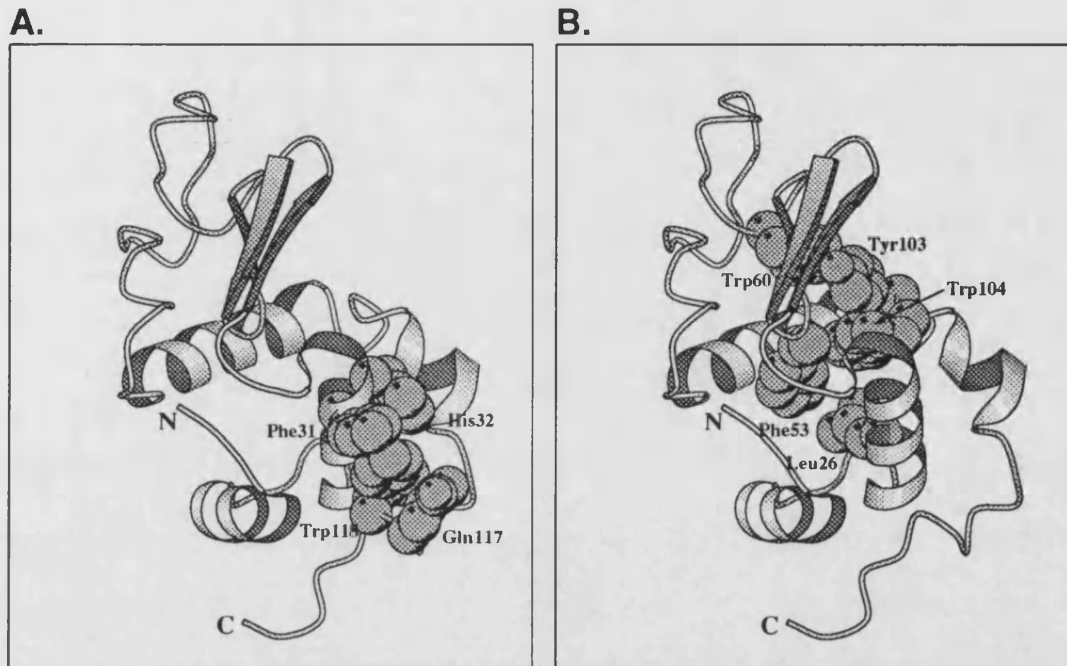


Figure 1.4 — Location of proposed GT binding sites in terms of LA's structure

Schematic representations of the three-dimensional structure of human LA showing the secondary structure and the locations of the two regions implicated in LA function in the LS complex.

**A.** Aromatic cluster I. The positions of the four conserved residues (Phe-31, His-32, Gln-117 and Trp-118) are illustrated; **B.** Aromatic cluster II ('hydrophobic box'). This cluster is composed of two invariant residues (Phe-53, Trp-104) and three further positions that contain buried hydrophobics (26, 60 and 103). This group of residues forms the core of LA. The orientation of LA is identical in **A** and **B** and the residues are illustrated as 'space-filling' representations. The figure was produced using *MOLSCRIPT* (Kraulis, 1991).

The results from chemical modification and trace labelling studies, as well the nature of the invariant residues in all LA sequences, suggest that LA's binding site for GT is composed of Phe-31, His-32, Gln-117 and Trp-118 (Warne *et al.*, 1974; Shewale *et al.*, 1984; Brew and Grobler, 1992). These residues are located on the surface of LA, adjacent to the lower cleft region, and form a predominantly aromatic cluster (aromatic

cluster I; Figure 1.4a). This cluster is flanked by two lysine residues (Lys-5 and Lys-114) whose reactivity is perturbed in the LS complex (Richardson and Brew, 1980). The 'hydrophobic box' (aromatic cluster II; Figure 1.4b) has also been linked to LA's function in the LS complex (Koga and Berliner, 1985). This region is distinct from aromatic cluster I and is part of the hydrophobic core that stabilises the LA fold. It is unclear whether this region is involved in LA's interaction site for GT as most of the residues are buried within the interior of the protein. In addition, the nitration of Tyr-103 does not appear to affect LA activity.

Although LA cannot bind saccharide substrates in isolation, the possible provision of a partial monosaccharide site in cleft of LA, that facilitates saccharide binding in the LS complex, remains an intriguing possibility. It is interesting that, out of the four exons that constitute the LA and LYZ genes, exon 2, which encodes the substrate binding residues of LYZ, is the most highly conserved (Hall *et al.*, 1982). In hen egg-white LYZ, there are six subsites (A-F) in the active site cleft that are capable of binding substrate (Imoto *et al.*, 1972). In the case of LA, the cleft region is shortened as subsites A and B are blocked by the sidechain of Tyr-103 (Warne *et al.*, 1974; Acharya *et al.*, 1989). Nonetheless, subsites D-F may be capable of supporting monosaccharide binding in the LS complex. Although this hypothesis is attractive, especially in terms of LA's evolutionary relationship with LYZ, there is no compelling evidence that such a situation prevails. Resonance energy transfer measurements between sites on LA and GT led O'Keefe and coworkers to conclude that the cleft of LA is probably located some distance from the acceptor binding site in the LS complex (O'Keefe *et al.*, 1980b).



Nevertheless, the existence of a partial functional monosaccharide binding site in the cleft cannot be ruled out.

### 1.3.3 Functional Sites in GT

Relatively little is known about the three-dimensional structure and functional sites in GT. cDNA and partial protein sequences have been determined for human, bovine and mouse GTs (for a review see Paulson and Colley, 1989). Sequence analysis indicates that, like other glycosyltransferases, the enzyme consists a short N-terminal domain, a hydrophobic transmembrane domain, and a stem region (~60 residues) and catalytic domain (~320 residues) that project into the lumen of the golgi. The N-terminal domain, absent from the soluble, proteolytically-cleaved forms of GT, is not required for catalytic function but may play a role in subcellular targeting (Evans *et al.*, 1994). Retention of the enzyme in the golgi apparatus is mediated by the transmembrane domain (Aoki *et al.*, 1992). The catalytic domain appears to be composed of two separate sub-domains; one involved in LA and acceptor substrate binding (N-terminal residues 79-250) and the other, comprising the C-terminal region (residues 275-402), which is involved in UDP-galactose binding (Yadav and Brew, 1990; 1991). A disulphide bond that connects the two catalytic sub-domains (Cys134-Cys247) is critical for GT and LS activity (Boeggeman *et al.*, 1993; Wang *et al.*, 1994). In addition, chemical modification and site-directed mutagenesis studies have demonstrated that a number of aromatic residues are critical for GT catalysis and/or UDP-galactose binding (Takase and Ebner, 1984; Aoki *et al.*, 1990; Zu *et al.*, 1995).

There are several lines of evidence that the galactosyl acceptor substrate and LA binding sites on GT are in close proximity and that they do not coincide with the UDP-galactose binding site. A photoaffinity analogue of UDP-galactose specifically labels GT's UDP-galactose binding domain in the presence and absence of acceptor substrates and LA (Lee *et al.*, 1983). LA and extended oligosaccharides bind to GT in a mutually exclusive manner (Khatra *et al.*, 1974; Bell *et al.*, 1976). Competitive inhibition is also observed between LA and  $\beta$ 1-glycosides of glucose and GlcNAc (O'Keefe *et al.*, 1980a; Lambright *et al.*, 1985). A methyl substituent is sufficient to elicit competitive binding. This observation suggests that, in the LS complex, a group on LA is only about 1.5Å from the  $\beta$ 1 hydroxyl of the galactosyl acceptor (Takase and Ebner, 1984).

Indirect evidence for the nature of the LA binding site comes from similar studies. Hydrophobic  $\beta$ 1-glycoside derivatives of GlcNAc and glucose cause as much as a 1000-fold decrease in the apparent  $K_m$  of the galactosyl acceptor while having minimal effects on the  $V_{max}$ . (Takase and Ebner, 1984). This effect is proportional to the size of the substituent. Therefore, it seems likely that the site adjacent to the  $\beta$ -OH position of the galactosyl acceptor, presumably the LA binding site, is relatively hydrophobic in nature. This is consistent with the view that the majority of the residues proposed to be involved in the interaction site on LA are hydrophobic in character (see Section 1.3.2).

#### 1.3.4 Monosaccharide Bridge Model of LA Action

A model has been proposed to explain LA's action in the LS complex that does not invoke any allosteric effects from LA binding (Lambright *et al.*, 1985; Brew and Grobler, 1992). The model is illustrated in Figure 1.5. It has been dubbed the

monosaccharide bridge model due to the stabilising nature of monosaccharide binding between LA and GT.

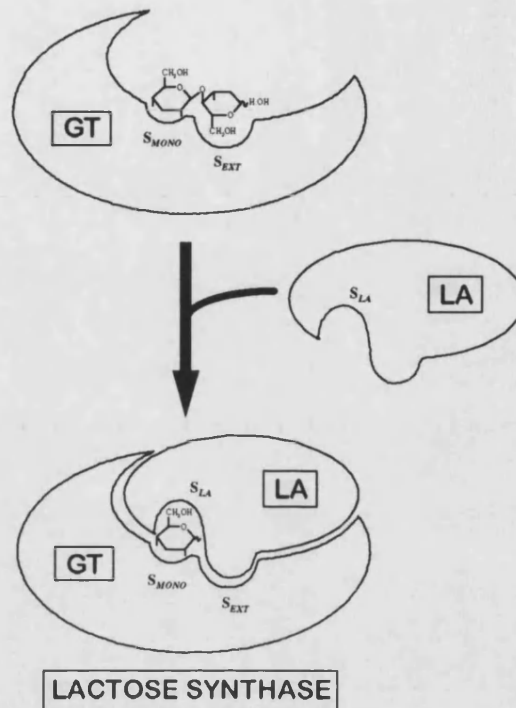


Figure 1.5 — Monosaccharide bridge model

A model that describes the molecular basis of LA's action in the LS complex is illustrated above. See text for details (Adapted from Brew and Grobler (1992)).

The galactosyl acceptor site of GT is envisaged to contain two or more subsites for the binding of monosaccharides; one ( $S_{MONO}$ ) capable of accommodating galactosyl acceptors such as free GlcNAc (and glucose albeit with low affinity) and another ( $S_{EXT}$ ) capable of binding an additional sugar in extended substrates. The binding site for LA (affinity control site) overlaps the  $S_{EXT}$  site, in a manner consistent with the mutually exclusive binding of LA and oligosaccharides, and brings a region of LA close to the  $S_{MONO}$  site. This site in LA ( $S_{LA}$ ) forms favourable stabilising interactions with

monosaccharide bound at  $S_{MONO}$  and gives rise to the observed synergism of binding between glucose and LA (Khatra *et al.*, 1974).

## 1.4 Present Work

The present work aims to clarify and extend our current understanding of the molecular mechanism of LA action in the LS complex. An ideal approach would be to determine the three-dimensional structure of the lactose synthase complex. Unfortunately this has been hampered by a variety of reasons including the heterogeneous nature of soluble GT purified from natural sources, problems associated with the solubility and lability of GT, and the relatively poor stability of the LS complex (Brew and Grobler, 1992). Although a number of these problems are now being addressed, such as the ability to produce relatively large quantities of soluble GT by recombinant methods (K. Brew, personal communication), crystallographic determination of the structure of the LS complex will not be accomplished in the near future. Consequently, information about the molecular basis of lactose synthesis must be obtained indirectly.

The 'monosaccharide bridge' model predicts that functional regions on the surface of LA, although closely interrelated, should be structurally distinct. If the model accurately describes the molecular action of LA in the LS complex, there should be one region on LA that is solely involved in binding to GT and another ( $S_{LA}$ ) that stabilises the binding of monosaccharides. Several potential functional sites in LA (aromatic cluster I and II) have already been implicated by chemical modification studies and comparative sequence analysis. In particular, a cluster of aromatic residues on the surface of LA (aromatic cluster I) appears to be critical for LA's action in the LS complex.

In collaboration with Dr. Keith Brew at the University of Miami Medical School, we have set out to ascertain which residues are involved in LA's interaction with GT and whether any regions in LA facilitate glucose binding in the LS complex. Site-directed mutagenesis, a powerful technique that facilitates the dissection of protein structure-function relationships, has already been used in Miami to study the functional importance of several residues on the surface of LA and to test the validity of the monosaccharide bridge model (Brew *et al.*, 1993). In the current work, these mutagenesis studies are expanded and the structural consequences of particular sidechain substitutions are investigated.

The following chapters document the progress made in identifying the functional regions of LA using protein engineering and correlating the resulting functional effects with the three-dimensional structure of LA.

## Recombinant Bovine LA

### 2.1 Previous Expression Systems

In general, detailed investigation of the structure-function relationships of a protein require that sufficient quantities of both wild type and variant forms of the protein are available. In such cases, bacterial expression systems can provide an ideal source of material for kinetic and structural studies. Suitable expression systems must allow the generation of relatively large quantities of homogeneously pure protein with the minimum of effort. Although several expression systems for LA have been reported in the literature, their success has been mixed. A yeast expression system for bovine LA yielded a product that was identical in amino acid sequence to the wild type milk protein and had a similar ability to modulate the affinity of GT. However, the final yields of this system were modest with expression levels of between 0.5 and 2mg/l (Viaene *et al.*, 1991). Similarly, an *E. coli* fusion system for goat LA (GOLA) only yielded about 1.5mg/l of protein after cleavage (Kumagai *et al.*, 1990). Furthermore, this product possessed only 12% of the activity expected for goat milk LA (Kumagai *et al.*, 1992).

A more successful bacterial expression system has been developed in Dr. Keith Brew's laboratory at the University of Miami School of Medicine (Wang *et al.*, 1989). The product, a fusion protein consisting of the mature form of bovine LA connected by a short linker to an amino terminal portion of cathepsin D, was produced at levels of about 50mg/l prior to the refolding and cleavage steps. The folded fusion protein had about 25% of the activity expected for the wild type protein but, after cleavage with trypsin, was essentially indistinguishable from bovine milk LA. However, the authors pointed out that one potential disadvantage of this system in the study of mutant LAs, is that mutations that destabilise the native conformation may exhibit less specific cleavage of the fusion protein. To reduce this possibility, the system was further developed so that it allowed a more specific cleavage of the folded fusion protein. A methionine residue was introduced prior to the coding region of the mature bovine LA using M13 site-directed mutagenesis. In addition, the coding region of LA was rendered resistant to CNBr cleavage by the replacement of methionine 90 with a valine. This substitution was considered to be relatively conservative as a valine is found in this position in several LAs from other species. Using this new construct (pC-LA), a recombinant form of bovine LA (M90V-LA) could be obtained, 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. The final yields were also improved by the introduction of this specific cleavage site and about 10mg folded protein could be consistently obtained per litre of bacterial culture. This expression system was subsequently used successfully to study the effects of various amino acid substitutions in aromatic cluster I of LA (Brew *et al.*, 1993).

## 2.2 Construction of a Non-fusion Expression System

### 2.2.1 Introduction

This was the current situation as the author travelled to Miami to purify sufficient quantities of recombinant and mutant LAs for X-ray crystallographic studies. However, it soon became apparent that the current expression system was not able to produce the quantities of protein required for structural analysis. Despite the reasonable final yields obtained for M90V-LA, the site-directed mutants could only be recovered in small amounts. These low yields appeared to be due to both the oxidative effects of CNBr and inefficient cleavage of the fusion product. As a result about 70% of the expressed product was lost during the purification procedure. Consequently, my arrival served to expedite the development of a non-fusion expression system that Keith and his postgraduate student had been considering for several months. Theoretically, a non-fusion system would avoid the CNBr cleavage step and should enable sufficient quantities of protein to be obtained both for kinetic and structural characterisation.

### 2.2.2 Construction of the Expression Vector

The expression vector was generated by cloning the coding sequence for bovine LA into the pET3a vector (Novagen, Madison, WI). The pET vector range is a powerful, high yield system for expressing recombinant proteins in *E. coli* (Studier *et al.*, 1990; Rosenberg *et al.*, 1987; Studier and Moffat, 1986). Target genes are introduced so that they are under the control of strong bacteriophage T7 transcription and translation signals. Subsequent protein expression is achieved by transferring the recombinant plasmid to a host that possesses a chromosomal copy of the T7 RNA polymerase gene.



Commonly used hosts are lysogens of the bacteriophage  $\lambda$ DE3, which contain the polymerase gene under the control of the inducible *lacUV5* promoter. Induction of the T7 polymerase gene with isopropylthio- $\beta$ -D-galactoside (IPTG) enables the target DNA in the expression vector to be specifically transcribed. The polymerase is so active that the target gene product can comprise more than 50% of the total cell protein a few hours after induction.

The cloning procedure is summarised in Figure 2.1. The coding region of bovine LA was amplified from the fusion vector (pC-LA) by the polymerase chain reaction (PCR) using primers designated NF-N and NF-C (see Appendix). These primers were designed to incorporate a unique *NdeI* restriction site at the 5' end and a *BamHI* site at the 3' end (after the stop codon) of the LA coding sequence. The *NdeI* restriction site also introduced an initiator methionine codon immediately prior to the first codon of LA. The amplification product was isolated by agarose gel electrophoresis and digested with *NdeI* and *BamHI*. This digested product was then cloned into pET3a vector that had been digested with the same restriction enzymes. The ligated product was transformed into competent DH5 $\alpha$  subcloning efficiency cells (Gibco BRL, Gaithersburg, MD) and the transformants were screened for correctly inserted DNA using *NdeI/BamHI* and *BamHI/SalI* double digestions (see Appendix). Twelve of the transformant colonies screened contained the correct insert.

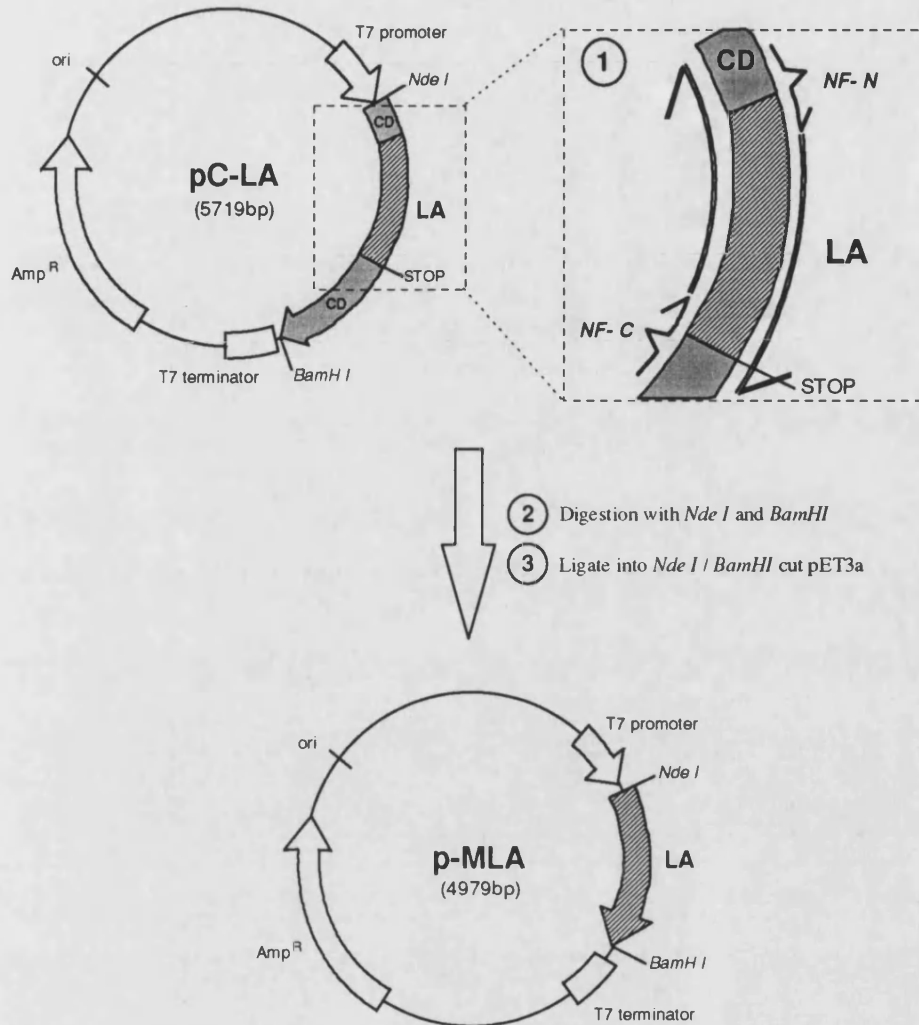


Figure 2.1 — Construction of p-MLA expression vector from pC-LA.

(1) The mature protein region of bovine LA was amplified from the pC-LA fusion protein vector using two oligonucleotide primers designed to introduce a 5' *Nde*I and a 3' *Bam*HI restriction site (CD represents the cathepsin D portion of the fusion protein). (2) The purified PCR product was digested using these restriction enzymes and (3) ligated into pET3a vector. See text and the Appendix for details.

The new expression vector (p-MLA) was transformed into 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 (Applied Biosystems) to confirm that the LA

coding region was correctly inserted and that no undesired mutations had been introduced during the PCR amplification.

### 2.3 Generation and Purification of Recombinant LA

#### 2.3.1 Expression and Initial Extraction

Cultures of *E. coli* strain BL21(DE3) transformed with p-MLA were grown overnight in Luria-Bertani (LB) medium containing 100µg/ml ampicillin at 37°C. An aliquot of the overnight culture was washed and diluted (1:100) in fresh medium containing 100µg/ml antibiotic and incubated at 37°C. When the  $A_{600}$  of the culture reached 1.0 (usually after 2hrs), expression of the host T7 polymerase gene was induced by adding IPTG to a final concentration of 0.4mM. Maximal levels of expression were observed at about 3hrs post-induction. After three hours, the cells were harvested by centrifugation at 4000rpm and the inclusion bodies, containing recombinant LA (mLA), were isolated using the lysozyme lysis method (Sambrook *et al.*, 1989; see Appendix). The inclusion body pellet was solubilised by incubation for two hours at 37°C in 20mM Tris-HCl pH 8.5 containing 8M urea and 20mM dithiothreitol (approximately 8–10mls per litre of bacterial culture). The resulting extract was loaded onto a column (26mm x 95mm) of Macro-Prep Q50 strong anion exchange support (Bio-Rad) that had been pre-equilibrated with running buffer (20mM Tris-HCl pH 8.5, containing 4M urea). The column was mounted on an EconoSystem™ (Bio-Rad) with a flow rate of 2.5ml/min. The column was washed with equilibration buffer until the absorbance at 280nm was reduced to the baseline level. Bound protein was then eluted with a linear gradient of buffer B (20mM Tris-HCl pH 8.5 containing 4M urea and 0.5M NaCl) at a rate of 1.1%

buffer B/min. Fractions of 2.5ml were collected and the elution process was constantly monitored by absorbance at 280nm (Figure 2.2). Analysis of the collected fractions by SDS polyacrylamide gel electrophoresis (SDS-PAGE) showed that the major peak, eluting at about 30min, contained mL A. This initial purification step was found to be highly effective in separating mL A from the inclusion body lysate and only small amounts of protein contaminants were present in the fractions containing mL A (Figure 2.3).

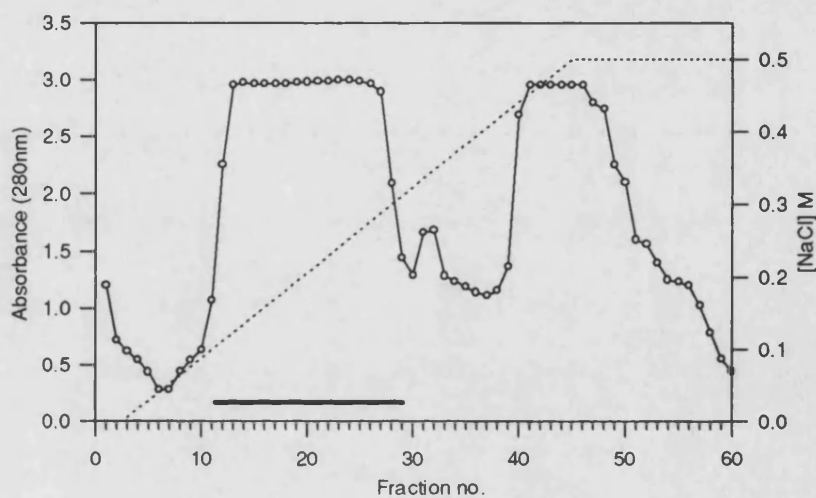


Figure 2.2 — **Elution of urea-solubilised mL A from the anion exchange matrix.** See text for details. The black bar represents the fractions (2.5ml) containing mL A. These were collected and folded to the native form using equilibrium dialysis. The continuous line corresponds to the absorbance at 280nm and the dotted line represents the concentration of NaCl used in the buffer B elution gradient.

### 2.3.2 Folding

Disulphide-containing proteins, such as LA, require an alkaline pH and the presence of thiol/disulphide compounds for (re-)folding and correct disulphide bond formation (Saxena and Wetlaufer, 1970; Creighton, 1980). In addition, LA is unusual in that it requires calcium in a 1:1 stoichiometry for efficient folding *in vitro* (Rao and Brew, 1989). Native mL A was generated by progressive dialysis steps against decreasing

urea concentrations in the presence of a redox folding buffer. The folding buffer (10X) contained 200mM Tris-HCl, 10mM CaCl<sub>2</sub>, 50mM β-mercaptoethanol and 5mM 2-hydroxyethylidisulphide with a pH adjusted to 9.0.

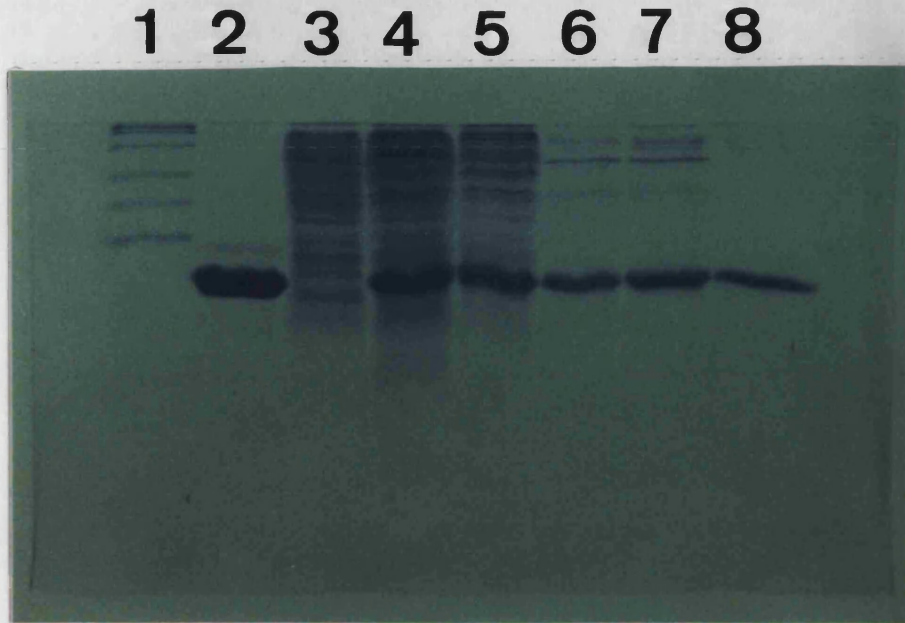


Figure 2.3 — **SDS-PAGE analysis of mLA at various stages during purification.**

The SDS-PAGE gel (20% polyacrylamide) above was run using standardised amounts of protein (1:20000) from various stages during the purification procedure. Lane 1, Molecular weight markers<sup>1</sup> (from bottom: lysozyme (18.5K), soybean trypsin inhibitor (27.5K), carbonic anhydrase (32.5K), ovalbumin (49.5K), bovine serum albumin (80K), phosphorylase *b* (106.5K)); Lane 2, 20µg commercially prepared bovine milk LA (Sigma); Lane 3, Uninduced expression cells; Lane 4, 3hrs post-induction; Lane 5, Urea solubilised inclusion bodies; Lane 6, Pooled Q50 purified fraction; Lane 7, Post dialysis refolding; Lane 8, Pooled G-75 purified fraction.

Fractions from the mono-Q separation containing expressed mLA were pooled and the protein concentration was adjusted to about 1mg/ml with folding buffer (1X) containing 2M urea. The protein concentration was estimated from the absorbance at 280nm assuming a value of 2.0 for  $E_{280}^{0.1\%}$  (Hill and Brew, 1975). For variant LAs, in

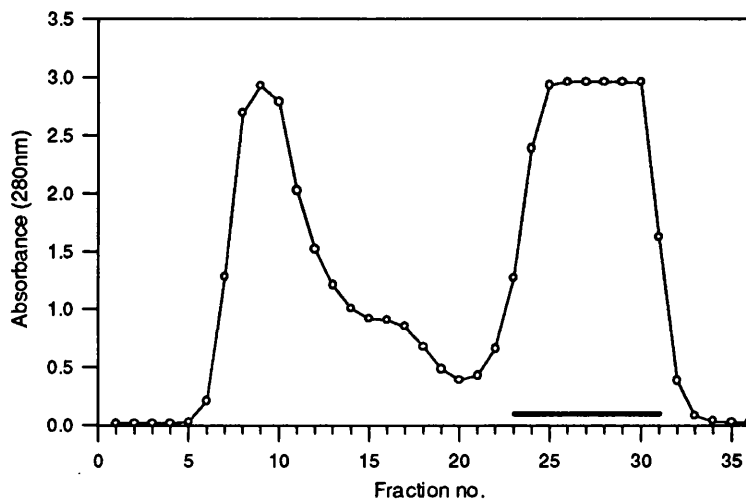
<sup>1</sup> In the case of the prestained SDS-PAGE standards (BioRad) used here, the covalently bound dye alters the molecular weight of the protein. The values given in parentheses therefore correspond to the apparent molecular weights of the markers. For example, lysozyme, which has a molecular weight similar to LA (~14.5KDa), runs with an apparent molecular weight of 18.5KDa in its prestained form.

which tyrosine or tryptophan residues had been inserted or removed, the  $E_{280}^{0.1\%}$  was calculated using the method described by Gill and von Hippel (1989). The diluted protein solution was then dialysed against folding buffer at 4°C. Dialysis was carried out for at least 6hrs against solutions of folding buffer containing 2M, 1M and 0M urea respectively. The resulting solution, containing refolded mLA, was then dialysed against 30mM ammonium bicarbonate and lyophilised. The urea (Sigma) stock solution was pre-filtered through a column containing a mixed-bed resin (AG501 X8(D), 20-50 mesh (Bio-Rad)) to remove cyanate and other contaminants.

### 2.3.3 Final Purification

mLA was finally purified by gel filtration with columns (80mm x 500mm) of Superdex G-75 equilibrated with 30mM ammonium bicarbonate containing 10 $\mu$ M CaCl<sub>2</sub>. The refolded mLA from a 6 litre preparation was typically dissolved in 10ml of 30mM ammonium bicarbonate and applied to the column. Elution was performed at room temperature (about 22°C) with 30mM ammonium bicarbonate containing 10 $\mu$ M CaCl<sub>2</sub> at a flow rate of 2.5ml/min. mLA eluted as a sharp single peak (Figure 2.4). The fractions containing mLA were pooled and lyophilised. The purity of the expressed protein was assessed by SDS-PAGE and analytical ion-exchange high performance liquid chromatography. HPLC separations were carried out with a Hewlett-Packard 1090 liquid chromatograph fitted with an ion-exchange column (see Appendix). Both bovine milk LA and mLA eluted after about 30 min.

The expression system was relatively efficient and final yields of mLA consistently exceeded 35mg/litre of bacterial culture. The final product ran as a single band on SDS-PAGE and had a mobility that was essentially identical to the bovine milk



**Figure 2.4 — Final purification of mLA by gel filtration.**

After treatment to generate native folding, the lyophilised material was dissolved in buffer and loaded onto a column containing Superdex G-75 (see text for details). The fractions containing mLA are marked by the black bar. This elution profile is typical of those obtained for the mutants that folded efficiently. Mutants that folded incorrectly exhibited large amounts of aggregated material in the void volume and an increased peak around fraction 16. SDS-PAGE shows that this latter peak (#16) contains LA but most likely represents an aggregated dimeric form of the protein.

protein (Figure 2.3). Amino-terminal sequencing of mLA produced a single sequence with methionine as the amino terminal residue followed by the expected amino terminal sequence of bovine LA. Sequencing was performed using an Applied Biosystems protein sequencer (model 470A) fitted with a phenylthiohydantoin-derivative analyser (model 120A) and a sequence assignment data analysis system (model 900A).

## 2.4 Generation and Expression of Mutant mLAs

### 2.4.1 Mutant Design

The mutations introduced into mLAs were directed to three separate regions of the structure (Table 2.1). The first region to be investigated comprised the invariant positions in aromatic cluster I (Phe-31, His-32, Gln-117 and Trp-118). These residues

Table 2.1— Primers used for the PCR mutagenesis.

Primer	Sequence	Orientation
F31E	5'-CTGTACCACG <b>GAA</b> CATACCAGTGG-3'	Coding
F31L	5'-CTGTACCACG <b>TTA</b> CATACCAGTGG-3'	Coding
F31S	5'-CTGTACCACG <b>TCT</b> CATACCAGTGG-3'	Coding
F31Y	5'-CTGTACCACG <b>TAT</b> CATACCAGTGG-3'	Coding
H32A	5'-ACCACGTTT <b>GCA</b> ACCAGTGG-3'	Coding
H32E	5'-ACCACGTTT <b>GAA</b> ACCAGTGG-3'	Coding
H32N	5'-ACCACGTTT <b>AAT</b> ACCAGTGG-3'	Coding
H32Y	5'-ACCACGTTT <b>TAT</b> ACCAGTGG-3'	Coding
Y103A	5'-GGCCAACCAG <b>GGC</b> GTTAATTCC-3'	Complementary
Y103P	5'-GGCCAACCAG <b>GGG</b> GTTAATTCC-3'	Complementary
W104A	5'-GCTTTATGGG <b>CGC</b> AAAGTAGTGAATTCC-3'	Complementary
W104L	5'-GCTTTATGGG <b>CAA</b> CAAGTAGTGAATTCC-3'	Complementary
W104Y	5'-GCTTTATGGG <b>AGT</b> AGTAGTGAATTCC-3'	Complementary
H107A	5'-CAGAACAGAGTGCTTT <b>AGC</b> GGCCAACCAG-3'	Complementary
H107Y	5'-CAGAACAGAGTGCTTT <b>ATA</b> GGCCAACCAG-3'	Complementary
H107W	5'-CAGAACAGAGTGCTTT <b>CCA</b> GGCCAACCAG-3'	Complementary
A109P	5'-CAGAACAGAG <b>TGG</b> TTTATGGGCC-3'	Complementary
L110E	5'-GCTTCTCAGAACA <b>CTC</b> TGCTTTATGGGC-3'	Complementary
L110H	5'-GCTTCTCAGAACA <b>ATG</b> TGCTTTATGGGC-3'	Complementary
L110R	5'-GCTTCTCAGAACA <b>CCG</b> TGCTTTATGGGC-3'	Complementary
Q117A	5'-GAGAAGCTGGAT <b>GCT</b> TGGCTCTG-3'	Coding
Q117D	5'-GAGAAGCTGGAT <b>GAT</b> TGGCTCTG-3'	Coding
Q117R	5'-GAGAAGCTGGAT <b>CGT</b> TGGCTCTG-3'	Coding
Q117S	5'-GAGAAGCTGGAT <b>AGT</b> TGGCTCTG-3'	Coding
W118H	5'-GGATCAG <b>CAI</b> TCTCTGTGAG-3'	Coding
W118Y	5'-GGATCAG <b>TAT</b> TCTCTGTGAG-3'	Coding

are conserved only in LA sequences and, as discussed in Chapter 1, have been implicated in LA function by a number of different studies. The choice of substitutions at these sites was based primarily on the nature of the residues found at the corresponding positions in lysozymes although additional changes were designed to examine the effects of different



sidechain types at a particular position. For example, His-32 was changed to an Asn and Tyr (as in several lysozymes), a Glu (to investigate the effect of introducing a charged group into the cluster) and an Ala (essentially removing the sidechain contribution).

The second group of mutations were directed at the 'hydrophobic-box' residues (aromatic cluster II) Tyr-103 and Trp-104. Other residues in cluster II (positions 26, 53, 60, 95) were not altered on account of their completely buried environments in the three-dimensional structures of both LA and lysozyme. The sidechain of Tyr-103 is located in the cleft region and results in the blockage of the A and B subsites that bind saccharide in lysozyme (Warne *et al.*, 1974; Acharya *et al.*, 1989). This tyrosine is present in all LAs sequenced to date apart from rabbit (His) and platypus (Ala). The corresponding position in the lysozymes (residue 107) is occupied by either an alanine or a proline (see Figure 1.1). Tyr-103 was replaced with either an Ala or Pro to investigate their influence on LA function. Changes at Trp-104, which is totally conserved in both LAs and lysozymes, were designed to ascertain the effect of decreased hydrophobic sidechain size at this position (W→Y, L, A).

The third set of mutations were made in a region adjacent to aromatic cluster I comprising residues 105-110 (hereafter referred to as 'the flexible loop region'). There is some evidence to suggest that this region is likely to be in close proximity to GT in the LS complex and it may even be an important interaction site. Studies have shown that Lys-108 is the major site of cross-linking in the LS complex (Brew *et al.*, 1975). Moreover, the conformation of this region is very flexible in the three-dimensional structure of HLA and adopts two alternate conformations depending on the conditions used to obtain crystals (Harata and Muraki, 1992). Substitutions at His-107 (H→A, Y,

W) were designed to favour the helical conformation, that is observed in HLA crystals grown at physiological conditions, in which the sidechain of residue 107 is buried. Tyr is found at this position in mouse LA and a Trp is present at this site in both platypus LA and all known lysozymes. Ala-109 was changed to a proline which is found at this position in several LAs. This substitution was based on the tenet that a proline would be incompatible with the observed helical conformation and such a change should favour the loop conformation seen in HLA crystals grown at pH 4.2. The final set of substitutions in this region were made at position 110 (L→E, H, R). These changes were designed to investigate the effects of various sidechain types. Much the same variability in sidechain character is seen in the corresponding position in lysozymes (Arg in human and chick LYZ and a His in several LYZs including bovine, trout and hoatzin).

### 2.4.2 PCR Mutagenesis

The single-site substitutions were introduced into p-MLA using the PCR Megaprimer method (Sarkar and Sommer, 1990). This site-directed mutagenesis method utilises three oligonucleotide primers to perform two separate PCRs. The product of the first reaction is subsequently used as a 'megaprimer' in the second amplification. This procedure is simpler than four primer/three PCR single site-directed mutagenesis method (Higuchi *et al.*, 1988) and is also less likely to introduce unwanted mutations as fewer PCR cycles are required. p-MLA was used as a template in all the amplifications. Depending on the location of the target amino acid, the first PCR was performed with either the T7 promoter or T7 terminator primer (Novagen) in conjunction with the mutagenic primer (see Table 2.1 for the sequences and orientation of the mutagenic primers). The product of the first PCR was directly isolated from the amplification

reaction using the Magic™ PCR Preps Purification Kit (Promega Co., Madison, WI). The isolated megaprimer was then used in the second amplification with the same template (p-MLA) and the cognate T7 primer to generate the full length product. The final PCR product was isolated, purified, digested and cloned into cut pET3a as described for mL A (see Section 2.2.2). Mutant expression vectors were characterised by restriction mapping and automated DNA sequencing.

All the sequence changes, except for those at positions 117 and 118, were made with relative ease using the above protocol. In four cases, an additional single base change was erroneously introduced by the mutagenesis procedure resulting in an extra, unwanted amino acid substitution. These were F31E-mLA, where a methionine was substituted for valine 8, and all the L110 variants where a lysine was substituted for leucine 115 in all cases (henceforth double mutants are referred to as F31E\* and L110\*). Changes at positions 117 and 118 proved particularly difficult to engineer using the primers that had been designed. On numerous occasions transformants were obtained but subsequent DNA sequencing and activity assays of the expression products revealed that the PCR mutagenesis had failed as the DNA sequences and protein activities were identical to the parent mL A. This was thought to be primarily due to the primer design. The 117 and 118 mutagenic primers, initially designed specifically for the fusion protein system, were orientated with respect to the coding strand of LA. As a result, the products of the first PCR reaction were relatively small (<80bp) and could not be isolated using the Magic™ PCR Preps Purification Kit as the reported recovery rate for this size of fragment is very low (<5%). These mutants were subsequently constructed successfully using appropriate modifications to the standard mutagenesis protocol. The

small megaprimers were separated by electrophoresis on a 12% polyacrylamide gel, isolated by Maniatis' 'crush-and-soak' method (Sambrook *et al.*, 1989) and used in the second PCR reaction to generate the full length product containing the desired mutation (J. Grobler, personal communication).

### 2.4.3 Expression, Folding and Purification

The variant LA vectors were transformed into BL21 DE3s and the mutant LAs were expressed and purified as described for mLA (see Section 2.3). Folding of the variant LAs was performed in a similar fashion to that used for mLA. In two cases (H32A and Y103P), 10% (v/v) glycerol was included in the folding buffer to try to increase the folding efficiency. Table 2.2 summarises the yields obtained at various stages during the purification procedure. Although initial levels of expressed protein for the mutant LAs were reasonably constant, the final yields varied considerably and the majority of the mutant LAs were recovered in much smaller amounts compared to mLA. SDS-PAGE had previously shown that mLA was the major protein component (>95%) after the initial anion exchange separation in urea with only low levels of protein contaminants (Figure 2.3). Therefore, as initial levels of expressed protein were fairly constant, the decreased yields appear to be due entirely to a reduction in the variant's folding efficiency *in vitro*. No specific attempts were made to alter the folding conditions for the mutants obtained in low yields, except for cases where 10% glycerol was added to the folding mixture. Particular substitutions appear to be incompatible with native folding, under the conditions employed, based on the minimal yields obtained: F31E\*, F31L, H32E, H32N, Y103A, W104A, W104L and W104Y. This inability to fold correctly was characterised by large amounts of aggregated protein eluting in the void

volume of the gel filtration separation. In contrast, the mutants that folded relatively efficiently (F31Y, H32Y, Y103P, H107's, A109P, L110\*'s) behaved in a similar fashion to mLA during gel filtration.

Table 2.2 — Yields of mLA and mutants at various stages during purification.

Protein	Total Protein post MacroPrep Q50	Total Protein post G75	Yield <sup>a</sup>
	<i>mg/litre of bacterial culture</i>		<i>%</i>
mLA	107	40	100
F31E*	78	8.8 (0) <sup>b</sup>	0 <sup>c</sup>
F31L	63	5.5 (0)	0 <sup>c</sup>
F31S	60	16.5 (1.0)	4 <sup>c</sup>
F31Y	71	25.5	95
H32A	55	10.0 (1.0)	5 <sup>c</sup>
H32A <sup>d</sup>	73	28.5 (5.7)	21 <sup>c</sup>
H32E	66	nd <sup>e</sup> (0)	0
H32N	58	nd (0)	0
H32Y	47	17.5	100
Y103P <sup>d</sup>	53	17.7	29
W104A	53	nd (0)	0
W104L	43	nd (0)	0
W104Y	59	nd (0)	0
H107A	79	18.1	61
H107W	94	16.3	46
H107Y	67	25.5	100
A109P	58	16.0	73
L110E*	63	4.7	20
L110H*	35	7.5	34
L110R*	34	7.2	59

<sup>a</sup> Yield for mLA after gel filtration was designated 100%. The yields for the mutants were calculated relative to mLA

<sup>b</sup> Values in parentheses are yields after ion exchange HPLC

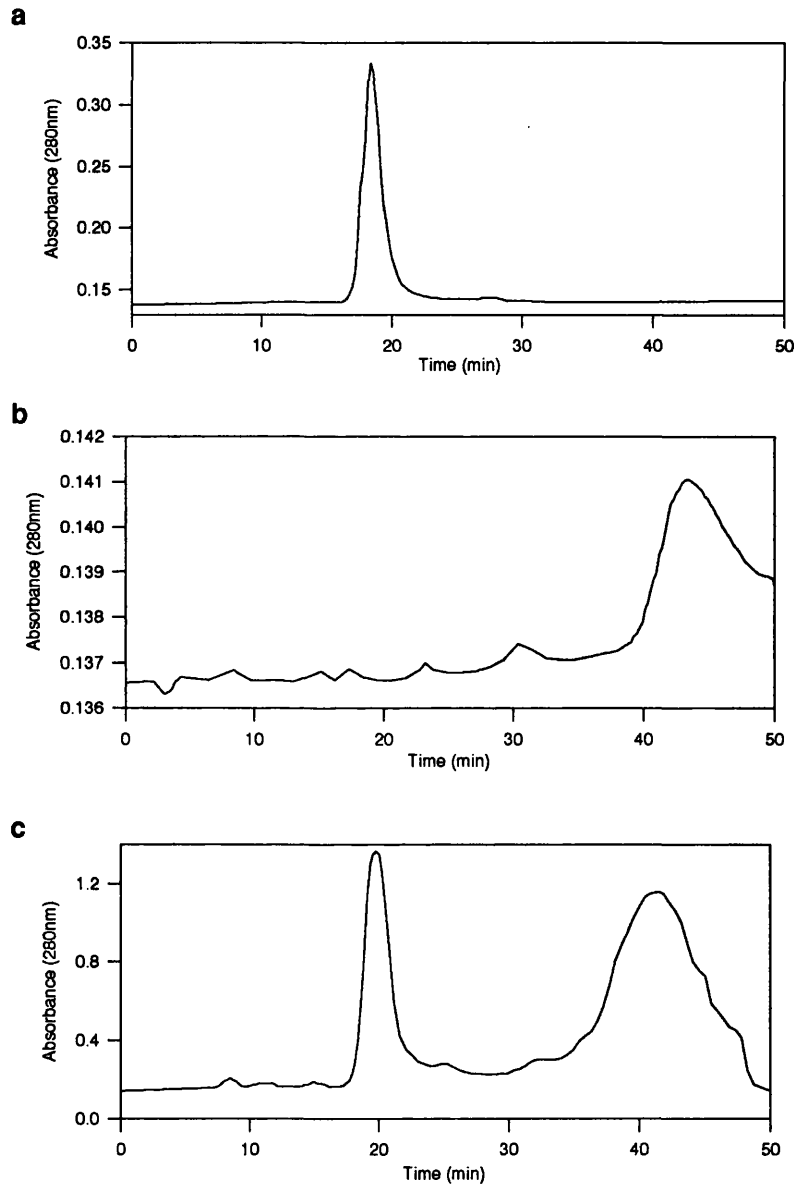
<sup>c</sup> Yields calculated after HPLC separation

<sup>d</sup> Folded in the presence of 10%(v/v) glycerol

<sup>e</sup> nd, no gel filtration step used. Mutants purified directly by HPLC ion exchange

mLA and the correctly-folded mutants eluted as single peaks on HPLC analysis at a similar NaCl concentration to that characteristic of bovine milk LA (Figure 2.5a). Other mutants had more complex HPLC elution profiles. Mutant mLAs with non-native folds, such as F31E\* and F31L, showed no peak eluting at the same time as mLA but a major

peak eluting at high NaCl concentration (Figure 2.5b). This is presumably due to the aggregation of the incorrectly folded protein. A second group of mutants, with lower folding efficiency (F31S and H32A), consisted of both correctly and incorrectly folded



**Figure 2.5 — Separation by HPLC anion exchange chromatography.**  
a. mLA; b. F31L-mLA; c. F31S-mLA. See text for details.

forms of the protein as HPLC analysis showed both the native and aggregated peaks (Figure 2.5c). F31S and H32A-mLA were subsequently purified exclusively by HPLC ion exchange to ensure the homogeneity of the final material. Sufficient amounts of protein were obtained for mLA, F31S, F31Y, H32A, H32Y, Y103P, H107A, H107W, H107Y, A109P, L110H\* and L110R\*. All the samples were judged to be homogeneous and of high purity (>99%) by SDS-PAGE and HPLC analysis.

### 2.4.4 Summary

The T7 polymerase expression system described above was found to be extremely powerful and allowed high yields of native and mutant mLAs to be obtained with relative ease. This system represents a significant improvement over those that have been previously described for LA (see Section 2.1). The Megaprimer method used to construct the site-directed mutants introduced several unintended mutations but this was a small sacrifice when balanced against the facility of the procedure. The problem appears to be associated with *Taq* polymerase's 3' terminal transferase activity which results in the addition of an extra base at the mutated end of the megaprimer. This effect could be overcome by taking extra precautions when designing primers so that any additional base added by the transferase activity can be accommodated without generating a protein sequence change. Alternatively, improved efficacy could be obtained with the use of a less error-prone polymerase.

Several conclusions can be made regarding the effect of various substitutions on the folding efficiency of recombinant LA. In general, replacement of aromatic sidechains in the clusters I and II with non-aromatics resulted in low final yields. This is presumably

due to either a destabilisation of the native conformation or to a more general effect of a particular substitution on the *in vitro* folding kinetics of the variant LAs. This observation is highlighted by the changes at positions 31 and 32 where only Tyr-substituted variants (F31Y and H32Y) were obtained in respectable final yields. In the case of H32A, addition of 10% glycerol to the folding mixture resulted in a 5-fold increase in the final yield of folded protein as reported for marginally stable mutants of lysozyme (Sawano *et al.*, 1992). Virtually all changes in aromatic cluster II tended to seriously affect the folding of LA. Although substitution of Tyr-103 with a proline resulted in reasonable final yields, two of the three substitutions made at position 104 (alanine and leucine) did not fold correctly. Some monomeric W104Y-mLA was isolated by gel filtration but preliminary characterisation using CD measurements suggest that this variant may have significant alterations in the conformation of its hydrophobic box (K. Brew, personal communication). In contrast, changes at positions 107, 109 and 110\* had minimal effects on LA's folding efficiency.

## 2.5 Kinetic Characterisation of mLA and the mLA Mutants

### 2.5.1 Kinetic Mechanism of GT and the Lactose Synthase Complex

The role of LA in the lactose synthase (LS) complex has already been briefly described in Chapter 1. In the complex, interaction of LA with GT modulates the affinity of GT for acceptor substrates, enhancing the binding of glucose and other monosaccharides while acting as a competitive inhibitor with respect to extended oligosaccharide substrates, such as N, N' diacetylchitobiose (ChB), and glycoproteins (Khatra *et al.*, 1974; Powell and Brew, 1976b; Bell *et al.*, 1976). Substrate inhibition is



also observed with glucose, and other weaker binding monosaccharides, at high concentrations of LA and/or acceptor due to the formation of dead-end inhibitory complexes (Morrison and Ebner, 1971b). The main consequence of this affinity modulation is to increase GT's affinity for glucose 1000-fold, thereby allowing the efficient synthesis of lactose at physiological concentrations of glucose.

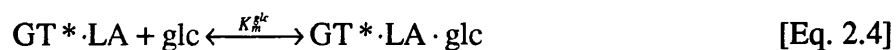
To date, no universal mechanism has been proposed that can account for all the steady-state kinetic data. Three alternative schemes, representing the mechanism of lactose synthase, have been suggested. The major differences between these schemes concern the point at which LA becomes involved and the level of randomness in the reaction. In the earliest studies, Morrison and Ebner suggested that the substrates of the transferase reaction add in an ordered manner with  $Mn^{2+}$  as the first substrate followed by the sequential addition of UDP-galactose, acceptor and finally LA (Morrison and Ebner, 1971a,c). A series of subsequent studies suggested that the reaction was best represented by a partially ordered mechanism, in which UDP-galactose binds to a  $GT \cdot Mn^{2+}$  complex prior to the random equilibrium binding of LA and glucose (Khatra *et al.*, 1974; Powell and Brew, 1976b). A third scheme proposes that lactose synthesis proceeds by a completely random equilibrium addition of substrates (UDP-galactose, LA and acceptors) to a  $GT \cdot Mn^{2+}$  complex (Bell *et al.*, 1976). In spite of the inherent differences and kinetic implications of the three schemes, there is a more general agreement that the modulatory effects of LA involve a random, but highly synergistic, binding with glucose to GT complexes. All three mechanisms propose that LA attaches to GT prior to the release of products and dissociates from the complex at the end of

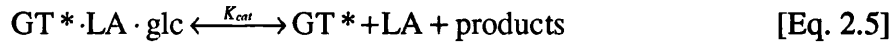
each catalytic cycle. In the case of larger acceptor substrates, such as ChB and ovalbumin, the binding of LA and acceptor is mutually exclusive.

## 2.5.2 Interpretation of Steady State Kinetic Measurements of LA Activity

The activity of mLA and the mutant mLAs in promoting glucose binding to bovine GT was determined by a radiochemical assay as in previous studies (Brew *et al.*, 1968; Khatra *et al.*, 1974). To simplify interpretation of the kinetic data from mLA and the mutants, the conditions used for the activity assays were designed to essentially saturate GT with  $Mn^{2+}$  and UDP-galactose. Therefore, regardless of which mechanism provides the best representation of the LS system, all the GT will be distributed in complexes containing both  $Mn^{2+}$  and UDP-galactose. The GT assay used here is described in full in the Appendix. Manganese ( $Mn^{2+}$ ), in the form of  $MnCl_2$ , was used at a concentration of 10mM; this compares to dissociation constants of  $Mn^{2+}$  from the two non-equivalent metal-binding sites on GT of 2 $\mu$ M and 2mM (Powell and Brew, 1976a). UDP-galactose was present at a concentration of 330 $\mu$ M (compared to  $K_m$  and  $K_d$  values of 60 and 25 $\mu$ M).

As GT is saturated with  $Mn^{2+}$  and UDP-galactose, the relevant part of the kinetic mechanism can be represented by the following equilibria:





where  $\text{GT}^*$  represents the  $\text{GT} \cdot (\text{Mn}^{2+})_2 \cdot \text{UDP-galactose}$  complex.

A low concentration of glucose was used in the assay (10mM compared to its  $K_m$  of 2M), so that enzyme complexes containing glucose but not LA (Eq. 2.2 and 2.3) were insignificant. In this case, if rapid equilibrium binding of LA and glucose is assumed (Khatra *et al.*, 1974; Bell *et al.*, 1976), the rate equation reduces to:

$$v = \frac{V_{max} [\text{LA}] [\text{glc}]}{K_i^{LA} K_m^{glc} + K_m^{glc} [\text{LA}] + [\text{LA}] [\text{glc}]} \quad [\text{Eq. 2.6}]$$

where  $V_{max}$  is the maximum velocity ( $k_{cat}[\text{GT}]$ ),  $[\text{LA}]$  and  $[\text{glc}]$  are the concentrations of LA and glucose,  $K_i^{LA}$  is the dissociation constant for LA from the  $\text{GT} \cdot \text{Mn}^{2+} \cdot \text{UDP-galactose} \cdot \text{LA}$  complex, and  $K_m^{glc}$  is the  $K_m$  for glucose at saturating levels of LA. In a rapid equilibrium mechanism, the  $K_m^{glc}$  is equal to the equilibrium dissociation constant of glucose from the  $\text{GT} \cdot \text{Mn}^{2+} \cdot \text{UDP-galactose} \cdot \text{LA} \cdot \text{glucose}$  complex.  $K_i^{LA}$  is the dissociation constant of LA from the same complex that produces competitive inhibition with respect to extended substrates (e.g. ChB), and can therefore be determined separately as the inhibition constant ( $K_i$ ) for that reaction.

The action of LA in activating the catalysis of lactose synthesis by GT, at a fixed concentration of glucose, reflects an increase in the enzyme complexes which can give rise to products (i.e.  $\text{GT}^* \cdot \text{LA} \cdot \text{glc}$ ). This is dependent on both the affinity of LA (and the mutants) for GT ( $K_i^{LA}$ ) and its effects on glucose binding ( $K_m^{glc}$ ). A secondary, double reciprocal plot of Eq. 2.6 (LS activity *versus*  $[\text{LA}]$ ) will have an intercept/slope ratio ( $1/K_m^{\text{app}}$ ) of:

$$\frac{\left(\frac{[\text{glc}]}{K_m^{\text{glc}}} + 1\right)}{K_i^{\text{LA}}} \quad [\text{Eq. 2.7}]$$

from which  $K_m^{\text{glc}}$  can be determined using the value of  $K_i^{\text{LA}}$  obtained from the inhibition of galactosyl transfer to ChB. Alternatively, the apparent  $V_{\text{max}}^2$  for lactose synthesis can be used to determine  $K_m^{\text{glc}}$  by comparison with the parameters obtained for a corresponding plot for bovine milk LA under the same conditions. The affinity of LA for the  $\text{GT}\cdot\text{Mn}^{2+}\cdot\text{UDP-galactose}$  complex can be measured independently from glucose binding by using LA as a competitive inhibitor of galactosyl transfer to extended substrates. For the ChB reaction (see Appendix), apparent  $K_i$  values can be obtained from the intercept/slope ratios of Dixon plots ( $1/v$  versus [LA]). The true  $K_i$  values can then be calculated using the relationship:

$$K_i(\text{true}) = \frac{K_i(\text{apparent})}{\left(1 + \frac{[\text{S}]}{K_m}\right)} \quad [\text{Eq. 2.8}]$$

where  $K_m$  is the  $K_m$  for ChB<sup>3</sup> and [S] is the concentration of ChB (0.5mM). This inhibition constant ( $K_i^{\text{LA}}$ ) can subsequently be used in Eq. 2.7 to calculate the  $K_m$  for glucose at saturating concentrations of LA.

### 2.5.3 Functional Properties of mLA and the Mutants

The effects of the amino acid substitutions were characterised by using the parent and mutant mLAs as activators of lactose synthesis and as inhibitors of galactosyl

---

<sup>2</sup>  $V_{\text{max}}(\text{apparent}) = \frac{V_{\text{max}}(\text{true})}{\left(1 + \frac{K_m}{[\text{glc}]}\right)}$

<sup>3</sup> separately determined to be 1.14mM under these conditions (see Appendix)

transfer to ChB. mLA is kinetically very similar to the bovine milk protein (BOLA). This is not particularly surprising as mLA only differs from the BOLA by the presence of an additional N-terminal methionine residue and a conservative internal methionine to valine change (M90V). Due to the way in which mLA was constructed, the additional amino terminal methionine residue, introduced by the cloning procedure, could be removed by CNBr cleavage to yield a product that differed from the milk protein at only one position. Nevertheless, it was deemed appropriate to investigate the structure-function relationships in LA without removing the NH<sub>2</sub>-terminal methionyl residues of the mLA mutants because of the similarity of mLA to the wild type protein in terms of kinetic and physical properties. However, only protein preparations that were 'native-like' based on chromatographic behaviour, homogeneity on HPLC ion exchange chromatography and near and far UV CD spectra were used for studying protein structure-function relationships. This was to ensure that any functional differences observed with the variant LAs only reflected local structural alterations.

The kinetic activities of mLA, Y103P-mLA and A109P-mLA are similar whereas all the other mutants are considerably less active than the parent protein (Table 2.3). Figure 2.6 shows the activities of mutants with substitutions for Phe-31 and His-32 as activators of lactose synthesis and inhibitors of galactose transfer to ChB. Half the substitutions at positions 31 and 32 are incompatible with the folding of LA under the conditions used here to generate the native protein. Inspection of the activity profiles for those variants of positions 31 and 32 that could be folded (F31S, F31Y, H32A, H32Y) show reduced lactose synthase activity even at relatively high concentrations of LA. As discussed in Section 2.5.2, changes in the apparent  $V_{max}$  of the lactose synthase reaction

Table 2.3 — Kinetic parameters determined for mLA and mutants with bovine GT.

Protein	$K_i^{LA}$	Change in $K_i^a$	$K_m^{glc}$	Change in $K_m^b$
	$\mu\text{M}$		$\text{mM}$	
Bovine LA <sup>c</sup>	17	1	0.9	0.9
	12.3	1	0.9	0.6
mLA	17±2	1	1.6±0.4	1
F31S	655	39	157 <sup>d</sup>	98
F31Y	38	2.2	10.5 <sup>d</sup>	7
H32A	509	30	nd <sup>e</sup>	nd
H32Y	201	12	213 <sup>d</sup>	133
Y103P	15	0.9	2.0	1.3
H107A	168	10	1.2 <sup>d</sup>	0.8
H107W	40.5	2.4	1.0 <sup>d</sup>	0.6
H107Y	138	8	0.7 <sup>d</sup>	0.4
A109P	15±3	0.9	1.7	1.1
L110E*	nd	nd	nd	nd
L110H*	632 (650) <sup>f</sup>	37 (38) <sup>f</sup>	2.7 (2.8) <sup>f</sup>	1.7 (1.7) <sup>f</sup>
L110R*	284 (168) <sup>f</sup>	17 (10) <sup>f</sup>	30.4 <sup>d</sup> (50) <sup>f</sup>	19 (31) <sup>f</sup>
Q117A	135	8	3.2	2
Q117D	57	3.4	2.1	1.3
W118H	1766	104	nd	nd
W118Y	1488	88	1.7	1.1

<sup>a</sup> Change in  $K_i$  is the value for  $K_i^{LA}$  divided by the corresponding value for mLA

<sup>b</sup> Change in  $K_m^{glc}$  divided by corresponding value for mLA

<sup>c</sup> Powell and Brew, 1974

<sup>d</sup>  $K_m^{glc}$  determined using  $V_{max}^{app}$

<sup>e</sup> nd, parameter could not be reliably determined from data

<sup>f</sup> Values in parentheses correspond to kinetic parameters for reconstructed 110 mutants.

reflect effects on LA's ability to promote glucose binding. This observation is best illustrated by the variant F31Y-mLA. It has a much lower ability to promote glucose binding (reflected by a 7-fold increase in  $K_m^{glc}$ ) but only a slightly worse affinity for GT when compared to mLA (2-fold). This general effect on LA's ability to promote glucose binding is also observed in the variants F31S, H32A and H32Y but is combined with major reductions in GT affinity (Table 2.3). Despite the low levels of activity observed for the variants at position 32, it appears that these mutants are even more deficient than the position 31 mutants in their effects on glucose binding. Encouragingly, these kinetic

results are consistent with previous mutagenesis studies carried out in the fusion protein system (Brew *et al.*, 1993).

As described in the previous section, alterations at the hydrophobic box residues 103 and 104 appear to be incompatible with the native folding of LA and only one of the five mutants (Y103P-mLA) allowed generation of the native protein. This mutation does not appear to significantly affect the function of LA (Table 2.3). Although Y103P-mLA had a higher affinity for GT, compared to mLA, its ability to promote glucose binding was slightly reduced and, consequently, it had a similar overall activity to mLA in the LS reaction (data not shown). The consequence of this substitution, in contrast to the Y103A-mLA which did not fold correctly, was initially of interest as both proline and alanine are observed in the corresponding position in the lysozymes. Furthermore, an alanine occupies this position in LA from platypus milk. However, automated sequencing of the Y103A-mLA mutant indicated that a single base deletion had occurred in the expression vector. Subsequent reconstruction of this mutant, using a different primer, yielded a 'native' product that had a greatly reduced activity in both kinetic assays as well as an altered CD spectra (K. Brew, personal communication). Therefore, it appears that the varying functional effects observed with the position 103 mutants may arise from differential effects on the stability of the native protein due to a destabilisation of the hydrophobic box region. This conclusion is consistent with recent data on the stability of mLA variants which demonstrates that Y103P-mLA has essentially the same stability as mLA. In contrast, the stability of other variants at positions 103 and 104 are considerably reduced (up to 2.6kcal/mol lower than mLA; K. Brew, personal communication). In

summary, the results from the substitutions in aromatic cluster II suggest that this region has a predominately structural role in LA.

In contrast to the substitutions at Phe-31 and His-32, changes in the flexible loop region (105-110) have major effects on LA's  $K_i^{LA}$  (Figures 2.7 and 2.8). Changes at His-107 have negligible effects on LA's ability to promote glucose binding but significant effects on affinity (Table 2.3). These variants were originally designed to restrict the local flexibility in this region by favouring the helical conformation where this residue is buried. Surprisingly, the alanine and tyrosine substitutions have the largest effect on LA's affinity for GT whereas substitution with the bulkier tryptophan sidechain has a minimal effect. It is difficult to reconcile these observations with the conformational flexibility of the 105-110 loop region. Hopefully crystallographic studies on the His-107 variants will demonstrate the structural consequences of these substitutions and facilitate interpretation of the kinetic results.

Substitutions at Leu-110 have more equivocal effects on LA function. Although the L110E\*-mLA can be refolded to a form similar to the wild type (e.g. similar HPLC profile and thermal stability), it has no detectable affinity for GT, even at high concentrations, and thus it was impossible to determine the effect on glucose binding (Figure 2.8). The changes to arginine and histidine were more informative. Both



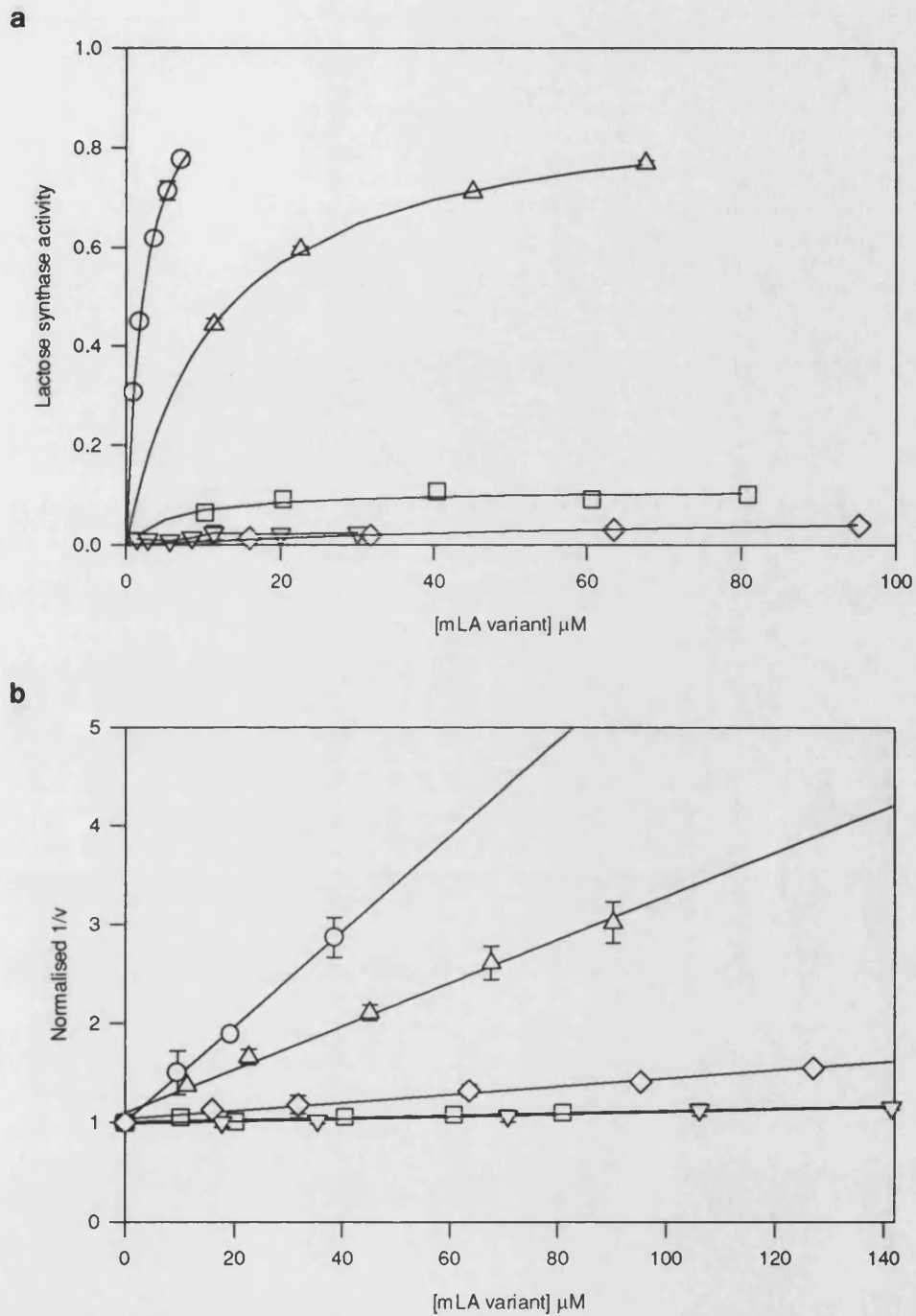


Figure 2.6 — The effects of substitutions for Phe-31 and His-32 on LA activity.

(a) activation of lactose synthase; (b) inhibition of galactose transfer to chitobiose. The results for the lactose synthase activity assay have been normalised against that of mLA. Data were fitted to the appropriate equations using the curve fitter function of SigmaPlot for Windows™. O, mLA; □, F31S; Δ, F31Y; ∇, H32A; ◇, H32Y.

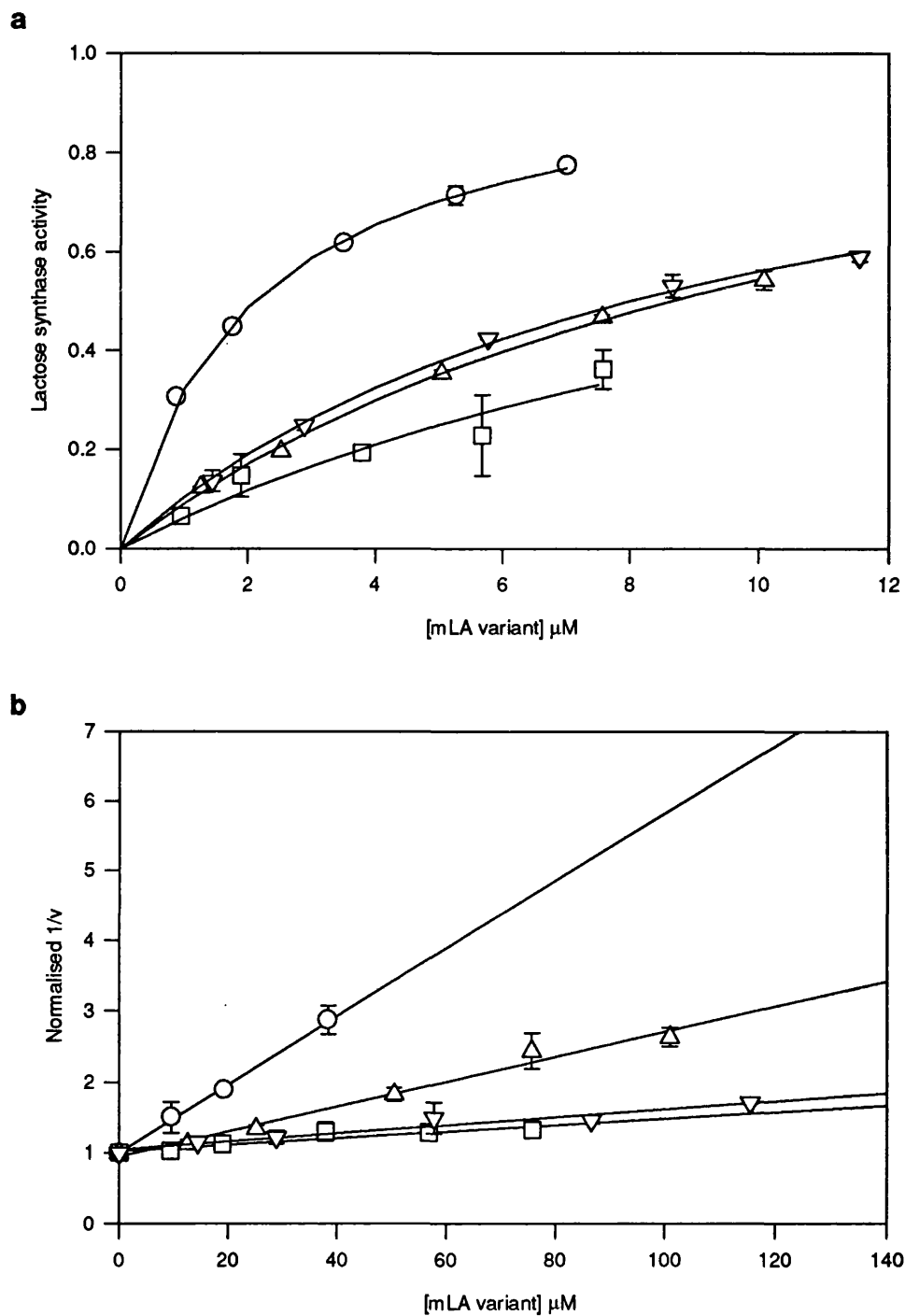


Figure 2.7 — The effects of substitutions for His-107 on LA activity.

(a) activation of lactose synthase; (b) inhibition of galactose transfer to chitobiose. The results for the lactose synthase activity assay have been normalised against that of mLA. Data were fitted to the appropriate equations using the curve fitter function of SigmaPlot for Windows™. ○, mLA; □, H107A, △, H107W; ▽, H107Y.

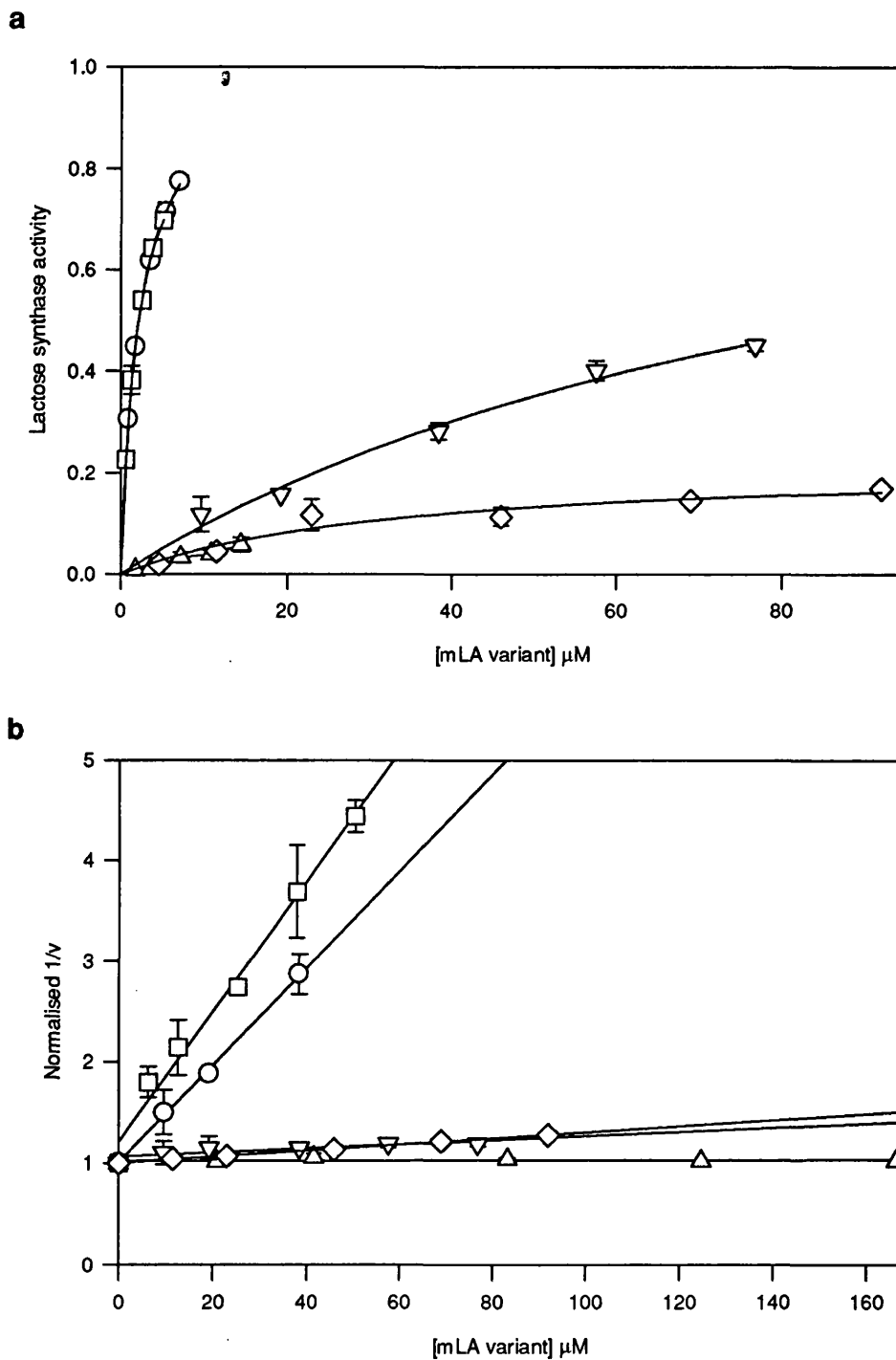


Figure 2.8 — The effects of substitutions for Ala-109 and Leu-110 on LA activity.

(a) activation of lactose synthase; (b) inhibition of galactose transfer to chitobiose. The results for the lactose synthase activity assay have been normalised against that of mLA. Data were fitted to the appropriate equations using the curve fitter function of SigmaPlot for Windows™. ○, mLA; □, A109P, △, L110E\*; ▽, L110H\*; ◇, L110R\*.

L110H\*-mLA and L110R\*-mLA have decreased affinities for GT (37-fold and 17-fold respectively) but, whereas the change to histidine has a minimal effect on LA's ability to promote glucose binding, L110R\* is about 20-fold worse than mLA with respect to glucose binding. This is visually apparent in Figure 2.8, where L110H\* is far more active in promoting lactose synthesis than L110R\* (Figure 2.8a), even though both L110H\* and L110R\* are poor inhibitors of galactosyl transfer to ChB (Figure 2.8b).

Interpretation of the results from changes made at position 110 is complicated by the additional mutation introduced during PCR mutagenesis. Sequencing of these mutants indicated that the conserved leucine at position 115 had been replaced by a lysine in all cases. These mutants have subsequently been reengineered so that they only contain the change at position 110. Characterisation of their kinetic behaviour shows that L110H-mLA and L110R-mLA behave in essentially the same way as the double mutants (K. Brew, personal communication; see Table 2.3). In contrast to the substitutions at position 110, the single change investigated at Ala-109 (A109P) had a negligible effect on either facet of LA function. Changes at positions 117 and 118 also had significant effects on LA's ability to bind to GT rather than on the promotion of glucose binding. These effects are far more pronounced in the position 118 variants and, in the case of W118H-mLA, the activity was so low that its influence on glucose binding could not be reliably determined from the data (J. Grobler, personal communication).

### 2.5.4 Significance of Kinetic Parameters and Assessment of Errors

The validity of the values obtained for the  $K_i^{LA}$  and  $K_m^{glc}$  for the various mutants is supported by the close agreement between the values obtained for the bovine milk protein and mLA with those obtained for bovine LA in more detailed kinetic studies

(Table 2.3). However, the low affinity for GT exhibited by several of the mutants could introduce significant errors into the calculated  $K_i^{LA}$  values. This potential error is compounded by limitations in the concentration range over which it was feasible to perform the inhibition studies due to the small quantities of folded protein obtained for some of the mutants. As the  $K_m^{glc}$  values are calculated using the  $K_i^{LA}$  values from the inhibition assays, any errors in the latter parameter will also be reflected in the former. It must therefore be emphasised that the mutant's kinetic parameters can only be used as a rough qualitative guide of the importance of a particular sidechain in LA function. Furthermore, the proximity of the sites on the surface of LA that interact with GT and glucose introduce a further convolution into any attempt to map them by site-directed mutagenesis. Any alteration in a component of one site might influence the functional properties of the adjacent site via steric, electrostatic or localised conformational effects.

However, as the values of  $K_m^{glc}$  and  $K_i^{LA}$  defined in Section 2.5 represent equilibrium dissociation constants (Khatra *et al.*, 1974; Bell *et al.*, 1976), they can be used to calculate Gibbs free energy changes (Table 2.4). In general, such  $\delta\Delta G$  values ( $\Delta G_{mLA} - \Delta G_{mutant} = -RT \ln \{K_{mLA}/K_{mutant}\}$ ), give the most reliable measurements of incremental changes in the respective binding energies that result from alterations in molecular contacts as a consequence of a particular sidechain substitution. An additional benefit of using such values is that errors in the calculated kinetic parameters become less significant. For example a 25% error in a 10-fold change in  $K_i^{LA}$  represents an error of about  $\pm 0.74$  kJ/mol in a  $\delta\Delta G$  value of 5.94 kJ/mol.

Due to the proximate nature of LA's functional sites, values for  $\delta\Delta G_{glc} - \delta\Delta G_{GT}$  ( $= -RT \ln \{K_m^{glc}(wt) * K_i^{GT}(mut) / K_m^{glc}(mut) * K_i^{GT}(wt)\}$ ) have also been calculated (Table

2.4). These values are a measure of the difference in the effects of a mutation on the free energy changes for the interaction with GT and the interaction of the GT-LA complex with glucose. Effects on glucose binding are highlighted by positive differences while negative values reflect a more pronounced effect on GT binding.

Table 2.4 — Effects of LA mutations on GT binding and ability to promote glucose binding in the LS complex.

mLA mutant	$\delta\Delta G_{glc}^a$	$\delta\Delta G_{GT}^a$	$\delta\Delta G_{glc} - \delta\Delta G_{GT}$
	kJ/mol		
F31S	<b>11.82<sup>b</sup></b>	<b>9.41</b>	<b><u>2.41</u></b>
F31Y	<b>4.85</b>	<b>2.07</b>	<b><u>2.78</u></b>
H32A	nd	<b>8.76</b>	nd
H32Y	<b>12.61</b>	<b>6.37</b>	<b><u>6.24</u></b>
Y103P	0.58	-0.31	0.89
H107A	-0.74	<b>5.90</b>	<b><u>-6.65</u></b>
H107W	-1.21	<b>2.24</b>	<b><u>-3.45</u></b>
H107Y	<b>-2.13</b>	<b>5.40</b>	<b><u>-7.53</u></b>
A109P	0.16	-0.31	0.47
L110H*	1.35 (1.44) <sup>c</sup>	<b>9.32 (9.39)<sup>c</sup></b>	<b><u>-7.97 (-7.95)<sup>c</sup></u></b>
L110R*	<b>7.59 (8.87)<sup>c</sup></b>	<b>7.26 (5.9)<sup>c</sup></b>	<b><u>0.33 (1.69)<sup>c</sup></u></b>
Q117A	<b>1.79</b>	<b>5.34</b>	<b><u>-3.55</u></b>
Q117D	0.7	<b>3.12</b>	<b><u>-2.42</u></b>
W118H	nd	<b>11.97</b>	nd
W118Y	0.16	<b>11.53</b>	<b><u>-11.37</u></b>

<sup>a</sup>  $\delta\Delta G$  values are  $\Delta G_{mLA} - \Delta G_{mutant} = -RT \ln \{K_{mLA}/K_{mutant}\}$

<sup>b</sup> Values in bold and/or underlined represent  $\delta\Delta G$ 's that exceed  $\pm 1.5$  kJ/mol

<sup>c</sup> Values in parentheses correspond to the reconstructed 110 mutants

When the effects of the substitutions are represented in this fashion one can clearly see that changes at Phe-31 and His-32 predominantly influence LA's ability to promote glucose binding (H32Y > F31Y > F31S > L110R). On the other hand, changes at His-107, Leu-110 (L110H), Gln-117 and Trp-118 modulate LA's ability to bind to GT (118Y > 110H > 107Y > 107A > 117A > 107W > 117D). It is interesting to note that when the effects of the L110R substitution are represented in this way, the difference  $\delta\Delta G_{glc} - \delta\Delta G_{GT}$  value is positive. This result, taken together with the effects on GT affinity of the L110H substitution, indicate that the introduced arginine sidechain has a significant

influence on LA's ability to promote glucose binding. The difference in effect observed with the histidine and arginine substitutions at 110 is probably caused by an electrostatic or steric effect. A more detailed discussion of these results, in terms of the three-dimensional structure of mLA, is given in Chapter 8.

### 2.5.5 Summary

The site-directed mutants of LA were constructed to probe the roles of aromatic clusters I and II and the 'flexible loop' region in the LS complex. As described in Chapter 1, both aromatic substructures are located in the cleft region of LA and have been implicated by a variety of studies as possible binding sites for GT (Shewale *et al.*, 1984; Koga and Berliner, 1985). Although the results from a considerable number of chemical studies suggest that aromatic cluster I is essential for LA function in the LS complex, the possibility that cluster II might play a functional role could not be excluded by previous data. However, the present results from the mutagenesis of Tyr-103 and Trp-104 in aromatic cluster II indicate that this region has a predominately structural role in LA. Substitutions at these positions clearly influence the folding kinetics of the protein as most of the aromatic cluster II variants did not fold correctly under the conditions used. Recent NMR studies have demonstrated that this region is particularly important for the stability of the molten globule state of LA. In the A-state, the sidechains of Tyr-103, Trp-104 and His-107 form a well-defined hydrophobic cluster that is different from the one observed in the fully folded protein (Alexandrescu *et al.*, 1993; Smith *et al.*, 1994). Nevertheless, the fact that Y103P-mLA has a similar stability to mLA and can be obtained in reasonable yields indicate that this region can accommodate other residue types without overly affecting the folding mechanics of the protein.

In contrast, mutations in both aromatic cluster I and the flexible loop region result in variant LAs that are functionally deficient. These deficiencies are manifested in three main ways:

- essentially normal affinity for GT but reduced ability to promote glc binding
- essentially normal ability to promote glc binding but reduced affinity for GT
- both reduced affinity for GT and ability to promote glc binding

Alterations in the flexible loop region at positions 107, 109 and 110 (L110H)

predominantly affect LA's ability to bind to GT. Position 110 is of considerable interest as the nature of the substituted sidechain at this site produces contrasting functional effects. The results from substituting a His and Arg for Leu-110 suggest that this residue probably lies extremely close to the monosaccharide binding site in the LS complex.

Although both substitutions effect LA's ability to bind to GT, only the longer Arg sidechain causes a decrease in the promotion of glucose binding. At present there is still insufficient information regarding the Leu-110 mutants and it is hoped that the structural studies will be able to shed more light on the causes of these mutants' functional differences.

The effects of substitutions in aromatic cluster I can be divided into two groups. Residues 31 and 32 appear to influence both the strength of binding to GT and the enhancement of glucose binding in the LS complex. In the case of F31Y-mLA, the effects on GT binding are minimal and the reduced LS activity of this variant is almost entirely attributable to a lowered ability to promote glucose binding. It appears that the additional phenolic hydroxyl present in the F31Y-mLA variant directly interferes with monosaccharide binding in the LS complex via a steric or electrostatic effect. On the



other hand, residues 117 and 118 only influence the strength of LA binding to GT.

Although the importance of His-32 and Trp-118 in LA function has been demonstrated (Prieels *et al.*, 1979; Schechter *et al.*, 1974; Bell *et al.*, 1975), the critical functional roles of Phe-31 and Gln-117 have not previously been demonstrated experimentally.

These results have significant relevance in terms of the molecular basis of LA's action in the LS complex. In the proposed 'monosaccharide bridge' model (see section 1.3.4), the interaction of LA with GT and its enhancement of glucose binding are associated with distinct but adjacent regions on the surface of LA (Brew *et al.*, 1979; Lambright *et al.*, 1985). In the mutagenesis experiments, we have been able to separately perturb, albeit to varying degrees, both aspects of LA's function as predicted by the model. If however, the action of LA on GT was of an allosteric nature, the two facets of LA action would undoubtedly be intrinsically linked. The proximity of the two functional regions on the surface of LA is highlighted by the fact that different mutations at the same site affect both facets of LA function by differing amounts. These results tentatively support the monosaccharide bridge model for LA action in the LS complex.

Substitutions in the flexible loop (comprising residues 105-110), adjacent to aromatic cluster I, suggest that this region, in conjunction with Gln-117 and Trp-118, forms an important interaction site between LA and GT in the LS complex. Further extensive mutagenesis experiments carried out in Miami indicate that certain substitutions at Ala-106 (A106K) and Lys-114 (K114N) perturb LA's ability to bind to GT in the LS complex (K. Brew, personal communication). In addition, truncation of the LA molecule after Trp-118 does not compromise its ability to bind to GT or promote glucose binding. These results confirm that LA's interaction site for GT is confined to a relatively small

surface area in the C-terminal portion (residues 105–118) of the molecule. The size of the proposed interaction site is consistent with LA's weak affinity for GT ( $k_d$   $10^{-5}$ M).

The mutagenesis studies have also provided the first direct evidence that residues from LA are involved in binding the galactosyl acceptor. This feature of LA function has long been suspected given its relationship with lysozyme. The critical roles of Phe-31, and His-32 in promoting monosaccharide binding in the LS complex suggests that LA may have retained some vestigial sugar binding properties during its evolutionary development. All three of these residues are located at the lower end of the cleft in a region which in lysozyme corresponds to subsite F (Blake *et al.*, 1967b; Perkins *et al.*, 1981). This functional correlation between LA and LYZ is considered in more detail in Chapter 8.

## Crystallisation

### 3.1 Introduction

In the structure determination process, the first potential stumbling block is usually obtaining crystals of the macromolecule being studied. Crystals must not only be grown but they must be of a suitable size and quality to enable high resolution X-ray diffraction analysis. In some cases, this is fairly straightforward but often the process can be fraught with problems and for many proteins suitable crystals prove to be elusive.

Why is this crystallisation step often such an obstacle? Unlike the crystallisation of small molecules, the success of macromolecular crystallisation depends firstly on a large numbers of parameters, such as temperature, pH and the nature of the buffers, precipitants and additives used to obtain crystals and, secondly, on the complex behaviour of macromolecular systems under these conditions. Proteins are inherently labile and particularly sensitive to their environment. These factors preclude the use of classical crystallisation techniques, such as evaporation or the addition of strong organic solvents. Crystals must therefore be grown using much gentler and restrictive methods. Furthermore, the physical forces behind crystal growth, particularly in the case of

biological macromolecules, are still poorly understood. Protein crystallisation is therefore still, to some extent, an art rather than a science and success is based on both luck and intuition as well as applying certain generalised rules that have been formulated from the vast number of previous crystallisation studies. A brief overview of some of the crystallisation screening methods is given below. For a more detailed exposition on the theoretical and practical aspects of macromolecular crystallisation, the reader is referred to several excellent reviews (Blundell and Johnson, 1976; McPherson, 1990; Giegé and Ducruix, 1992; Weber, 1991).

#### 3.1.1 Screening Methods

Optimal growth conditions for protein crystals are very difficult to predict *a priori* due to the vast number of considerations that need to be taken into account. The inordinate number of variables that influence crystal growth and the often limited amounts of a protein preclude an exhaustive, systematic screening of all possible conditions. Intelligent and intuitive experimental design and subsequent evaluation of the individual and collective trials is therefore essential. Many different methods have been devised to efficiently screen various conditions that may be important when trying to crystallise a new macromolecule. A number of these have been based on incomplete factorial designs that simultaneously evaluate a number of different possible conditions (Carter and Carter, 1979; Carter, 1992). These methods attempt to reduce the possible number of variables by identifying important 'factors' via simultaneous random screening with different reagents, in different mother liquors and at different pHs. A further simplification of this approach has been to bias the initial matrix design towards crystallisation conditions that have been reported for other macromolecules. This sparse

matrix method, or 'Magic 50' as it has become known in our laboratory, has proved to be particularly successful in identifying initial crystallisation conditions for a number of unrelated proteins (Jancarik and Kim, 1991).

A more classical, but related approach to identifying conditions that are conducive to crystallisation is to investigate the optimal pH for protein solubilisation and hence crystallisation, optimal concentration of a particular precipitant, and the effect of temperature. Successful crystal growth depends on reducing the protein's solubility in solution and appropriate crystallisation agents must be selected to gradually force the molecule to a state of minimum solubility. pH is a particularly important variable as it can be used to alter the protein's charge and hence solubility. These various factors can be evaluated in simple grid screens (McPherson, 1990; Weber, 1991). A typical grid matrix can explore the effects of a coarse range of pHs and precipitant concentrations at various temperatures. As with the other crystallisation screens mentioned above, finer grid screens can then be constructed around any promising conditions. Crystal quality can often be further optimised by the addition of small amounts of electrostatic crosslinkers and organic molecules such as isopropanol and detergents (McPherson *et al.*, 1986; Cudney *et al.*, 1994).

#### 3.1.2 Sample Purity

Although crystallographers have little control over many of the parameters that influence crystal growth, they can ensure that the biological macromolecule being studied is highly pure. Not only should the starting material be free of contaminants, but it must also be conformationally homogeneous i.e. free from denatured forms of the protein.

These precautions will ensure that all the molecules have the same surface properties, such as charge distribution, since this will influence the internal structure of the crystal and hence the final crystal quality. Often, unsuccessful attempts to obtain high quality crystals for diffraction studies can be attributed entirely to the purity of the starting material. In the case of proteins obtained by recombinant methods, maintaining the homogeneity of the sample during purification is made easier as the action of proteases, that often give rise to microheterogeneity, is limited. In terms of the present study, the only major concern about sample purity was the presence of misfolded forms of LA. Particular care was taken during purification of mLA and the mLA variants to remove these and only protein preparations that appeared 'native-like' were used for the kinetic and crystallisation experiments.

#### 3.1.3 Crystalline Forms of LA

Historically, crystallisation has often been used in the purification of proteins from relatively crude samples. LA was first crystallised using ammonium sulphate in 1899 by Wichmann and this technique was later used to isolate and identify LA as the 'B' protein of the lactose synthase complex (Brodbeck *et al.*, 1967). Several crystal forms of bovine LA have been obtained (Inman and Bryan, 1966; Fenna, 1982a). Although several of these were suitable for X-ray diffraction studies, determination of the bovine three-dimensional structure was hindered by non-isomorphism of the derivative crystals (Phillips *et al.*, 1987). Often, in situations where initial attempts to define the structure of a particular protein have been unsuccessful, significant progress has been made by switching to other sources of the protein (Kumar *et al.*, 1992). This proved to be the case with the elucidation of the three-dimensional structure of LA.

After a brief dabble with goat LA (Aschaffenburg *et al.*, 1972a,b), attention was focused on LA from human and baboon milk (Aschaffenburg *et al.*, 1979; Fenna, 1982b).

Nevertheless, although suitable crystals could be grown, it was only after several frustrating years searching for heavy atom derivatives, that the structures of baboon and then human LA were finally determined (Smith *et al.*, 1987; Acharya *et al.*, 1991). The following sections document my attempts to obtain suitable crystals for the refractory bovine milk LA (BOLA).

## 3.2 Initial Crystallisation Screens

### 3.2.1 Experimental Design

As mentioned above, several different crystal forms, suitable for high resolution diffraction analysis, can be grown for BOLA (Fenna, 1982a). Even though both the hexagonal and trigonal crystal forms obtained in this study were reported to be suitable for further studies, no significant progress has been made (reflected by a total absence of reports in the literature over the last 12 years). Therefore, a variety of screening methods have been employed to try to obtain some novel crystal forms that are more amenable to structure analysis. These have included coarse screens around the published conditions of human and baboon LA, commercially available sparse matrix screens similar to those of Jancarik and Kim (1991), and grid screens using both ammonium sulphate (AS) and polyethylene glycol (PEG) 4000 as precipitating agents. Details of the various screening procedures are given in the Appendix and Figure 3.1.

Crystallisation trials were carried out using the vapour diffusion method with either hanging or sitting drops (McPherson, 1982). Initial trials were performed using

commercially available BOLA purified from cow's milk (Sigma; cat. # L5385). This relatively impure material (approx. 85% LA by SDS-PAGE), was purified further using

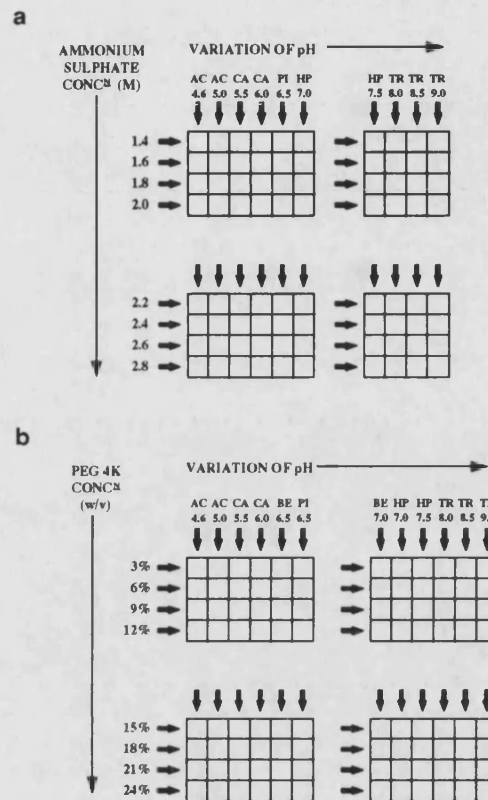


Figure 3.1 — Design of grid screens

**a.** Ammonium sulphate; **b.** PEG 4K. The grid screens were designed to evaluate different pHs (varied horizontally) and the precipitant concentrations (varied vertically). The trials were set up as follows: Each well, represented by each box above, contained 0.7ml of reservoir solution made up of the both buffer and precipitant at the appropriate concentrations. Drops, consisting of 3 $\mu$ l of protein stock (20mg/ml) and 3 $\mu$ l of reservoir solution, were suspended over the reservoir from siliconised coverslips. All buffers were present at a concentration of 0.1M. The buffers abbreviated above are AC, sodium acetate; CA, sodium cacodylate; BE, (N,N-bis[2-hydroxyethyl]-2-aminoethanesulphonic acid (BES); PI, 1,4-piperazine diethanesulphonic acid (PIPES); HP, (N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid] (HEPES) and Tr, Tris-HCl.

HPLC anion exchange chromatography as described for mLA (see Appendix). The elution profile exhibited one major peak and several side peaks (data not shown). These side peaks probably correspond to either contaminants in the preparation (e.g.  $\beta$ -



lactoglobulin) or molecular isoforms of LA (Baumy and Fauquant, 1989). Protein from the major peak (HPLC-BOLA) was collected, dialysed against distilled water, lyophilised and stored at  $-20^{\circ}\text{C}$ . Subsequent crystallisation trials were performed using highly purified samples of mLAs and several of the variant mLAs obtained in Miami (see Chapter 2). Crystallisation conditions were also screened for goat LA (GOLA) using material kindly supplied by Dr. Keith Brew (University of Miami Medical School). Protein stock solutions (20–40mg/ml) were prepared by dissolving the lyophilised material in MilliQ water or 10mM Tris-HCl, pH 8.0 and centrifuging at 13000rpm for 10 min to remove any undissolved particulate matter. These protein stocks could be stored at  $-20^{\circ}\text{C}$  for at least one month without any noticeable effect on the crystallisation process.

In several cases, crystals of sufficient size could not be obtained from ‘primary’ vapour diffusion experiments. Macroseeding techniques were used to improve the size of the initial crystals. In this technique, small pre-formed crystals are introduced into a pre-equilibrated protein and precipitant solution. The precipitant concentration should be at a level at which spontaneous nucleation is prevented but, will allow controlled crystal growth if an external nucleation centre is added. Various different seeding methods have been described and the reader is referred elsewhere (Stura and Wilson, 1992). In the case of the monoclinic III form of mLAs, needles were carefully removed from their parent sitting drop and washed several times with an appropriate reservoir solution. The crystals were then ‘etched’ in a similar solution containing a lower concentration of PEG (about 1% lower) to produce a clean growth surface. These crystals were then introduced into sitting drops that had been equilibrated for 3-5 days (10mg/ml,  $16^{\circ}\text{C}$ ) against a reservoir

solution that did not yield spontaneous crystal growth within the duration of the pre-equilibration period. This seeding process was repeated until the crystals had reached a size sufficient for X-ray analysis.

#### 3.2.2 Bovine Milk LA (BOLA)

Several crystal forms were obtained for the commercially available bovine milk LA when the various crystallisation screens were carried out (Table 3.1). Two of these forms (trigonal and hexagonal) have been previously grown using batch methods (Fenna, 1982a). The other crystal forms of LA obtained in the screening process have not been reported prior to this work.

The first form was identified using the sparse matrix screen (condition #42: 0.05M Potassium phosphate, 20% PEG 8K). Lozenge-shaped crystals appeared after about 3 days and reached a maximum size (approximate dimensions of 0.7mm x 0.5mm x 0.3mm) after 7-10 days. Further characterisation showed that these crystals were orthorhombic, space group  $P2_12_12$ , having systematic absences for  $h00$  and  $0k0$  reflections with  $h$  and  $k$  odd. Diffraction was observed to at least  $2.5\text{\AA}$  using the in-house area detector (AD). The crystal's unit cell dimensions are  $a=72.3\text{\AA}$ ,  $b=105.1\text{\AA}$ ,  $c=117.7\text{\AA}$ . Assuming that there are eight molecules in the crystallographic asymmetric unit (AU), the volume occupied per dalton of protein ( $V_M$ ) is  $1.97\text{\AA}^3$  per dalton; this is well within the range commonly found for protein crystals (Matthews, 1968).

Table 3.1 — Crystal Forms of BOLA

Crystal form	Growth conditions <sup>a</sup>	Space Group	Cell dimensions (Å)	N <sup>b</sup>	V <sub>M</sub> <sup>c</sup>	Comments
Rhombohedral	20mM CaCl <sub>2</sub> 29.5% PEG 4K 0.2M PIPES <sup>d</sup> , pH 6.6 20mg/ml, 37°C	ND <sup>e</sup>	a=b=112.9 c=111.2 γ=120°	—	—	Crystals are small and diffract weakly.
Orthorhombic	0.05–0.08M KP 15–20% PEG 8K 10–15mg/ml, 16–20°C	P2 <sub>1</sub> 2 <sub>1</sub> 2	a=72.3 b=105.1 c=117.7	8	1.97	Diffract to at least 2.5Å spacings. Crystals can also be grown using sodium acetate (pH 5.0)
Monoclinic	1.8M AS 10mM CaCl <sub>2</sub> 0.1M PIPES, pH 6.5–6.6 40mg/ml, 37°C	P2 <sub>1</sub>	a=63.6 b=197.9 c=137.0 β=93.4°	24	2.53	Occasionally obtained using the trigonal conditions. Crystals diffract poorly.
Trigonal	1.8M AS 10mM CaCl <sub>2</sub> 0.1M PIPES, pH 6.5–6.6 20–40mg/ml, 37°C	P321	a=b=93.8 c=66.9 αβ=90° γ=120°	2 or 3	2.99 or 1.99	Large crystals grown with HPLC purified material diffract to beyond 2Å
Hexagonal	1.8M AS 10mM CaCl <sub>2</sub> 0.1M PIPES, pH 6.5 20–40mg/ml, 37°C	P622	a=b=93.8 c=66.9 αβ=90° γ=120°	1	2.99	Growth under the conditions given is highly variable. Crystals diffract to medium resolution

<sup>a</sup> Growth conditions refer to that of the reservoir. Protein concentration is that prior to setting up the drop

<sup>b</sup> N, number of molecules in the asymmetric unit.

<sup>c</sup> V<sub>M</sub>, Å<sup>3</sup>/dalton (Matthews, 1968).

<sup>d</sup> Abbreviations used: PIPES, 1,4-piperazine diethanesulphonic acid. KP, Potassium dihydrogen orthophosphate. AS, ammonium sulphate.

<sup>e</sup> ND, space group not determined precisely due to poor and incomplete data.

Similar crystals were also observed in the PEG 4K grid screen at 16°C (10mg/ml protein, 0.1M sodium acetate pH 5.0, 10mM CaCl<sub>2</sub>, 15–18% PEG 4K) and could be obtained using a variety of different PEG polymer chain lengths (3350-8000). Unfortunately, due to the large number of molecules in the crystallographic asymmetric unit, this form is unsuitable for further detailed structural studies.

A second crystal form was obtained using both *Crystal Screen I* and the PEG 4K grid screen at 16°C (0.1M Tris pH 8.0-9.0, 10mM CaCl<sub>2</sub>, 18-24% PEG 4K). These crystals typically grew as clusters of thin fragile needles. Unfortunately, all attempts to improve these crystals failed and further analysis was not possible. The third new crystal form was rhombohedral (R3 or R32) with unit cell dimensions  $a=b=112.9\text{\AA}$ ,  $c=111.2\text{\AA}$ ,  $\alpha=\beta=90^\circ$   $\gamma=120^\circ$ . These hexagonal rods grew at 37°C but diffracted poorly to a maximum resolution of 4Å. Attempts to improve crystal quality using macroseeding were largely unsuccessful and so further studies on this crystal form were abandoned.

The two most promising forms of BOLA were the trigonal and hexagonal crystals described by Fenna (1982a). Unlike the new crystal forms, these appear to be suitable for detailed diffraction analysis. Trigonal (*P321*) crystals were obtained over a wide range of ammonium sulphate concentrations at 37°C using the ammonium sulphate (AS) grid screen. Large triangular ellipsoid crystals were obtained using hanging drops, consisting of 5µl 40mg/ml HPLC-BOLA stock and 5µl reservoir solution, equilibrated against a reservoir of 0.1M PIPES pH 6.5 containing 1.8M AS and 10mM CaCl<sub>2</sub>. These crystals appear after about 7 days and typically reach a maximal size of 0.7mm in width and 0.3mm thick after 14 days. The unit cell dimensions are:  $a=b=93.7\text{\AA}$ ,  $c=66.9\text{\AA}$  and the crystals diffract strongly to Bragg spacings of at least 2Å. The space group

assignment was confirmed by examining pseudo precession plots using the program *HKLVIEW* (CCP4, 1994). Theoretical solvent calculations, based on the unit cell volume and molecular weight of BOLA (14200Da), suggested that each crystallographic AU contains either two (59% solvent) or three (38% solvent) LA molecules. Both cases, lying at the upper and lower limits normally observed for globular proteins (Matthews, 1968), are possible and could not be distinguished at this stage. On several occasions, crystals of a different morphology to the trigonal form grew in the same drop. These crystals are monoclinic but have extremely large unit cell dimensions and an estimated 24 molecules per AU. Hexagonal crystals (*P622*) were more difficult to obtain but grew under essentially the same conditions as the trigonal form (Table 3.1). The hexagonal cell dimensions are very similar to the trigonal form and there is only one LA molecule per AU. Although they did not diffract as strongly as the trigonal form, reflections were observed to at least 2.9Å.

#### 3.2.3 Recombinant LA (mLA)

mLA differs from the wild type protein (BOLA) only by the presence of an additional methionine residue at its N-terminus and an internal methionine to valine substitution (M90V). These changes have minimal effects on the protein's kinetic and spectral properties and it was hoped that the majority of the crystal forms seen with BOLA (Table 3.1) could be repeated for mLA. In hindsight, this assumption was misguided. Neither the orthorhombic (*P2<sub>1</sub>2<sub>1</sub>2*), nor the trigonal (*P321*) forms seen for BOLA have been reproduced for the recombinant protein. Despite numerous trials, mLA has not yet been crystallised in the trigonal form. Nonetheless, large hexagonal crystals,

Table 3.2 — Crystal Forms of mL A

Crystal form	Growth conditions <sup>a</sup>	Space Group	Cell dimensions (Å)	N <sup>b</sup>	V <sub>M</sub> <sup>c</sup>	Comments
Orthorhombic	0.2M AS 30% PEG 8K 0.1M Caco <sup>d</sup> , pH 5.23 20mg/ml, 16°C	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	a=58.3 b=115.0 c=153.5	8	2.26	Diffract to 3Å but have a large number of molecules in the AU
Monoclinic I	0.15M NaAc 22-25% PEG 4-8K 0.1M BTP, pH 8.0-8.2 20mg/ml, 16°C	<i>P</i> 2 <sub>1</sub>	a=78.0 b=60.6 c=117.2 β=97.2°	8	2.4	Crystals diffract poorly
Monoclinic II	0.15M NaAc 22-25% PEG 6K 0.1M BTP, pH 8.0 10μl toluene 20mg/ml, 16°C	<i>P</i> 2 <sub>1</sub>	a=78.1 b=60.7 c=172.4 β=97.0	12	2.38	Crystals diffract poorly
Monoclinic III	0.15M NaAc 10mM CaCl <sub>2</sub> 22-25% PEG 4K 0.1M Tris-HCl, pH 8.0 20mg/ml, 16°C	<i>P</i> 2 <sub>1</sub>	a=58.0 b=60.6 c=76.7 β=96.9	4	2.37	Diffract to 3Å in-house but are very sensitive to temperature and X-rays. Larger crystals can be grown using MgAc
Hexagonal	1.8M AS 10mM CaCl <sub>2</sub> 0.1M PIPES, pH 6.5 1-5% glycerol 20-40mg/ml, 37°C	<i>P</i> 622	a=b=93.8 c=67.6 αβ=90 γ=120	1	3.02	Crystals diffract to at least 2.5Å.

<sup>a</sup> Growth conditions refer to that of the reservoir. Protein concentration is that prior to setting up the drop

<sup>b</sup> N, number of molecules in the asymmetric unit.

<sup>c</sup> V<sub>M</sub>, Å<sup>3</sup>/dalton (Matthews, 1968).

<sup>d</sup> Abbreviations used: PIPES, 1,4-piperazine diethanesulphonic acid; KP, Potassium dihydrogen orthophosphate; AS, ammonium sulphate; Caco, Sodium cacodylate; NaAc, Sodium acetate; BTP, *bis*-Tris propane; MgAc, Magnesium acetate.

identical to those seen with BOLA, can be reproducibly grown for mL A under conditions where the calcium levels are saturating (Table 3.2). The quality of these crystals is further improved by the addition of small amounts of glycerol (1-5% v/v). The inability of mL A to adopt the packing required for trigonal crystal formation was initially ascribed to mL A's additional N-terminal methionine residue. However, subsequent crystallisation trials with a sample of mL A that had been treated with CNBr, to specifically remove the additional N-terminal residue, exclusively yielded the hexagonal form. These crystals diffracted to Bragg spacings of at least 2.5Å and looked promising for further analysis. Although removal of the N-terminal methionine did not influence the space group of the crystals, there was a noticeable improvement in crystal size and quality when the CNBr-treated mL A was used. Pseudo precession plots displayed with *HKLVIEW* (CCP4) were used to confirm the previous space group assignment.

In addition, several other crystal forms have been grown for mL A using the various screening procedures (Table 3.2). Orthorhombic crystals can be grown for mL A but they are different from those obtained with BOLA. These crystals were grown using conditions similar to *Crystal Screen I* #15 (0.2M ammonium sulphate, 0.1M sodium cacodylate pH 5.23, 25% PEG 8K) at 16°C. Further analysis showed that the space group was  $P2_12_12_1$  with unit cell dimensions of:  $a=58.3\text{\AA}$ ,  $b=115.0\text{\AA}$ ,  $c=153.5\text{\AA}$ . Based on theoretical solvent content calculations ( $V_M=2.26$ ), this form is estimated to contain eight molecules per asymmetric unit. As with the orthorhombic form of BOLA, these crystals are probably not suitable for further study.

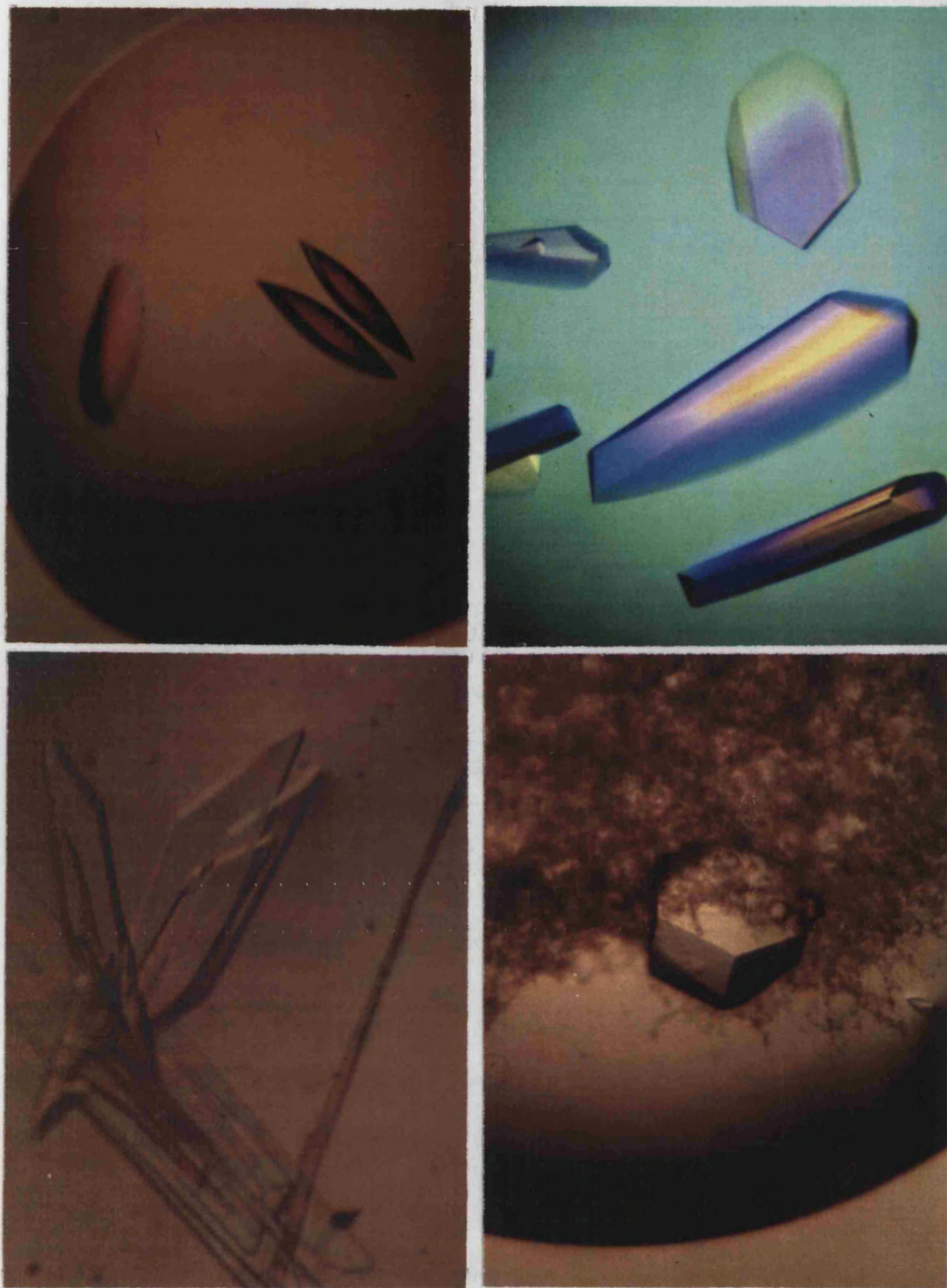


Figure 3.2 — Crystal Forms of BOLA and mLA.

The morphology of some of the crystal forms obtained for bovine milk LA and the recombinant LA are shown. The forms are trigonal ( $P321$ ) BOLA (*top left*), orthorhombic ( $P2_12_12$ ) BOLA (*top right*), monoclinic III ( $P2_1$ ) mLA ( $Mg^{2+}$  form; *bottom left*) and hexagonal ( $P622$ ) mLA (*bottom right*).



A second crystal form was also observed in the sparse matrix screen at 16°C. *Crystal Screen I*'s condition #22 (0.2M sodium acetate, 0.1M Tris-HCl pH 8.5, 30% PEG 4K) produced clusters of very fine needles. Extensive refinement of this condition yielded a form that appeared to be amenable to further studies. These crystals have a 'blade-like' morphology – relatively broad in one dimension but very thin in the other two (Figure 3.2). Initial attempts to improve the crystal size succeeded in increasing the broadness but had little effect on the other two dimensions. It became apparent that a more radical change in growth conditions was required and so a range of different buffers were tested. When *bis*-Tris propane was used instead of Tris-HCl, a significant improvement in all the crystal dimensions was seen. These larger crystals are monoclinic, space group  $P2_1$ , having systematic absences for  $0k0$  reflections with  $k$  odd. This form, termed monoclinic I, has unit cell dimensions of  $a=78.0\text{\AA}$ ,  $b=60.6\text{\AA}$ ,  $c=117.2\text{\AA}$ ,  $\beta=97.2^\circ$  and an estimated 8 molecules per AU. A related form (monoclinic II) has approximately 12 molecules per AU; a result of a 50% increase in the  $c$  dimension of the unit cell (Table 3.2). Both these monoclinic forms (I and II) were clearly unsuitable for further study due to the complexity of their crystallographic AUs.

Finally, the crystallisation condition that yielded the original needles was re-examined. Repeated macroseeding was used to improve the crystal size sufficiently to allow further characterisation. Analysis showed that these crystals (monoclinic III) have unit cell dimensions of:  $a=58.0\text{\AA}$ ,  $b=60.6\text{\AA}$ ,  $c=76.7\text{\AA}$ ,  $\beta=96.9^\circ$ , giving a unit cell volume of  $2.70 \times 10^5 \text{\AA}^3$ . If there are 4 molecules per asymmetric unit, the  $V_M$  is  $2.37\text{\AA}^3$  per dalton which is well within acceptable limits (Matthews, 1968). Careful refinement of the growth conditions, by replacing sodium acetate with magnesium acetate and switching

from hanging to sitting drops, resulted in an appreciable increase in crystal size.

Subsequent analysis indicated that this  $Mg^{2+}$  form was essentially isomorphous with the monoclinic III  $Na^+$ -form with only a subtle change in cell dimensions. Unfortunately, both forms often exhibit contact twinning on planes parallel to the  $b$  axis and extreme care was required when selecting crystals for diffraction analysis. Nonetheless, these crystals were suitable for further studies and diffracted to at least  $3\text{\AA}$  on the in-house area detector.

#### 3.2.4 mLA Mutants

Due to the limited amount of material available for most of the mLA variants and the poor success of the screening procedures in identifying new forms of mLA, crystallisation trials on the mutants were initially confined to those used to obtain the monoclinic III form. Although two mutants (H32Y, H107W) crystallised under these conditions, the crystals were very small and severely twinned. Subsequently, it became apparent that the monoclinic III form was not ideal for studying the mutant proteins due to a combination of reasons that are described in the following chapters. As no suitable form of mLA had been obtained, cross-seeding techniques could not be used (Stura and Wilson, 1992). Screening using both *Crystal Screen I* and *II* failed to produce any useable crystals. However, good quality crystals were subsequently grown at  $37^{\circ}\text{C}$  using the hexagonal conditions (0.1M PIPES pH 6.5, 1.8-1.95M AS, 10mM  $\text{CaCl}_2$ , 1-5% glycerol). Crystals can be grown for the mLA mutants F31Y, H32Y, Y103P, A109P and H107Y using hanging drops consisting of  $3\mu\text{l}$  20mg/ml protein stock and  $3\mu\text{l}$  reservoir solution. Characterisation of the A109P-mLA crystals showed that they are identical to

the hexagonal form of both BOLA and mLA. Diffraction was observed to at least 2.5Å on the in-house detector.

### 3.2.5 Goat LA (GOLA)

Initial crystals were obtained using condition #20 from *Crystal Screen I* (0.2M AS, 20-25% PEG 4K, 0.1M sodium acetate pH 4.5) and a 20mg/ml GOLA stock solution. The crystals appeared as clusters of thin plates within 7 days and reached their maximum size after several weeks. Unfortunately, the growth conditions were not particularly reproducible and considerable variation in crystal quality was observed under the same condition. The crystal size and quality could sometimes be improved by the addition of 0.5% (w/v)  $\beta$ -octylglucoside. Further analysis showed that these crystals are monoclinic ( $P2_1$ ), having systematic absences for  $0k0$  reflections with  $k$  odd. This crystal form is similar to one grown by dialysis under low salt conditions (Aschaffenburg *et al.*, 1972a). The unit cell dimensions are:  $a=32.5\text{\AA}$ ,  $b=89.7\text{\AA}$ ,  $c=45.2\text{\AA}$ ,  $\beta=94.5^\circ$  giving a unit cell volume of  $1.32 \times 10^5 \text{\AA}^3$ . Each of the two AU's probably contains two LA molecules ( $V_M=2.31\text{\AA}^3$  per dalton). Despite their small size, these crystals diffracted strongly to better than 2.5Å resolution and were suitable for detailed analysis.

### 3.3 Summary

The crystallisation trials were a mixed success. Numerous different crystal forms could be obtained for both wild type and recombinant bovine LA but, in nearly all cases, the crystals were not ideal for further study. This was often due to the size and complexity of their crystallographic asymmetric units and their poor diffraction quality. The uncommonly large numbers of molecules in most of the crystal forms studied

appears to reflect bovine LA's aggregation properties in solution (Hill and Brew, 1975; Kronman, 1989). The primary aim of the present structural studies was to determine whether the altered functional behaviour of the mLA variants is due to local conformational changes at the mutation site. Therefore, crystal forms that contain large numbers of molecules in the AU are not ideal as any subtle conformational changes resulting from the mutations will probably be masked due to the numerous intermolecular contacts. This is particularly pertinent in this work as the majority of the mutation sites are located on the surface of the LA molecule. Nonetheless, the crystallisation trials produced a couple of crystal forms that were suitable for detailed analysis. The hexagonal form is the most promising as good quality crystals can be obtained for both mLA and a number of the variants.

During the crystallisation trials, it became apparent that the recombinant mLA had an altered behaviour in solution, compared to BOLA. mLA exhibits a reduced solubility compared to the wild type protein, that is particularly noticeable at low pH. This change in mLA's physical properties might result from the additional methionine residue present at its N terminus. Although this additional hydrophobic residue does not affect the recombinant LA's function in the lactose synthase complex (see Chapter 2), it might make the protein more susceptible to aggregation and limit the number of possible crystal forms. Removal of the additional methionine has no effect on the crystals obtained with the recombinant protein at 37°C. However, widespread trials with the CNBr-treated mLA have not been performed due to the lack of success in finding a suitable crystal form for the wild type protein. Another factor that might contribute to the different behaviour of mLA and BOLA is the use of lyophilisation in the purification

of mLA and the mutants. Lyophilisation can affect the conformation of macromolecules and laser Raman scattering studies have shown that lyophilisation alters the three-dimensional structure of LA (Yu, 1974). Obviously the effect of such damage on the crystallisation of LA requires further investigation.

## Data Collection and Processing

### 4.1 Diffraction Theory

X-ray diffraction is the classical method for determining the three-dimensional structures of proteins and nucleic acids at atomic resolution. This visualisation process is analogous to that carried out with an optical microscope. The sample is illuminated by electromagnetic radiation of a suitable wavelength and the resulting diffraction pattern is recorded and recombined to generate an image of the sample. X-rays are used because they have a wavelength that is close to the interatomic distance of most atoms. This relationship is important because the variation in diffraction intensity with direction, resulting from the scattering of X-rays by the electron clouds of atoms within a molecule, arises because the path differences taken by the scattered X-ray beams are of the same magnitude as the separation of the atoms within the molecule.

In order to study atomic structure by X-ray diffraction, the sample must first be crystallised. This is because single molecules diffract X-rays weakly and only a little structural information can be obtained. Crystals, on the other hand, consist of regular arrays of molecules arranged in identical orientations so that the diffracted beams are

amplified to produce strong, detectable X-ray beams. The resultant diffraction pattern therefore reflects a whole series of molecules within the crystal. Consequently, the final X-ray structure is a spatially averaged representation of the protein. A brief overview of the principles of X-ray diffraction is given below but, for a more detailed description, the reader is referred to several excellent texts (Blundell and Johnson, 1976; Rhodes, 1993; Drenth, 1994).

#### 4.1.1 Laue Conditions

Diffraction is only observed from a crystal when the phase difference between the waves scattered by successive unit cells is equal to an integral multiple of  $2\pi$ . These conditions are mathematically expressed by the Laue equations:

$$\mathbf{a} \cdot \mathbf{S} = h$$

$$\mathbf{b} \cdot \mathbf{S} = k$$

$$\mathbf{c} \cdot \mathbf{S} = l$$

[Eq. 4.1]

where  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  are the unit cell vectors,  $\mathbf{S}$  is the diffraction vector and  $h$ ,  $k$ , and  $l$  are integers (reflection indices).

#### 4.1.2 Bragg's Law

Another representation of these diffraction conditions is given by the well known Bragg's Law.

$$n\lambda = 2d\sin\theta$$

[Eq. 4.2]

where  $n$  is an integer,  $\lambda$  is the wavelength,  $d$  is the interplanar spacing and  $\theta$  is the incidental angle of the X-ray beam. W. L. Bragg visualised X-ray diffraction as a simple

reflection of the X-ray waves against an imaginary series of planes within the crystal.

Bragg's Law states that if the path difference for waves reflected by successive sheets of atoms is a whole number of wavelengths, the resultant waves will combine to produce a strongly diffracted beam. In other words, beams diffracted from a set of parallel crystal planes with interplanar spacing  $d_{hkl}$  will interfere constructively only when the incidental angle of the X-rays ( $\theta$ ) satisfies the Bragg equation (i.e. resulting in an integral number ( $n$ ) of wavelengths). As the angle of diffraction ( $\theta$ ) is inversely related to the interplanar spacing ( $d_{hkl}$ ), planes that are close together will produce scattered beams at large angles. This equation is central to our understanding of diffraction and it can be used to predict the position of any diffracted ray.

### 4.1.3 Reciprocal Lattice and Ewald Sphere

A crystal can be regarded as a three-dimensional lattice that produces a three-dimensional X-ray diffraction pattern. There is an inverse relationship between the diffraction pattern and the crystal lattice. The reciprocal (diffraction) lattice, derived from the crystal lattice, is a theoretical, imaginary lattice that is useful for constructing the directions of diffraction by a crystal. In the case of orthogonal unit cells, the reciprocal unit cell has axes  $\mathbf{a}^*$  lying along  $\mathbf{a}$ ,  $\mathbf{b}^*$  lying along  $\mathbf{b}$  and  $\mathbf{c}^*$  along  $\mathbf{c}$  and units of  $\text{\AA}^{-1}$ .

The principle of diffraction, embodied by Bragg's Law and the Laue equations, is best illustrated by a geometric construction proposed by Ewald (Figure 4.1). As the crystal is rotated in the X-ray beam, the reciprocal lattice moves about a fixed origin. Diffraction arises when each point in the reciprocal lattice passes through a sphere, termed 'the sphere of reflection'.



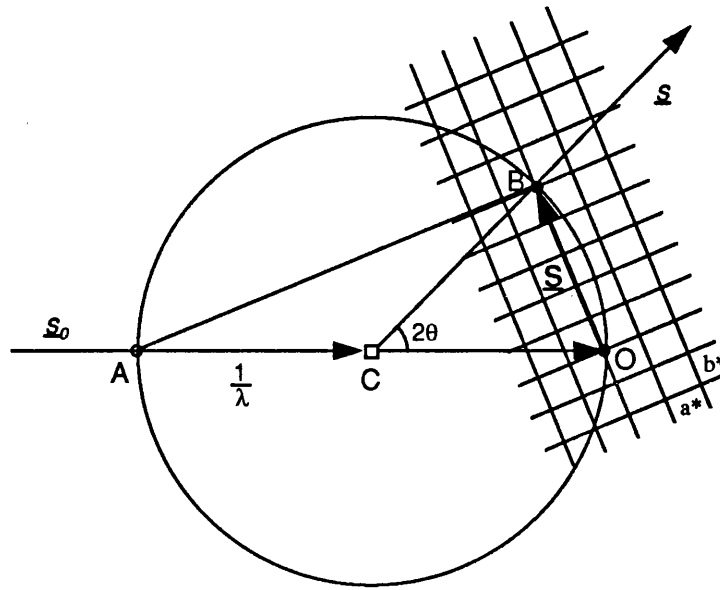


Figure 4.1 — The Ewald construction

See text for details. The sphere has radius  $1/\lambda$ .  $\underline{s}_0$  indicates the direction of the incident beam;  $\underline{s}$  indicates the direction of the scattered beam. The grid represents a portion of the  $a^*b^*$  plane of a reciprocal lattice.  $\underline{S}$  is the reciprocal lattice vector  $\underline{S} = ha^* + kb^* + lc^*$ .

The crystal (C) is taken as the centre of a sphere with radius  $1/\lambda$ . The origin of the reciprocal lattice (O) is placed at the point at which the X-ray beam ( $\underline{s}_0$ ), travelling in the direction AC, exits the sphere after passing directly through the crystal. As the crystal is rotated, diffraction will occur in the direction  $\underline{s}$  when the reciprocal lattice point B ( $h,k,l$ ) passes through the sphere. The point B represents the  $050$  reflection arising from the planes of spacing  $d_{050}$ . As the angle  $OCB = 2\theta$  and the vector  $\underline{S}$  (OB) is perpendicular to the diffracting crystal plane, it can be shown that  $OB = 2 \times CO \times \sin\theta = 2 \times (1/\lambda) \times \sin\theta$ . Furthermore, OB has a length  $1/d_{050}$ , so  $1/d_{050} = 2 \times (1/\lambda) \times \sin\theta$  which is Bragg's Law (Eq. 4.2). Therefore under these circumstances Bragg's Law is fulfilled and a reflection is observed.

#### 4.1.4 Electron Density Equation

What is the relationship between the observed diffraction pattern and the atomic structure? One of the drawbacks of using X-rays to visualise atomic structure is that there is no physical means to focus the diffracted beams and reconstruct the image as can be done in the optical light microscope. In order to reconstitute the image, the diffraction pattern must be mathematically recombined in the correct phase relationship. This can be achieved using the electron density equation:

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} F(hkl) e^{-2\pi i(hx + ky + lz)} \quad [\text{Eq. 4.3}]$$

where  $V$  is the volume of the unit cell,  $hkl$  are indices of the diffracted ray arising from planes  $d(hkl)$ , and  $F(hkl)$  is the structure factor of the diffracted ray. The structure factor  $F(hkl)$  is a complex quantity that is used to represent the amplitude,  $F(hkl)$ , and phase,  $\alpha(hkl)$ , of a particular reflection:

$$\begin{aligned} F(hkl) &= F(hkl) e^{i\alpha(hkl)} \\ &= \sum_{n=1}^N f_n e^{-B_n(\sin^2\theta/\lambda^2)} e^{2\pi i(hx_n + ky_n + lz_n)} \end{aligned} \quad [\text{Eq. 4.4}]$$

where  $N$  is the number of atoms in the unit cell,  $x_n$ ,  $y_n$ ,  $z_n$  are the fractional coordinates of the  $n$ th atom,  $f_n$  is the atomic scattering factor and  $B_n$  is the Debye-Waller temperature ( $B$ -) factor for the  $n$ th atom. The  $B$ -factor accounts for reductions in reflection intensities due to the thermal motion of atoms ( $B = 8\pi^2\bar{u}^2$  where  $\bar{u}^2$  is the isotropic mean square amplitude of vibration perpendicular to the reflecting planes).

The amplitude can be determined directly from the intensity  $I(hkl)$  of each reflection recorded in the diffraction pattern (where  $|F(hkl)|^2 \propto I(hkl)$ ). In contrast, the

phase angle  $\alpha(hkl)$  of each diffracted X-ray beam, relative to the unscattered beam, is lost on recording the diffraction pattern and must be calculated indirectly. This so-called 'phase problem' can be overcome by using methods such as isomorphous replacement (SIR and MIR), multiwavelength anomalous diffraction (MAD) and molecular replacement (MR). In the following chapter, determination of protein phase angles using the MR method is described in more detail.

Therefore, if the amplitudes and phases are known for all reflections, the electron density can be calculated for each point  $xyz$  in the unit cell (Eq. 4.3). This electron density can then be interpreted to reveal the three-dimensional structure of the molecule. For a number of reasons, initial phase estimates often contain a significant error. Consequently, once the initial structure has been determined, structure factors can be calculated from the positional and thermal parameters of the atoms (Eq. 4.4) and used to improve the protein phases through refinement.

## 4.2 In-House Data Collection

### 4.2.1 Introduction

One of the most important advances in modern day crystallography has been the development of sophisticated recording devices (electronic area detectors and image plates) that allow the rapid acquisition of X-ray diffraction data. These instruments have revolutionised protein X-ray crystallography by considerably reducing crystal exposure times, the time taken to process data and as a consequence the time required for data collection.

Electronic area detectors (multi-wire proportional counters) are based on gas-filled ionisation chambers. The gas-filled chamber has two cathodes and an anode, each consisting of either horizontal or vertical parallel wires about 1-2mm apart. X-ray photons enter the detector via a beryllium window and cause the ionisation of xenon gas atoms. The liberated electrons ionise neighbouring atoms resulting in a collision cascade that produces about 300 ion and electron pairs per absorbed X-ray photon ( $\lambda=1.55\text{\AA}$ ). The subsequent secondary ionisation produced as the electrons are accelerated in an electric field between the first cathode and the anode, results in a measurable signal. The electrons hit the anode and the ions hit the second cathode. Each 'hit' is registered as one count and the position of the incident photon is electronically determined.

Area detectors are very sensitive and, combined with rapid data processing, allow data to be collected approximately 50 times faster than film based methods. Area detectors also differ from film-based methods in that they can scan through a diffraction spot every  $0.2^\circ$ . This is made possible by the immediate processing of data from each frame. This contiguous, incremental measurement of each diffraction spot results in a low background and allows the construction of a three-dimensional profile for each reflection. This is particularly advantageous in the measurement of weak reflections. During an experiment, the detector's acquisition and analysis software automatically combines the experimental data into a complete frame of information. Each 'frame' contains the reflection positions and intensities gathered by the detector at a specific set of goniometer angles over a specified period of time.

Due to non-uniform sensitivity and the geometric distortion produced by the system, careful calibration of the detector is required. Variation in sensitivity over the

detector surface is measured at a particular crystal-to-detector distance by collecting a 'flood field' from an  $^{53}\text{Fe}$  source. Geometric distortion is measured using a 'brass plate', a precision-tooled grid of holes that covers the detector screen. Spots from the grid are accumulated using the  $^{53}\text{Fe}$  source and this information is used by the data processing software to account for the distortion.

### 4.2.2 Experimental Methods

X-ray diffraction data were collected at room temperature on the in-house Siemens X-1000 Area Detector System. X-rays were generated using a Siemens copper rotating anode (0.3mm x 0.3mm effective source size) typically operating at 40kV and 70mA.  $\text{CuK}_\alpha$  radiation ( $\lambda=1.5418\text{\AA}$ ) was selected using a graphite monochromator with a limiting collimator aperture of 0.5mm in diameter. Crystals were mounted in thin-walled quartz capillaries and stored at 16°C until required. Before any experimental data were collected, the crystal-to-detector distance was set according to the crystal's unit cell dimensions (usually maximum cell edge / 10) and the detector calibrated as described above. Crystals were mounted on a goniometer with plasticene and aligned in the X-ray beam so that the crystal remains in the beam during rotation about the spindle. Frames of data were recorded with the crystal oscillating through 0.25° steps using a variable exposure time. X-ray data from each crystal were processed with the *XDS* package of programs. Datasets from different crystals were merged together using *XSCALE* to obtain a final set of scaled intensities.

### 4.2.3 Data Processing and Analysis

Diffraction data were processed using the programs *XDS* and *XSCALE* (Kabsch, 1988a,b; Kabsch, 1993). This software allows rapid data processing even in cases where the space group and unit cell dimensions are initially not known. The individual steps of data processing are summarised in Figure 4.2. The program is run as a batch file with the user supplying various information such as the brass plate image, location and name of the diffraction data,  $2\theta$  swing angle, coordinates for the centre of the beam and details of the crystal spacegroup and cell dimensions. The program carries out a number of user-defined steps. *XYCORR* calculates a look-up table of spatial correlations using the brass plate image. Due to the nature of the detector geometry, spots cast by the brass plate deviate from their theoretical positions and so the observed distortions in both the  $x$  and  $y$  directions are calculated. *INIT* estimates the initial background signal by first analysing the readings at each pixel over the first ten oscillation images and then smoothing these values according to a known function (Kabsch, 1988b). *COLSPOT* locates strong diffraction spots for determination of the crystal orientation. *IDXREF* calculates the reduced cell, indexes the spots found by *COLSPOT* and automatically rates all 44 of the possible lattice types (Kabsch, 1993). At this stage, the user can obtain the probable lattice type and cell dimensions for an unknown crystal. Individual three-dimensional reflection profiles are collected by *COLPROF* from the remainder of the rotation images, and the background at each pixel is updated. *COLPROF* also periodically refines experimental parameters such as the unit cell dimensions, beam centre and crystal-to-detector distance. *PROFIT* estimates the intensity of each reflection from its three-

dimensional profile by integration. In *CORRECT*, these integrated reflection intensities are corrected by the Lorentz, polarisation and air absorption factors.

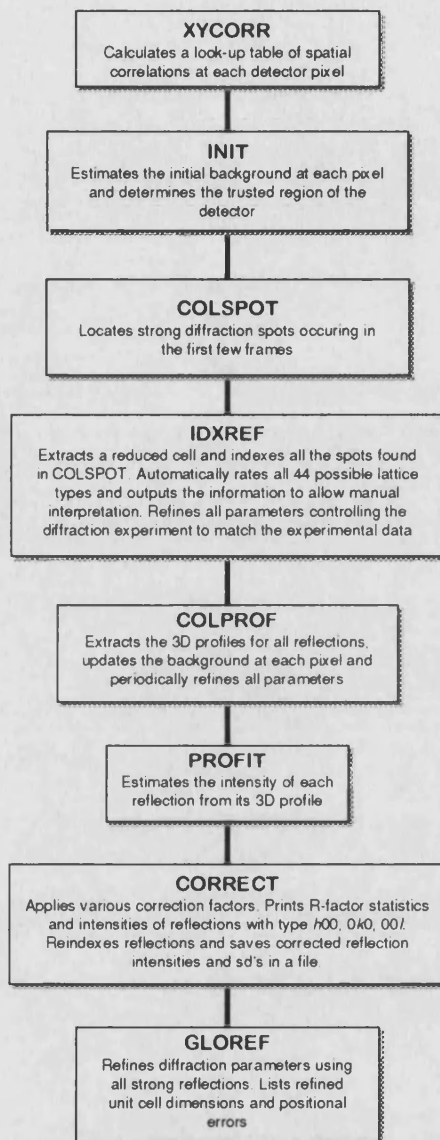


Figure 4.2 — Data processing using *XDS*

The data is scaled and merged to produce a dataset containing only reflections of one asymmetric unit. Various statistics and other information are output to aid in the determination of the correct space group if it is not already known. All reflections are

reindexed and their corrected intensities and standard deviations are output. Finally, **GLOREF** refines the experimental parameters using all the strong reflection spots.

The final stage of data processing is to analyse and merge data from different data collection runs or from different crystals. This was carried out using the program **XSCALE**. This program performs in a similar fashion to **CORRECT** in **XDS** but provides additional statistics and can also be used to merge and scale separate datasets. The output statistics include correlation coefficients for the various datasets and merging and statistical *R*-factors. The *R*-factor for symmetry equivalent reflections ( $R_{sym}$ ) is:

$$R_{sym} = \frac{\sum_{hkl} \sum_i |I_{hkl,i} - I_{average,hkl}|}{\sum_{hkl} \sum_i |I_{hkl,i}|} \quad [\text{Eq. 4.5}]$$

Structure factor amplitudes were calculated from the **XSCALE** output using either **X2L** (in-house program, G. Taylor) or **TRUNCATE** (CCP4).

#### 4.2.4 Diffraction Data

Due to the many different crystal forms observed with BOLA, mL A and mL A mutants, a large number of different X-ray datasets have been collected. Only those that have subsequently been used for structure solution are reported below.

##### 4.2.4.1 BOLA Trigonal (*P*321)

The triangular ellipsoid crystals were aligned so that the triangular face was normal to the X-ray beam. Data were collected from two crystals using a crystal-to-detector distance of 120mm and an exposure time of 150-180sec/frame. The  $2\theta$  swing was fixed at  $25^\circ$  so that diffraction at the edge of the detector screen corresponded to a



resolution of approximately 2Å. Initially 504 frames were collected from the first crystal; this crystal was then translated and a further 134 frames were collected. Finally, a third dataset, comprising of 242 frames, was collected from the second crystal.

Table 4.1 — *XSCALE* merging statistics for trigonal AD-LAB1 data.

Resolution shell (Å)	Total No. Observed Reflections	Total No. Unique Reflections	$R_{sym}$ (%)	Total No. Possible Reflections	Completeness ( $I > 0\sigma$ ) (%)
∞ – 10.0	387	170	3.5	226	75.2
10.0 – 6.0	2127	718	4.2	722	99.4
6.0 – 4.0	6362	2104	4.2	2113	99.6
4.0 – 3.0	11166	4016	5.9	4027	99.7
3.0 – 2.5	11491	4970	11.5	5001	99.4
2.5 – 2.4	3190	1518	16.6	1528	99.3
2.4 – 2.3	3600	1802	18.0	1826	98.7
2.3 – 2.2	3870	2134	19.0	2175	98.1
2.2 – 2.1	3997	2429	24.7	2582	94.1
2.1 – 2.0	3210	2208	29.2	3141	70.3
∞ – 2.0	49400	22069	7.0	23341	94.6

Table 4.2 — *AGROVATA* merging statistics for trigonal AD-LAB2 data.

Dmin (Å)	$R_{merge}(I)$	$R_{cum}$	$I/\sigma$	Total No. Observed Reflections	Total No. Unique Reflections	Multiplicity	Completeness ( $I > 0\sigma$ ) (%)
8.47	0.029	0.029	21.0	837	307	2.7	88.6
6.00	0.037	0.033	17.4	1892	594	3.2	100
4.90	0.039	0.036	16.9	2509	752	3.3	100
4.25	0.038	0.037	16.5	2864	865	3.3	99.2
3.80	0.043	0.039	15.3	3372	994	3.4	100
3.47	0.050	0.041	13.1	3592	1075	3.3	99.2
3.21	0.059	0.044	11.6	3840	1160	3.3	99.5
3.00	0.077	0.046	8.9	4050	1265	3.2	100
2.83	0.093	0.049	7.6	4099	1320	3.1	99.5
2.69	0.114	0.053	6.1	4178	1400	3.0	99.8
2.56	0.128	0.056	5.3	4204	1467	2.9	100
2.45	0.151	0.059	4.7	4280	1533	2.8	99.8
2.36	0.174	0.062	4.0	4265	1590	2.7	99.6
2.27	0.192	0.065	3.5	4263	1667	2.6	100
2.19	0.207	0.068	3.4	4015	1698	2.4	99.2
2.12	0.269	0.071	2.6	3820	1751	2.2	99.1
2.06	0.276	0.073	2.5	3394	1727	2.0	95.3
2.00	0.330	0.075	2.0	2409	1482	1.6	79.2
∞–2.0	–	0.075	8.8	61883	22647	2.8	97.7

As there are two alternative, non-equivalent ways that diffraction spots can be indexed in the trigonal space group, particular care was taken when the three datasets were merged. To avoid possible problems in later stages, two different scaled datasets were produced. The first (AD-LAB1) consisted of only reflections from the first 504 frames collected from the first crystal. The *XSCALE* merging statistics for this dataset show a reasonable  $R_{\text{sym}}$  and completeness to  $2\text{\AA}$  (Table 4.1). The second dataset (AD-LAB2) comprised all the data. Due to disagreements in the scaling of intensities between the two crystals, it was necessary to swap the  $h$  and  $k$  indices of the data collected from the second crystal. Failure to do this resulted in a  $R_{\text{merge}}(I)$  of 11.4% and the rejection of more than 10% of the data. Data from all three crystals were merged using *ROTAVATA / AGROVATA* (CCP4) with the unique  $hkl$  files from *XDS*. Inclusion of the extra data had a minimal effect on the merging  $R$ -factor and also improved the completeness in the low and high resolution shells (Table 4.2).

### 4.2.4.2 mLA Monoclinic III ( $\text{Na}^+$ Form)

Numerous datasets were collected for this crystal form of the recombinant LA. During the data collection process it became apparent that the maximum resolution that could be obtained using the in-house facilities was about  $3\text{\AA}$ . An initial dataset (AD-LAB1), intended for use in structure solution with molecular replacement (see following chapter), was collected from four crystals. Extreme care was taken in crystal mounting as these monoclinic crystals typically grew in closely stacked clusters. The crystal-to-detector distance was 120mm ( $2\theta=10^\circ$ ) and the exposure per frame was 180sec. Although all four crystals diffracted to  $2.9\text{\AA}$ , the higher resolution data was weak and had poor merging statistics. Therefore, only data to  $3.5\text{\AA}$  was used in the initial dataset

(Table 4.3). A subsequent dataset (AD-LAB2) was collected, that had better correlations between high resolution reflections (3–3.5Å shell). This dataset was used in the preliminary refinement of the molecular replacement solution (see Chapter 7).

Table 4.3 — *XSCALE* merging statistics for in-house mLA monoclinic data (AD-LAB1).

Resolution shell (Å)	Total No. Observed Reflections	Total No. Unique Reflections	$R_{sym}$ (%)	Total No. Possible Reflections	Completeness ( $I > 0\sigma$ ) (%)
∞ – 10.0	1203	291	4.8	313	93.0
10.0 – 8.0	1153	274	6.3	290	94.5
8.0 – 6.0	2314	740	8.8	787	94.0
6.0 – 5.0	2380	913	10.2	996	91.7
5.0 – 4.5	1937	783	11.5	855	91.6
4.5 – 4.0	2971	1234	14.8	1355	91.1
4.0 – 3.9	767	320	21.6	358	89.4
3.9 – 3.8	776	345	26.7	394	87.6
3.8 – 3.7	926	397	21.1	453	87.6
3.7 – 3.6	880	397	25.7	460	86.3
3.6 – 3.5	1071	478	26.8	549	87.1
∞ – 3.5	16378	6172	10.4	6810	90.6

#### 4.2.4.3 mLA Hexagonal (*P622*)

The hexagonal recombinant LA crystals diffracted significantly better than those obtained for the wild type protein. Diffraction data were collected to a resolution of 2.9Å from three crystals using a crystal-to-detector distance of 120mm, a  $2\theta$  swing of 10° and an exposure time of 250sec/frame. A total of 528 frames of data were collected from these crystals. As for previous crystal forms, diffraction in the higher resolution shells (2.9–3.1Å) was poor and not included in the final merged dataset. The final dataset had a reasonable  $R_{sym}$  and was essentially complete to 3.2Å (Table 4.4).

Table 4.4 — *XSCALE* merging statistics for mLA hexagonal data.

Resolution shell (Å)	Total No. Observed Reflections	Total No. Unique Reflections	$R_{\text{sym}}$ (%)	Total No. Possible Reflections	Completeness ( $I > 0\sigma$ ) (%)
$\infty - 10.0$	1369	136	5.4	140	97.1
10.0 – 8.0	1257	113	5.7	113	100
8.0 – 6.0	2581	296	7.7	296	100
6.0 – 5.0	2560	355	9.0	355	100
5.0 – 4.0	5448	786	9.6	786	100
4.0 – 3.5	4868	780	13.1	780	100
3.5 – 3.4	1249	210	17.1	210	100
3.4 – 3.3	1354	243	17.1	243	100
3.3 – 3.2	1383	270	19.5	272	99.3
$\infty - 3.2$	<b>22069</b>	<b>3189</b>	<b>9.9</b>	<b>3195</b>	<b>99.8</b>

Table 4.5 — *XSCALE* merging statistics for GOLLA monoclinic data.

Resolution shell (Å)	Total No. Observed Reflections	Total No. Unique Reflections	$R_{\text{sym}}$ (%)	Total No. Possible Reflections	Completeness ( $I > 0\sigma$ ) (%)
$\infty - 10.0$	637	147	4.9	152	96.7
10.0 – 8.0	499	130	5.2	132	98.5
8.0 – 6.0	1282	383	6.2	387	99.0
6.0 – 4.0	4884	1523	7.1	1543	98.7
4.0 – 3.0	8552	2948	11.1	2983	98.8
3.0 – 2.5	8337	3585	18.7	3728	96.2
2.5 – 2.4	1893	960	26.3	1172	81.9
2.4 – 2.3	1151	739	21.4	1357	54.5
2.3 – 2.2	766	638	8.9	1647	38.7
2.2 – 2.1	793	697	8.8	1941	35.9
2.1 – 2.0	617	591	26.0	2366	25.0
$\infty - 2.0$	<b>29429</b>	<b>12359</b>	<b>9.6</b>	<b>18280</b>	<b>67.6</b>

#### 4.2.4.4 GOLLA Monoclinic ( $P2_1$ )

Numerous crystals were examined due to the contact twinning exhibited by the GOLLA monoclinic crystals. In the end, data from four crystals were scaled and merged (Table 4.5). Although diffraction was observed to Bragg spacings of  $2.0\text{\AA}$ , only one of the four crystals chosen for final merging diffracted to this resolution. The other three crystals diffracted to a maximal resolution of  $2.3\text{\AA}$ . The resultant dataset was essentially complete to a resolution of  $2.5\text{\AA}$ . At higher resolution, the completeness falls off sharply so that between 2.1 and  $2.0\text{\AA}$  only a quarter of the possible reflections are represented.

## 4.3 Synchrotron Data Collection

### 4.3.1 Introduction

The use of high energy synchrotron sources in X-ray crystallography has had a considerable impact on the field. Synchrotron radiation is produced when electrons and positrons move at relativistic energies in a circular path. This radiation is generated in synchrotrons, where powerful dipole magnets steer trajectories of charged particles in a circular orbit within the storage ring. These particles emit a continuum of synchrotron radiation whose spectral range depends on the particle's energy and the field strength of the bending magnets. The emitted radiation beam typically covers a wide spectrum, has a high spectral brilliance and is extremely collimated.

Synchrotron radiation has several advantages over conventional X-ray tube sources for use in X-ray diffraction studies. Firstly, its high intensity facilitates data collection from weakly diffracting specimens, such as very small crystals or those with extremely large unit cell dimensions. Secondly, as synchrotron radiation consists of a smooth, continuous spectrum it can be tuned. Short wavelengths are often selected (about 1Å) because at these lower wavelengths, absorption by the air and crystal are reduced. This latter factor, combined with decreased exposure times, results in an appreciable reduction in the radiation damage to the protein crystal (approximately 5-fold). Radiation damage is also decreased as the emitted radiation is pulsed so that the crystal is subjected to short, intense bursts of X-rays. At the Synchrotron Radiation Source (SRS) at the Daresbury Laboratory, 160 circulating electron bunches produce 120ps bursts followed by a 2ns pause. The variable wavelength is also useful for optimising the anomalous scattering signal for particular atoms present in the crystal.

This tunability can now be used in MAD phasing, a powerful new technique which eliminates the need to screen for heavy atom derivatives (Hendrickson, 1991).

### 4.3.2 mLA Monoclinic ( $\text{Mg}^{2+}$ Form)

It was hoped that the shorter wavelength available at the SRS would allow the collection of higher resolution diffraction data for the monoclinic (III) form of mLA to augment the in-house data. As described in section 4.2.4.2, in-house diffraction data could only be collected to a maximum resolution of  $3\text{\AA}$ . This was probably due to the thin nature of the crystals and their extreme susceptibility to radiation damage.

Preliminary tests on Station 9.5 at the SRS showed that the monoclinic crystals diffracted to better than  $2.3\text{\AA}$ . This improvement in resolution was due in part to the improved signal to noise ratio, resulting from the natural collimation and spectral purity of synchrotron radiation, but also to increases in crystal size produced by manipulating the crystallisation conditions (see Chapter 3). Subsequent data collection was mainly hindered by problems associated with 'invisible' contact twinning of the crystals and their susceptibility to radiation damage. The crystal twinning made it impossible to ascertain whether a crystal was affected prior to exposure. As a result, over sixty crystals had to be examined, before a nominally complete dataset could be obtained.

Data was collected on all three protein crystallography stations at the SRS. The station parameters for each trip are summarised in Table 4.6. On the majority of occasions, short wavelengths could be used and this had the effect of increasing the crystal's lifetime in the beam. On Station 7.2, most of the data was not as reliable as that from other stations due to rapid radiation damage caused by the relatively high, fixed

wavelength ( $\lambda=1.488\text{\AA}$ ). Diffraction data was collected from crystals cooled to  $15^\circ\text{C}$  and recorded on MAR Research image plates (30cm on stations 9.5 and 9.6; 18cm on station 7.2). Most of the crystals diffracted to a resolution of  $2.3\text{\AA}$  but the higher resolution

Table 4.6 — Station statistics.

Data Collection Trip	8/9/94	23/10/94	2/11/94	24/11/94	20/12/94
Station	9.6	7.2	9.5	9.6	9.5
Crystal-to-detector distance (mm)	315	131	431	369	455
Wavelength ( $\text{\AA}$ )	0.882	1.488	0.9457	0.87	0.9199
Exposure time <sup>a</sup> (sec)	90	40	90	100-120	60-90
No. crystals used <sup>b</sup>	1	1	1	3	2
Rotation per image (degrees)	2.0	1.5	1.5	1.5	2.5
$I_{\text{max}}$ (mA)	210	280	290	170	170
$\text{ICR}_{\text{max}}$ (V)	0.8	1.2	1.5	2.0	1.4

<sup>a</sup> Two oscillations per image

<sup>b</sup> Number of crystals used refers to the data that were used in the final dataset

reflections were often extremely weak. The program *DENZO* was used to process and reduce the diffraction data. Integrated intensities from different crystals were scaled and merged using the program *SCALEPACK* (Otwinowski, 1993; Gewirth, 1995). The final set of merged and scaled intensities were converted to structure factor amplitudes using *TRUNCATE* (CCP4).

The choice of which data to include in the final merging and scaling was a difficult one. Ideally, the final dataset ought to be a balance between data quality and completeness. In this case, data quality had to be compromised, to some extent, so that a relatively complete dataset could be obtained. A total of ninety-three images (derived from eight crystals) were used in the final merging with *SCALEPACK* (Table 4.7). Eight of these were subsequently rejected due to high merging *R*-factors.

Table 4.7 — Summary of each image used in final merging.

Image No.	No. of obs	$\langle I/\sigma \rangle$	Mosaicity	$R_{\text{merge}}(I)$	Image No.	No. of obs	$\langle I/\sigma \rangle$	Mosaicity	$R_{\text{merge}}(I)$
1	626	8.7	0.331	0.089	48	1267	5.8	0.308	0.095
2	718	8.6	0.352	0.067	49	1226	5.3	0.320	0.108
3	683	7.9	0.343	0.065	50	1164	6.6	0.369	0.100
4	735	8.7	0.328	0.052	51	988	6.1	0.444	0.111
5	684	8.0	0.344	0.060	54	457	6.9	0.563	0.106
6	692	8.1	0.351	0.056	55	568	6.9	0.553	0.081
7	736	7.9	0.281	0.046	56	571	6.6	0.530	0.064
8	745	8.1	0.279	0.046	57	573	7.0	0.567	0.063
9	718	8.0	0.333	0.049	58	563	6.9	0.565	0.057
10	711	7.7	0.322	0.049	59	550	6.5	0.550	0.059
11	622	7.5	0.302	0.048	60	575	7.1	0.527	0.058
12	700	6.0	0.253	0.065	61	547	7.2	0.573	0.053
13	711	6.4	0.317	0.066	62	559	6.9	0.548	0.055
14	777	7.0	0.338	0.052	63	552	6.9	0.500	0.058
15	708	6.8	0.325	0.064	64	568	6.7	0.540	0.068
16	766	6.9	0.367	0.058	65	561	6.7	0.518	0.058
17	748	7.3	0.330	0.044	66	545	6.6	0.470	0.065
18	728	7.0	0.348	0.063	67	567	6.5	0.474	0.070
19	733	7.8	0.342	0.059	68	546	6.1	0.455	0.078
20	734	7.5	0.359	0.071	69	466	6.7	0.462	0.073
21	776	8.1	0.320	0.071	70	470	6.4	0.510	0.094
22	748	7.9	0.321	0.053	71	463	6.0	0.623	0.081
23	745	7.4	0.387	0.081	72	335	5.8	0.758	0.086
24	669	7.6	0.382	0.056	73	327	5.8	0.824	0.080
25	734	7.6	0.333	0.079	74	402	6.0	0.551	0.119
26	761	7.1	0.372	0.092	75	369	6.7	0.612	0.095
27	643	7.7	0.346	0.068	76	470	6.8	0.418	0.112
28	498	6.9	0.565	0.069	77	542	6.1	0.371	0.100
29	673	6.5	0.569	0.093	78	576	6.4	0.308	0.091
30	734	6.6	0.533	0.094	80	517	7.0	0.319	0.087
31	675	6.2	0.517	0.083	81	579	6.7	0.300	0.085
32	730	5.9	0.475	0.089	82	521	6.2	0.315	0.112
33	547	5.6	0.520	0.097	83	549	6.8	0.305	0.086
39	409	7.2	0.457	0.064	84	574	7.3	0.345	0.091
40	536	6.3	0.341	0.088	85	493	6.6	0.343	0.093
41	538	6.1	0.303	0.102	86	546	5.8	0.274	0.089
42	534	5.7	0.350	0.094	87	627	6.1	0.310	0.083
43	463	4.8	0.331	0.112	88	645	5.9	0.260	0.096
44	1126	6.3	0.278	0.080	89	634	5.8	0.331	0.081
45	1250	6.2	0.311	0.093	90	634	5.5	0.307	0.106
46	1206	6.5	0.354	0.085	91	631	5.3	0.253	0.096
47	1225	6.1	0.329	0.084	92	656	5.8	0.230	0.101
					93	540	5.4	0.317	0.085
<b>TOTAL</b>	<b>56008</b>	<b>6.7</b>	<b>0.400</b>	<b>0.078</b>					



Table 4.8 — Summary of merging statistics for the SRS monoclinic data.

Resolution Shell (Å)	Total No. unique reflections	$R_{\text{merge}}(I)$	Completeness ( $I > 0\sigma$ ) (%)
99.00 – 5.71	1282	0.059	79.1
5.71 – 4.53	1408	0.063	89.7
4.53 – 3.96	1420	0.067	90.9
3.96 – 3.60	1442	0.078	92.4
3.60 – 3.34	1430	0.088	92.3
3.34 – 3.14	1436	0.119	92.6
3.14 – 2.99	1410	0.141	91.5
2.99 – 2.86	1288	0.150	83.9
2.86 – 2.75	1176	0.116	75.6
2.75 – 2.65	1142	0.147	74.0
2.65 – 2.57	1142	0.203	74.1
2.57 – 2.49	1115	0.215	72.4
2.49 – 2.43	1053	0.246	69.1
2.43 – 2.37	972	0.268	62.2
2.37 – 2.32	913	0.300	60.8
2.32 – 2.27	525	0.335	34.3
2.27 – 2.22	311	0.294	20.0
2.22 – 2.18	259	0.394	16.7
2.18 – 2.14	273	0.445	17.8
2.14 – 2.10	253	0.444	16.5
<b><math>\infty</math> – 2.10</b>	<b>20250</b>	<b>0.078</b>	<b>65.4</b>

All the data were included and no sigma cutoffs were used. The cell dimensions and crystal orientation were refined for each batch of data but the mosaicity of each image was refined individually. The final mosaicity values varied from 0.25-0.8°. The merging statistics are given in Table 4.8. The completeness of the data decreases gradually but drops off significantly at resolutions higher than 2.3Å. The merging statistics reflect the weak nature of the high resolution reflections and the decreased signal-to-noise ratio. The poor merging statistics may also reflect the effects of radiation damage on the high angle reflections. Attempts to merge the synchrotron data with the low resolution in-house data were unsuccessful due to the differences in unit cell dimensions between the Na<sup>+</sup> and Mg<sup>2+</sup> monoclinic III crystal forms.

### 4.3.3 Summary

Undoubtedly, the monoclinic mLA data is not of the highest quality but it does represent a significant improvement over the 'in-house' datasets in terms of resolution. The data quality primarily reflects the properties of the monoclinic crystals. Crystal twinning was compounded by a short lifetime in the beam. The dramatic fall-off in the diffracted intensities at high Bragg angles is probably entirely due to packing disorder within the monoclinic crystals. This is reflected in the mosaicity values for the data which have an average of  $0.4^\circ$ . There is no doubt that cryocooling would have improved the quality of the final dataset (Petsko, 1975; Garman, 1985). However, the crystals were extremely sensitive and we had no facilities available at Bath to flash freeze the crystals prior to the synchrotron trips.

The considerable amount of time that was invested in collecting high resolution data was justified by the fact that the structure of the monoclinic form had been solved at  $3\text{\AA}$  (see following chapter). Furthermore, at the time the data collection was in progress, this monoclinic form was the most promising in terms of studying the structures of the recombinant and variant LAs. However, in hindsight it was probably not the best choice and more time should have been invested in studying other crystal forms. Nonetheless, the monoclinic mLA structure is the only structure of bovine LA that has been determined to date and has provided a useful insight into properties of the recombinant protein.

## Molecular Replacement

### 5.1 The Molecular Replacement Method

Molecular replacement (MR) can be used to obtain an initial phasing model for an unknown structure. This technique involves finding the orientation of a 'search model', a structure homologous to the unknown, in the unit cell of the unknown crystal that produces the best agreement between the calculated and observed diffraction data. Once this orientation has been found, initial phase estimates can be calculated from the reorientated search model and used for crystallographic refinement. In most cases, a suitable search model can be identified on the basis of sequence identity with the target structure as proteins, homologous in amino acid sequence, tend to adopt very similar tertiary folds. MR has therefore become an extremely useful technique for protein phase angle determination due to the rapid increase in the number of successful protein structure determinations.

In the simplest case, with one molecule in the crystallographic asymmetric unit, three rotational and three translational parameters fully describe the orientation of the search model in the unit cell of the unknown crystal. A full six-dimensional search is

impractical due to the enormous number of trial orientations and positions for the phasing model. However, the search can be reduced to a three-dimensional rotation search ('rotation function') followed by a three-dimensional translation search ('translation function') without overly affecting the end result. A number of different approaches have been proposed for these two searches since the pioneering studies of Rossmann and Blow (Rossmann and Blow, 1962; Rossmann, 1972; Lattman, 1985; Machin, 1985; Dodson *et al.*, 1992).

### 5.1.1 Patterson Function

The cornerstone of the MR method is the Patterson function (Patterson, 1934). The Patterson function  $P(u, v, w)$  is a Fourier summation with intensities as coefficients and with all phase angles equal to zero. The value at a particular point  $\mathbf{u}(u, v, w)$  is calculated from the product of the electron density at position  $\mathbf{x}$  and the electron density at position  $\mathbf{x} + \mathbf{u}$ .

$$\begin{aligned}
 P(uvw) &= \int_V \rho(xyz)\rho(x+u, y+v, z+w)dv \\
 &= \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)] \quad [\text{Eq. 5.1}]
 \end{aligned}$$

where  $u$ ,  $v$  and  $w$  are relative coordinates in the unit cell and  $V$  is the volume of the unit cell. As the Patterson function is the Fourier transform of  $|F(hkl)|^2$  it may always be calculated from a set of experimental diffraction intensities.

In essence, the Patterson function represents a map of the interatomic vectors. These vectors are of two types: relatively short self-Patterson vectors arising from pairs of atoms in the same molecule and longer cross-Patterson vectors arising from pairs of

atoms in different molecules related by the crystal symmetry. Therefore, as the self vectors are independent of the position of the molecule in the unit cell, they can be used to obtain the rotational relationship between the known and unknown structures. In contrast, the intermolecular vector set are dependent on both the molecular structure and its orientation and position in the unit cell. Consequently, if the orientation is known, the intermolecular vectors can be used to determine the position of the molecule relative to the crystal's symmetry elements.

### 5.1.2 Rotation Function

The rotational relationship between the search model and the unknown structure can be calculated by a Patterson overlap function ( $R$ ).

$$R(\alpha, \beta, \gamma) = \int_U P(\mathbf{u}) \times P_r(\mathbf{u}_r) d\mathbf{u} \quad [\text{Eq. 5.2}]$$

where  $P(\mathbf{u})$  is the Patterson function,  $P_r(\mathbf{u}_r)$  is the rotated Patterson and  $U$  is the volume in the Patterson map where the self-Patterson peaks are located. This product function depends on the rotation angles  $(\alpha, \beta, \gamma)$  and will have a maximum value for the correct overlap. The rotated Patterson  $P_r(\mathbf{u}_r)$  can either be of the same crystal lattice (self-rotation function) or a different crystal lattice (cross-rotation function). The self-rotation function is particularly useful for detecting local (non-crystallographic) symmetry within the asymmetric unit. Rossmann and Blow's (1962) original formulation of the rotation function is,

$$R(\alpha, \beta, \gamma) = \frac{U}{V^3} \sum_{\mathbf{h}} \sum_{\mathbf{h}'} |F(\mathbf{h})|^2 |F(\mathbf{h}')|^2 \times \mathbf{G}[-(\mathbf{h} + [\mathbf{C}^{-1}] \mathbf{h}')] \quad [\text{Eq. 5.3}]$$

where  $|F(\mathbf{h})|^2$  is the intensity of the unknown structure,  $|F(\mathbf{h}')|^2$  is the intensity of the known structure,  $C^{-1}$  is the transpose of the rotation matrix  $C$  and  $G$  is the Fourier transform of the shape function  $U$ . For convenience,  $U$  is assumed to be spherical and its size is chosen to maximise the number of self-Patterson vectors used in the calculation and to limit the number of short intermolecular vectors. The size of the shape function is often critical for the success of the rotation function. Short intermolecular Patterson vectors can seriously affect the calculation and their influence is reduced by placing the search model in a large, artificial triclinic unit cell that has no crystallographic symmetry.

Strong reflections tend to dominate the calculation of  $R$  (Eq. 5.3). Normalisation of the structure factors can compensate for the fall off in  $f$  with resolution giving a more even distribution of  $|F|^2$ . This prevents the rotation (and translation) function from being dominated by a few large  $|F|^2$  terms. The normalised structure factor is defined as,

$$E_{\mathbf{h}} = \frac{F_{\mathbf{h}}}{\left(\overline{|F|^2}\right)^{1/2}} \quad [\text{Eq. 5.4}]$$

where  $\left(\overline{|F|^2}\right)^{1/2}$  is the root mean square value of the structure factor amplitudes with  $\sin\theta$  values close to that of  $F_{\mathbf{h}}$ . The effective size of the calculation is reduced as both low and high resolution data can be excluded. Low resolution terms (below 10Å) are relatively insensitive to rotation and are primarily dominated by the solvent. On the other hand, higher resolution reflections (>3.0Å) will differ markedly between homologous structures as they reflect precise conformations of residues. These factors effectively limit the rotation search to a resolution range of about 10 to 3Å. The original Rossmann

and Blow function has subsequently been modified by a number of workers to improve its speed and accuracy (Lattman, 1985 and references therein).

### 5.1.3 Translation Function

Once the known Patterson has been correctly orientated by the rotation function, the relative translation needed to overlap one molecule onto the other in real space must be determined. The translation function is inherently less robust than the rotation function because the only difference between correctly and incorrectly positioned molecules are the intermolecular vectors. The success of the translation function is therefore highly dependent on the accuracy of the solution to the rotation function. The different methods that are used to determine the position of a properly orientated molecule, can be divided into two main groups.

#### 5.1.3.1 Patterson Searches

In this method, a translation function is calculated that gives the correlation between a set of cross-Patterson vectors for the search model and the observed Patterson function of the crystal. The most commonly used translation functions ( $T$ ,  $T1$ ,  $T2$ ) are those derived by Crowther and Blow (1967). These are most easily comprehended in their real space formulation. The basic  $T$  function compares both intra- and intermolecular vectors of the unknown and model structures<sup>1</sup>.

---

<sup>1</sup> Notations used:  $P_c$ , calculated Patterson function for the model structure;  $P_o$ , experimental Patterson function for the unknown structure;  $P_{jk}$ , calculated intermolecular Patterson for the asymmetric units  $j$  and  $k$  of the search model;  $P_{ll}$ , calculated intramolecular Patterson for the  $l$ 'th subunit of the search model;  $\mathbf{t}$ , translation vector for the identity asymmetric unit;  $T_{jk}$ , pairwise translation function for asymmetric units  $j$  and  $k$ ;  $T1_{jk}$ ,  $T$  function with subtraction of intramolecular vectors;  $T2$ ,  $TO$  function with subtraction of intra- and known intermolecular vectors;  $\mathbf{u}$ , vector in Patterson space;  $\mathbf{v}$ ,

$$T_{jk}(\mathbf{v}) = \int_V P_o(\mathbf{u})P_{jk}(\mathbf{u}, \mathbf{v})d\mathbf{u} \quad [\text{Eq. 5.5}]$$

In the  $TI$  function, the intramolecular self-Patterson vectors are calculated for each molecule in the unit cell and removed from the function.

$$TI_{jk}(\mathbf{v}) = \int_V (P_o(\mathbf{u}) - \sum_l P_{ll}(\mathbf{u})) P_{jk}(\mathbf{u}, \mathbf{v}) d\mathbf{u} \quad [\text{Eq. 5.6}]$$

Both the  $T$  and  $TI$  translation functions are two-dimensional and for a complete positional solution, the various two-dimensional translation functions must be combined. This is especially complicated in high symmetry spacegroups and for this reason the  $T$  and  $TI$  translation functions are not widely used. The  $T2$  function, derived originally by Crowther and Blow (1967), is now preferred as it represents a full, three-dimensional function utilising all symmetry operators simultaneously.

$$T2(\mathbf{t}) = \int_V (P_o(\mathbf{u}) - \sum_l P_{ll}(\mathbf{u}))(P_c(\mathbf{u}, \mathbf{t}) - \sum_l P_{ll}(\mathbf{u}))d\mathbf{u} \quad [\text{Eq. 5.7}]$$

Subsequent formulations of the  $T2$  function enable all vector components of the translation to be determined in a single calculation (Tickle, 1985; Rius and Miravittles, 1986). Further modifications have allowed contributions from molecules, related by non-crystallographic symmetry, to be taken into account when searching for additional 'subunits' (Driessen *et al.*, 1991; Tickle, 1992).

### 5.1.3.2 Correlation Searches

In these methods the correctly orientated molecule is systematically translated through the asymmetric unit and a measure of the correctness is calculated for each sample point. Usually structure factors are calculated and compared with the observed

---

intermolecular vector between local origins of asymmetric units  $j$  and  $k$ ;  $V$ , unit cell volume (see Appendix I of Tickle, 1992).



structure factors using an *R*-factor or correlation coefficient as a function of the molecular position. The correlation coefficient tends to be a better indicator than the *R*-factor because it is independent of the relative scale of the observed and calculated data. One disadvantage of these kinds of searches is that they are computationally very expensive as every sample point evaluated requires a loop over all reflections. This is especially relevant when the asymmetric unit contains more than one molecule. The translation function implemented in *X-PLOR* is one of the more commonly used correlation searches (see section 5.2.2).

### 5.2 Software

Two main implementations of the MR method have been used in the current work. These two packages differ significantly and I will briefly describe them before documenting my attempts to solve the structures of the various crystal forms of LA.

#### 5.2.1 Automatic Molecular Replacement (*AMoRe*)

As its name suggests *AMoRe* allows fast, automatic exploration of many potential rotation function solutions (Navaza, 1994; Navaza, 1992). This strategy has been particularly successful especially in cases where the correct orientation was not at the top of rotation function (Saludjian *et al.*, 1992; Lescar *et al.*, 1993). An overview of the *AMoRe* procedure is given in Figure 5.1.

*AMoRe* consists of four main programs: *TABLING*, *ROTING*, *TRAIING* and *FITING*.

*TABLING* produces the arrays of Fourier coefficients of the model densities corresponding to the search molecules. Unlike many MR packages, this is the only step where the

atomic coordinates of the search model are used. Subsequent structure factor calculations are performed by simple interpolation.

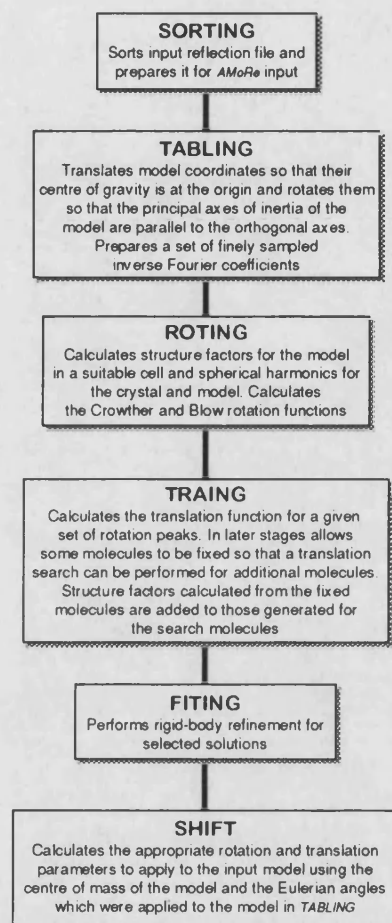


Figure 5.1 — The *AMoRe* procedure

*ROTING* calculates Crowther's fast rotation function but uses numerical integration, rather than Fourier-Bessel expansions in the radial variable, to give more accurate results (Crowther, 1972; Navaza, 1987). However, unlike Crowther's formulation of the fast rotation function, there are no limitations on the maximum radius of integration for a given resolution of data. *ROTING* also uses a novel algorithm that enables enhancement of the rotation peaks by skipping low angular resolution contributions (Navaza, 1990).

**TRAINING** performs the Crowther and Blow Patterson  $T2$  overlap function (Crowther and Blow, 1967; Harada *et al.*, 1981). In cases where there is more than one molecule in the crystallographic asymmetric unit, **TRAINING** allows the position of models already placed to be fixed while secondary translation searches are performed to locate the remaining molecules. Potential solutions are output at each stage and can be monitored by a  $R$ -factor<sup>2</sup>

$$R_f = \sum_{\mathbf{H}} m_{\mathbf{H}} |F(\text{obs})_{\mathbf{H}}| - |F_{\mathbf{H}}(\alpha, \beta, \gamma, x, y, z)| \times \left( \sum_{\mathbf{H}} m_{\mathbf{H}} |F(\text{obs})_{\mathbf{H}}| \right)^{-1} \quad [\text{Eq. 5.8}]$$

and correlation coefficient.

$$C_c = \sum_{\mathbf{H}} m_{\mathbf{H}} \Delta |F(\text{obs})_{\mathbf{H}}| \Delta |F_{\mathbf{H}}(\alpha, \beta, \gamma, x, y, z)| \\ \times \left\{ \left[ \sum_{\mathbf{H}} m_{\mathbf{H}} (\Delta |F(\text{obs})_{\mathbf{H}}|)^2 \right] \times \left[ \sum_{\mathbf{H}} m_{\mathbf{H}} [\Delta |F_{\mathbf{H}}(\alpha, \beta, \gamma, x, y, z)|]^2 \right] \right\}^{-1/2} \quad [\text{Eq. 5.9}]$$

The final stage of the MR procedure is to improve the rotational and translational parameters. **FITING** performs a fast, least-squares rigid-body refinement procedure, first developed by Huber and Schneider (1985), and subsequently modified for use in **AMoRe** (Castellano *et al.*, 1992). During the procedure, optimal overall scale and temperature factors are determined before minimisation of the positional parameters of the search model. **AMoRe** is part of the CCP4 suite of programs (CCP4, 1994).

---

<sup>2</sup> Notation used:  $\mathbf{H}=(h,k,l)$ , reciprocal-space crystal vector;  $F(\text{obs})_{\mathbf{H}}$ , Fourier coefficient of the crystal electron density;  $F_{\mathbf{H}}$ , Fourier coefficient of the model electron density;  $(\alpha,\beta,\gamma)$ , Euler rotation angles;  $(x,y,z)$ , fractional translation coordinates;  $\Delta |F_{\mathbf{H}}|$ , stands for  $|F_{\mathbf{H}}| - \langle |F_{\mathbf{K}}| \rangle_{\mathbf{K}}$ . (see Navaza, 1994)

### 5.2.2 X-PLOR Molecular Replacement

The *X-PLOR* package includes programs to perform rotation and translation functions (Brünger, 1992a). The implemented rotation search employs the real-space Patterson method (Huber, 1985, Brünger, 1992a). Model Patterson vectors, typically calculated in a large orthorhombic box ( $100\text{\AA}^3$ ), are selected according to length and rotated using Eulerian angles  $(\theta_1, \theta_2, \theta_3)$  (Rossmann and Blow, 1962). Product correlations with the crystal Patterson are computed by linear eight-point interpolation. Lattman's pseudo-orthogonal Eulerian angles<sup>3</sup>  $(\theta_+, \theta_-, \theta_2)$  are used to sample possible orientations of the search model (Lattman, 1985). The rotation search is restricted to an asymmetric unit that is dependent on the Patterson symmetry of the probe and unknown (Rao *et al.*, 1980). The rotation peaks are sorted with respect to their product correlation value (*RF*) and then subjected to a cluster analysis. This analysis effectively reduces the number of peaks that need to be tested. Patterson correlation (*PC*) refinement can then be used to filter the rotation peaks with the aim to 'promote' the correct peak to the top of the rotation function output (Brünger, 1990b). In essence, this procedure performs a three-dimensional search to find the optimal coordinates for each of the selected rotation function peaks. A standard linear correlation coefficient (*PC*) is calculated between the squares of the normalised observed and calculated structure factors, as a function of the coordinates of the centre of gravity of the search model,

$$PC(\Omega) = \frac{\langle |E_x|^2 |E_m(\Omega)|^2 - \langle |E_x|^2 \rangle \langle |E_m(\Omega)|^2 \rangle \rangle}{\sqrt{\left( \langle |E_x|^4 - \langle |E_x|^2 \rangle^2 \rangle \langle |E_m(\Omega)|^4 - \langle |E_m(\Omega)|^2 \rangle^2 \rangle \right)}} \quad [\text{Eq. 5.10}]$$

<sup>3</sup> Lattman's Eulerian angles are defined as  $\theta_+ = \theta_1 + \theta_3$ ,  $\theta_- = \theta_1 - \theta_3$ ,  $\theta_2 = \theta_2$ .

where  $E_x$  denotes normalised observed structure factors, and  $E_m(\Omega)$  denotes the normalised structure factors of the search model placed in the unit cell of the crystal in an arbitrary position and orientated according to the rotation matrix ( $\Omega$ ) without considering the crystallographic symmetry mates. The angle brackets denote an averaging over the set of observed reflections expanded to  $P1$ . Refinement is carried out against  $-PC$  and solutions close to the correct orientation should give a relatively large correlation coefficient at the end of the refinement.

This refinement of the MR model prior to the translation function suppresses noise peaks in the original rotation function and ultimately reduces the number of orientations that need to be tested in subsequent translation searches. The  $PC$ -refined model's position in the crystallographic unit cell is then determined by a translation function based on the correlation coefficient of Fujinaga and Read (1987). The equation used is identical to that for  $PC$ -refinement (Eq. 5.10) except that  $E_m$  is replaced with  $E_{\text{calc}}$  (the normalised structure factor of the search model and its symmetry mates) and the averaging operations are performed over an asymmetric unit of diffraction data. The highest peaks of the translation function are output along with a packing value (Brünger, 1992a). Potential MR solutions can be improved and evaluated using rigid-body refinement with the free  $R$ -factor (Brünger, 1992b). This translation function is generally more sensitive than those that make use of Patterson functions but has the drawback of being extremely slow (Brünger and Nilges, 1993).

## 5.3 Results

High resolution, three-dimensional crystal structures have been previously determined for LA (Acharya *et al.*, 1989; Acharya *et al.*, 1991). Both the observed similarity between these two structures and the high level of homology within the LA family suggested that, in the majority of cases, phasing using MR would be fairly straightforward. Unless stated otherwise, the same search model was used in all cases. The MR search model was based on the crystal structure of human LA (Acharya *et al.*, 1991). Sequence differences between the human and bovine protein (31 of the 123 residues) were incorporated into the crystal structure and the altered sidechains were subjected to 10 cycles of real-space refinement using *FRODO* (Jones, 1985). The C-terminal tripeptide (121–123) was omitted from the final model as it is flexible and poorly defined in both the human and baboon crystal structures. The temperature factors for human LA were also included in the model. *B*-factors tend to be closely correlated between related structures and their inclusion in the model should result in the generation of more accurate structure factor amplitudes. In the following sections, the MR strategies used for the various crystal forms of LA are described.

### 5.3.1 Recombinant LA (mLA)

#### 5.3.1.1 Monoclinic III Form

The in-house diffraction data was sufficiently complete within the resolution range typically required for MR (Table 4.3). As there are an estimated four molecules per AU, the first stage in the MR process was to try to identify the non-crystallographic

symmetry (NCS) operators that relate the four molecules. To do this, the crystal Patterson was rotated against itself using a so-called self-rotation function.

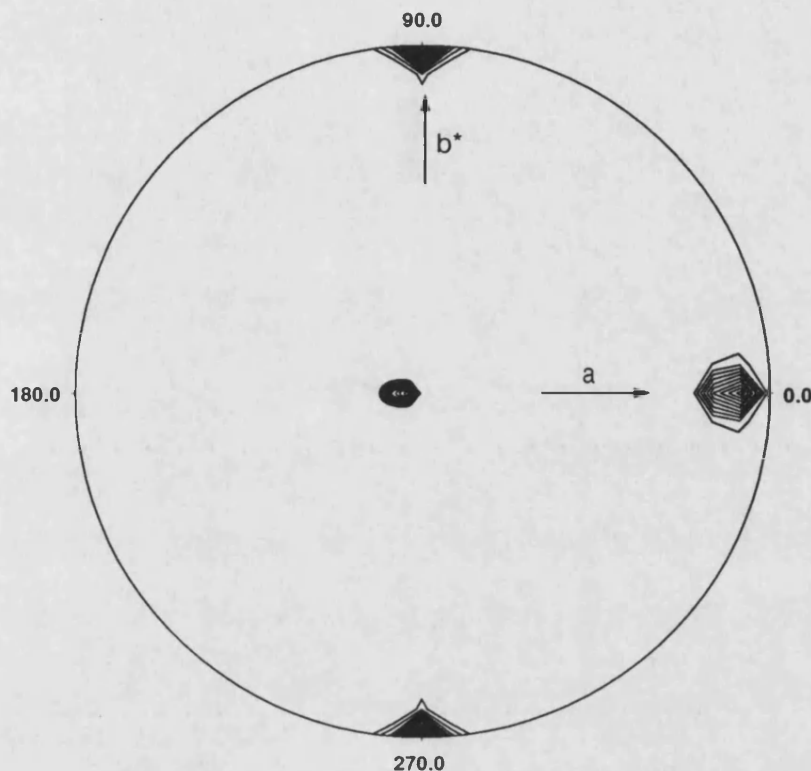


Figure 5.2 — Self-rotation function for monoclinic mLA.

Stereographic projection of the  $\kappa = 180^\circ$  section of the Patterson self-rotation function for monoclinic III data ( $\text{Na}^+$  form). The self rotation function was calculated using data between 8 and  $4\text{\AA}$  (3658 reflections) with an integration radius of  $21\text{\AA}$ . The data was sharpened using a temperature factor of  $-20\text{\AA}^2$ . The stereographic plot is contoured from 30% to 100% (origin peak) in 5% steps. The section has  $\omega=0$  or  $180^\circ$  at the centre,  $\omega=90^\circ$  around the edge and  $\phi$  as marked around the periphery.

#### 5.3.1.1.1 Self-Rotation Function

A Patterson self-rotation search was performed using the program *POLARRFN* written by W. Kabsch (CCP4). Spherical polar angles, rather than Eulerian angles, are used as they clearly indicate the presence of non-crystallographic symmetry. The self-rotation function was carried out by varying  $\phi$  and  $\omega$  in steps of  $5^\circ$  in the range  $\phi=0$  to  $180^\circ$ ,  $\omega=0$  to  $180^\circ$  while  $\kappa$  was held fixed at either  $90^\circ$  (fourfolds) or  $180^\circ$  (twofolds).

Using observed reflections between 8 and 4Å, a very clear peak (77.9% of the origin), representing a strong non-crystallographic two-fold symmetry operator, was observed at  $\phi=0$ ,  $\omega=83.8^\circ$ ,  $\kappa=180.0^\circ$  (and by symmetry  $\phi=0$ ,  $\omega=173.8^\circ$ ,  $\kappa=180.0^\circ$ ) (Figure 5.2). This result suggested that the four molecules in the AU were arranged as a pair of dimers.

Table 5.1 — Top cross-rotation function peaks.

The columns correspond to the peak numbers, orientation Euler angles ( $\alpha, \beta, \gamma$ ) and the correlation coefficient (CC) respectively. The correct orientations are shaded. The mean and sigma of the rotation function map were 0.0 and 3.34 respectively.

Peak	$\alpha$	$\beta$	$\gamma$	CC
1	7.31	92.51	199.55	24.2
2	188.14	106.27	201.93	21.9
3	136.85	70.35	141.18	18.3
4	253.84	100.45	134.40	16.8
5	71.66	96.84	121.05	14.9
6	309.41	14.89	43.41	13.5
7	277.91	29.21	125.10	13.4
8	314.03	59.59	150.68	13.2

### 5.3.1.1.2 Cross-Rotation and Translation Functions

In the cross-rotation function, calculated with *AMoRe* (CCP4 v2.6), using data between 8-4Å and all model Patterson vectors with lengths of between 0 and 21Å, three of the correct orientations could be clearly distinguished (Table 5.1). A translation function (TF), using data between 8 and 4Å, confirmed the correctness of these rotation peaks. Subsequent two-, three- and four-body TF's unambiguously defined the four positions of search molecule within the AU. Rigid-body refinement with *FITING*, using data between 15 and 3.5Å, gave a final *R*-factor of 39.2% and correlation coefficient of 56.6% for the quartet of solutions (Table 5.2). The four correct orientations corresponded to the first, second, third and eighth peak in the rotation function.



Table 5.2 — Summary of translation functions peaks.

The columns correspond to the orientation Euler angles ( $\alpha, \beta, \gamma$ ), the positional parameters  $T_A, T_B, T_C$  (fractions of the unit cell) and the correlation coefficient (CC) and  $R_f$  respectively. For each translation function, the four correct solutions are followed by the highest incorrect solution (separated by a dashed line and in *italics*). In the n-body translation function ( $n=2,3,4$ ) the  $R$ -factor and correlation coefficient are only given for the 'free' peaks. Abbreviations used: TF-n, n-bodied translation function; RBR, rigid-body refinement.

	$\alpha$	$\beta$	$\gamma$	$T_A$	$T_B$	$T_C$	CC	$R_f$
TF-1	7.31	92.51	199.55	0.13816	0.00000	0.09694	18.7	48.5
	188.14	106.27	201.93	0.36184	0.00000	0.07143	18.7	48.4
	136.85	70.35	141.18	0.35526	0.00000	0.42347	11.3	49.9
	314.03	59.59	150.68	0.18421	0.00000	0.05612	12.1	50.1
	<i>309.41</i>	<i>14.89</i>	<i>43.41</i>	<i>0.06579</i>	<i>0.00000</i>	<i>0.15816</i>	<i>10.5</i>	<i>50.3</i>
TF-2	7.31	92.51	199.55	0.13816	0.00000	0.09694	-	-
	188.14	106.27	201.93	0.85783	0.92133	0.56822	33.0	45.0
	136.85	70.35	141.18	0.67328	0.05518	0.02817	26.6	46.7
	314.03	59.59	150.68	0.32199	0.87382	0.58885	23.5	47.3
	<i>253.84</i>	<i>100.45</i>	<i>134.40</i>	<i>0.26601</i>	<i>0.60526</i>	<i>0.99235</i>	<i>18.4</i>	<i>48.6</i>
TF-3	7.31	92.51	199.55	0.13816	0.00000	0.09694	-	-
	188.14	106.27	201.93	0.85783	0.92133	0.56822	-	-
	136.85	70.35	141.18	0.67326	0.05527	0.02719	42.5	42.0
	314.03	59.59	150.68	0.32130	0.87482	0.58839	38.6	43.3
	<i>253.84</i>	<i>100.45</i>	<i>134.40</i>	<i>0.97101</i>	<i>0.17589</i>	<i>0.83280</i>	<i>29.3</i>	<i>46.9</i>
TF-4	7.31	92.51	199.55	0.13816	0.00000	0.09694	-	-
	188.14	106.27	201.93	0.85783	0.92133	0.56822	-	-
	136.85	70.35	141.18	0.67326	0.05527	0.02719	-	-
	314.03	59.59	150.68	0.32242	0.87449	0.58768	47.9	40.7
	<i>253.84</i>	<i>100.45</i>	<i>134.40</i>	<i>0.26694</i>	<i>0.61235</i>	<i>0.99490</i>	<i>37.7</i>	<i>44.0</i>
RBR	8.94	93.24	200.60	0.14005	0.00014	0.09883	56.6	39.2
	188.91	106.72	202.42	0.85742	0.92123	0.57144	56.6	39.2
	136.56	68.62	140.98	0.67546	0.05527	0.02943	56.6	39.2
	315.77	57.45	150.28	0.31643	0.87690	0.58792	56.6	39.2

### 5.3.1.1.3 The Solution

The four rotation and translation solutions were applied to the initial search model and the positions of the resultant molecules were visually examined for close contacts using *FRODO*. The four molecules are arranged as two non-equivalent dimer pairs in the crystallographic asymmetric unit. The same non-crystallographic symmetry operator, running parallel to the  $x$  axis, relates both dimer pairs (Figure 5.3). The four molecules packed together very well and there were no significant clashes between either themselves or their symmetry equivalents. Further indications to correctness of the

solution were given by the presence of large positive peaks of electron density, in the initial  $\sigma_A$ -weighted  $2F_o - F_c$  map (Read, 1986; 1990), in positions corresponding to the calcium binding site of each molecule. In addition, extra density was observed at the N-terminus of each molecule and the density for the sidechain at position 90 was truncated. These are all features of mLA that were not included in the search model. High resolution refinement of this structure is described in Chapter 7.

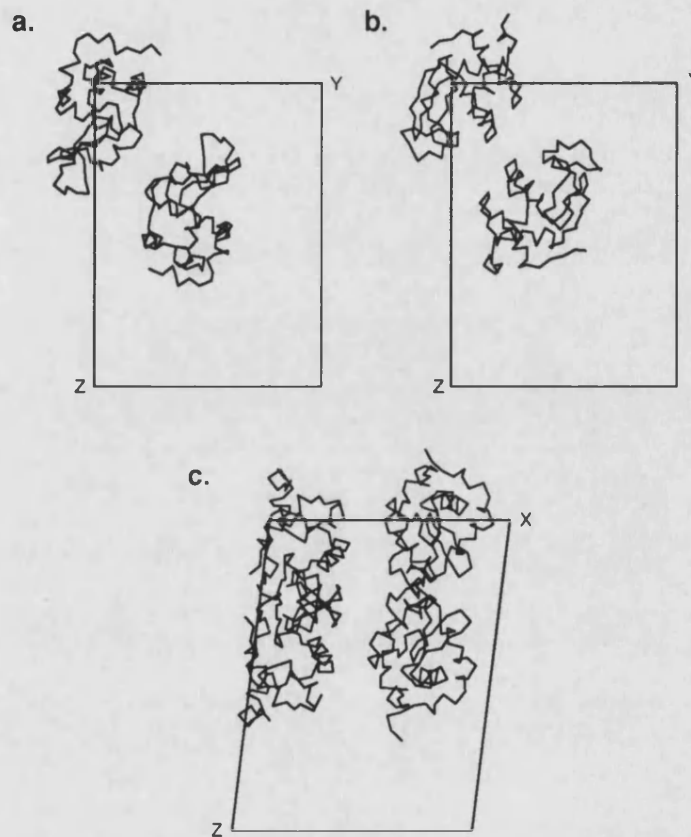


Figure 5.3 — Arrangement of molecules in the monoclinic asymmetric unit  
 a. Molecules A and B and b. molecules C and D viewed down the local two-fold. c. contents of the AU viewed perpendicular to the local symmetry axis. Molecules A and B are located nearest to the z axis.

### 5.3.1.2 Hexagonal Form (A109P-mLA)

Attempts to solve the structure of this form of mLA have been largely unsuccessful. Diffraction data from hexagonal crystals of the variant LA (A109P-mLA) was used for the majority of these studies due to its higher quality ( $R_{sym}=8.8\%$ ) and overall completeness ( $99.4\% \infty 2.9\text{\AA}$  ( $94.3\% 3.0\text{--}2.9\text{\AA}$ )  $I>0\sigma$ ). Nevertheless, the rotation and translation peaks obtained with this data are identical to those observed when using data from either native or mLA hexagonal crystals.

It was not necessary to perform a self-rotation function for this data because the hexagonal crystals only contain one molecule per AU. A series of cross-rotation functions performed over different resolution ranges (between 15 and  $3\text{\AA}$ ) using different model Patterson vectors (between 0 and  $25\text{\AA}$ ) gave a single, consistent solution that was 4.7–5.1 $\sigma$  above the mean and 0.9–1.3 $\sigma$  above the next highest peak. The peak had Euler rotation angles ( $\alpha,\beta,\gamma$ ) of  $1.5\pm 0.5^\circ$ ,  $71.6\pm 0.1^\circ$  and  $138.0\pm 0.2^\circ$  respectively (Table 5.3). A similar peak was obtained when normalised structure factors were used instead of  $F_s$ .

Table 5.3 — Top peaks for a typical rotation function against the hexagonal data.

The columns correspond to the peak numbers, orientation Euler angles ( $\alpha,\beta,\gamma$ ) and the correlation coefficient (CC) respectively. This data is taken from a rotation function using data between 8 and  $4\text{\AA}$  and a Patterson sphere of  $21\text{\AA}$ . The mean and sigma of the rotation function map were 0.0 and 3.59 respectively.

Peak	$\alpha$	$\beta$	$\gamma$	CC
1	0.57	71.54	137.92	17.5
2	34.01	58.75	72.59	13.1
3	35.18	61.46	191.02	11.5
4	18.56	57.03	307.09	10.6
5	27.42	70.34	135.98	10.6
6	38.5	25.88	1.50	10.5
7	35.13	26.91	8.50	10.4
8	2.01	76.40	256.03	10.3

However, in subsequent translation functions, no correct solutions were obtained for any of the top 50 peaks from the rotation function regardless of the resolution range used. In all cases, the correlation coefficients were unexpectedly low after rigid-body refinement ( $CC_{\max}=36\%$ ), despite reasonable  $R$ -factors (43–48%). Visual examination of position of the search model in the unit cell, after applying the calculated rotational and translational elements, revealed that every MR solution located the model on either the two- or six-fold symmetry axes. To ensure that the rotation function gave consistent results, the search model was rotated and used in *ROTING* with a series of different resolution ranges and Patterson spheres. These searches gave a clear, consistent peak ( $\alpha=52^\circ$ ,  $\beta=72^\circ$ ,  $\gamma=217^\circ$ ), that was identical to that obtained with the original model. The failure of the translation function was very puzzling. Translation functions ( $T2$  and  $TO/O$ ), calculated using the rotation peaks obtained in *AMoRe* and the program *TFFC* (CCP4), gave similar results to *TRAINING* with all the peaks located on the crystallographic symmetry axes.

To clarify the situation, the rotation and translation functions were calculated with *X-PLOR*. The rotated search model was placed in a orthorhombic cell (105Å on the edge) and the computed model Patterson vectors were selected according to length. Cross-rotation peaks were systematically identified using a variety of resolution ranges, Patterson vector lengths and  $PC$  refinement strategies. The searches were restricted to an asymmetric unit of  $\theta_+=0 - 720^\circ$ ,  $\theta_2=0 - 90^\circ$ ,  $\theta_-=0 - 60^\circ$  (Rao *et al.*, 1980).  $PC$ -refinement of the overall orientation ( $\Omega$ ) of the molecule consisted of 15 cycles of Powell conjugate gradient energy minimisation. A consistent peak was obtained after  $PC$ -refinement:  $\theta_1=51.5\pm 0.9$   $\theta_2=72.0\pm 0.7$   $\theta_3=25.5\pm 1.1$  (RF value=1.4 $\sigma$ ). Comparison of this peak with the one obtained in *AMoRe* revealed that they were similar and related by

symmetry. Encouraged by the fact that two very different rotation function methods had given the same peak, attempts were made to determine the position of the orientated, *PC*-refined model using the *X-PLOR* translation function.

Translation functions were performed over a series of resolution ranges (15-4Å, 15-3.5Å, 12-6Å, 8-4Å) using a sampling interval in *x*, *y* and *z* set to a third of the high resolution limit. The asymmetric unit used for the translation function was from 0 – 1 on *x* and *y* and 0 – 0.5 on *z*. Potential solutions were first evaluated on packing criteria to eliminate peaks that gave rise to considerable mainchain overlaps with their symmetry mates. Reasonably packed solutions were then analysed by a rigid-body refinement strategy using the free *R*-factor. This involved 30 cycles of rigid-body refinement using data between 15 and 3.5Å resolution (10% flagged for  $R_{\text{free}}$  calculation). Several reasonably packed solutions were identified but initial,  $\sigma_A$ -weighted  $2F_o - F_c$  maps were extremely poor and showed no indications of a bound ion in the calcium binding site.

There are a number of possible reasons why the molecular replacement method can fail to give a satisfactory solution. These include the quality of the experimental data, the suitability of the search model and the symmetry of the target Patterson function. In this case, doubts about the search model are not a major concern as it is highly homologous to the target structure and has been used to solve five different LA structures. The experimental data quality is acceptable ( $R_{\text{sym}}=8.8\%$ ) and virtually all the reflections within the resolution limits required for MR have been observed. Therefore, the failure to find a correct solution is probably due to the high symmetry of the target crystal (*P6/mmm*). High symmetry can give rise to heightened noise levels in both the rotation and translation functions resulting in difficulties in peak identification.

Nevertheless, rotation functions using both *AMoRe* and *X-PLOR* appeared to give a clear and identical peak. It may be possible that the failure of the subsequent translation function was due to a freak crystallographic situation such as the alignment of the Patterson vector set parallel to a crystallographic symmetry element. Whatever the reason, it appears that additional phasing information, such as a single heavy atom derivative, is required to solve this form of LA.

### 5.3.2 Bovine Milk LA

Several attempts have been made to solve the structure of the trigonal (*P321*) form using the both *AMoRe* and *X-PLOR*. Both strategies have suggested potential solutions but, as will be discussed, problems were encountered during their preliminary refinement.

The first ambiguity with this crystal form was the number of molecules present in each crystallographic asymmetric unit. Fenna (1982a) reported that the asymmetric unit contains two molecules based on the unit cell volume and the molecular weight of BOLA. This assumption is perfectly acceptable and gives a theoretical solvent content (59%) which is around the upper range commonly found for protein crystals (Matthews, 1968). However, initial attempts to solve the structure with *AMoRe*, consistently suggested that there may be three molecules per asymmetric unit. Self rotation functions with *POLARRFN* (CCP4) did not however show any non-crystallographic relationship within the asymmetric unit. This was unexpected as the diffraction pattern clearly indicates the presence of a non-crystallographic 2-fold axis (Fenna, 1982a).

Table 5.4 — Top rotation function peaks.

The columns correspond to the peak numbers, orientation Euler angles ( $\alpha, \beta, \gamma$ ) and the correlation coefficient (CC) respectively. The potential 'correct' orientations are shaded. This data is taken from a rotation function using data between 8 and 4 Å and a Patterson sphere of between 4 and 21 Å. The mean and sigma of the rotation function map were 0.03 and 3.07 respectively.

Peak	$\alpha$	$\beta$	$\gamma$	CC
1	0.65	108.34	316.86	16.6
2	70.85	108.89	315.68	15.0
3	53.51	110.06	315.90	14.2
4	96.41	57.44	71.43	12.0
5	83.02	154.48	182.51	11.9
6	30.30	58.24	72.48	11.8
7	18.50	109.72	313.96	11.7
8	91.83	109.88	315.01	11.7
9	101.93	61.54	190.83	10.8
10	86.34	119.05	13.02	10.2

Table 5.5 — Summary of translation functions peaks.

The columns correspond to the orientation Euler angles ( $\alpha, \beta, \gamma$ ), the positional parameters  $T_A, T_B, T_C$  (fractions of the unit cell) and the correlation coefficient (CC) and  $R$ -factor (Rf) respectively. For each translation function, the three potential solutions are followed by the two next highest peaks (separated by a dashed line and in *italics*). In the  $n$ -body translation function ( $n=2,3$ ) the  $R$ -factor and correlation coefficient are only given for the 'free' peaks (these values are meaningless for the FIXed peaks). Abbreviations used: TF- $n$ ,  $n$ -bodied translation function; RBR, rigid-body refinement.

	$\alpha$	$\beta$	$\gamma$	$T_A$	$T_B$	$T_C$	CC	Rf
TF-1	0.65	108.34	316.86	0.17640	0.02857	0.24280	19.8	48.1
	70.85	108.89	315.68	0.97985	0.93021	0.47922	13.2	50.4
	53.51	110.06	315.90	0.02866	0.02776	0.43952	16.6	49.8
	<i>18.50</i>	<i>109.72</i>	<i>313.96</i>	<i>0.02060</i>	<i>0.12677</i>	<i>0.38418</i>	<i>13.1</i>	<i>49.4</i>
	<i>83.02</i>	<i>154.48</i>	<i>182.51</i>	<i>0.07976</i>	<i>0.08160</i>	<i>0.37679</i>	<i>14.6</i>	<i>51.0</i>
TF-2	0.65	108.34	316.86	0.17640	0.02857	0.24280	-	-
	70.85	108.89	315.68	0.79664	0.51787	0.90695	24.8	47.5
	53.51	110.06	315.90	0.49623	0.83291	0.57211	23.7	47.3
	<i>91.83</i>	<i>109.88</i>	<i>315.01</i>	<i>0.99538</i>	<i>0.98421</i>	<i>0.12000</i>	<i>20.5</i>	<i>48.4</i>
	<i>18.50</i>	<i>109.72</i>	<i>313.96</i>	<i>0.16555</i>	<i>0.02698</i>	<i>0.24434</i>	<i>20.4</i>	<i>48.9</i>
TF-3	0.65	108.34	316.86	0.17640	0.02857	0.24280	-	-
	70.85	108.89	315.68	0.79664	0.51787	0.90695	-	-
	53.51	110.06	315.90	0.49727	0.83249	0.57170	32.2	46.1
	<i>18.50</i>	<i>109.72</i>	<i>313.96</i>	<i>0.16107</i>	<i>0.02690</i>	<i>0.24191</i>	<i>26.0</i>	<i>47.7</i>
	<i>83.02</i>	<i>154.48</i>	<i>182.51</i>	<i>0.01382</i>	<i>0.95996</i>	<i>0.73759</i>	<i>23.9</i>	<i>48.1</i>
RBR	1.77	107.57	316.92	0.17437	0.02626	0.23794	-	-
	69.91	107.83	315.78	0.79405	0.51812	0.90945	-	-
	55.21	108.04	317.75	0.49840	0.82951	0.56981	46.3	42.8
	<i>17.53</i>	<i>110.11</i>	<i>316.78</i>	<i>0.15544</i>	<i>0.03180</i>	<i>0.23986</i>	<i>39.8</i>	<i>45.1</i>
	<i>96.45</i>	<i>55.64</i>	<i>70.17</i>	<i>0.08782</i>	<i>0.89404</i>	<i>0.94715</i>	<i>34.1</i>	<i>46.6</i>

### 5.3.2.1 *AMoRe*

Cross rotation functions using a variety of different resolution shells, Patterson spheres and normalised structure factors ( $E$ 's) as well as  $F$ 's consistently gave similar results. All the searches resulted in three rotation peaks that were distinct from the rest (Table 5.4). The same results were obtained using either the AD-LAB1 or AD-LAB2 datasets. Subsequent translation searches (8-4Å) suggested a good orientation for three molecules in the asymmetric unit. A three-body rigid-body refinement, using data between 15 and 3.5Å, yielded a solution with a correlation coefficient 12% higher and an  $R$ -factor approximately 4% lower than the next best peak (Table 5.5).

The molecular packing of this solution was encouraging and showed no major clashes between the three molecules and their symmetry mates. In contrast, packing analysis of the other potential 'solutions' from *TRAINING* showed major mainchain overlaps. Although conventional positional refinement, simulated annealing and grouped  $B$ -factor refinement reduced the conventional  $R$ -factor from 44.0 to 25.9% (using data between 10 and 2.5Å with NCS restraints), the free  $R$ -factor showed no significant improvement ( $R_{\text{free}}^{\text{in}}=44.5$ ,  $R_{\text{free}}^{\text{out}}=44.1$ )<sup>4</sup>. Further refinement confirmed that something was wrong with the *AMoRe* solution. Cycles of refinement and rebuilding resulted in a stagnation of both the conventional and free  $R$ -factors (about 24% and 40% respectively). The electron density maps were unexpectedly poor and there was also a poor correlation of the  $B$ -values for bonded atoms.

---

<sup>4</sup> See Chapter 7 for description of crystallographic refinement using *X-PLOR*. The free  $R$ -factor was calculated against a random 10% of the data (1156 reflections).



Table 5.6 — Evaluation of potential solutions using the free  $R$ -factor.

The top solutions after a two-body translation function (with Peak 1 and 2's positions fixed) are listed. The conventional  $R$ -factor and free  $R$ -factor (10% data) are given for each of the peaks after 35 cycles of rigid-body refinement in *X-PLOR* (using 15-3.5Å data, 3 separate bodies). The  $R$  and  $R_{\text{free}}$  for peaks 1 and 2 alone are 47.5 and 47.0 respectively.

Peak	$\alpha$	$\beta$	$\gamma$	$T_A$	$T_B$	$T_C$	$R$	$R_{\text{free}}$
1	0.26	107.58	317.65	0.17623	0.02961	0.24242	—	—
2	70.92	107.22	316.16	0.79868	0.51721	0.90884	—	—
3	55.05	108.51	317.61	0.49613	0.83239	0.57342	45.0	46.3
4	95.97	58.70	72.30	0.83536	0.04848	0.47640	48.1	47.1
5	89.01	120.82	252.24	0.02693	0.15387	0.02686	48.1	47.1
6	85.21	120.60	12.52	0.11916	0.98804	0.53134	48.1	49.9
7	101.65	59.00	191.19	0.08164	0.15596	0.44166	48.4	48.0
8	20.23	107.88	315.99	0.30405	0.70783	0.03478	47.7	53.2
9	85.56	60.57	193.24	0.83969	0.78453	0.47048	48.6	50.4
10	66.50	44.27	18.83	0.09552	0.90519	0.90285	48.2	50.0
11	94.93	121.44	8.19	0.87682	0.98493	0.26228	48.7	48.3
12	87.19	150.90	167.07	0.99468	0.07863	0.14346	49.1	48.4
13	15.60	57.77	308.88	0.30665	0.09602	0.86811	48.3	46.9

Table 5.7 — Top rotation function peaks after PC refinement.

Model Patterson were computed in an orthorhombic box of 120x120x120Å. The rotation search was restricted to an asymmetric unit of  $\theta_1=0-720$ ,  $\theta_2=0-120$ ,  $\theta_3=0-90$ . PC-refinement consisted of 10 cycles of Powell energy minimisation.

Rotation Function		PC Refinement				
Resolution Range (Å)	Maximum Patterson vector (Å)	Resolution Range (Å)	Top Rotation Function Peaks			RF value ( $\sigma$ )
			$\theta_1$	$\theta_2$	$\theta_3$	
20-4	30	20-4	358.28	74.42	316.29	0.95
			356.22	73.84	267.63	1.03
			357.04	75.76	285.78	0.99
			353.26	76.48	336.11	0.99
10-4	30	10-4	358.51	74.30	315.68	1.40
			356.31	73.58	267.70	1.43
			357.33	75.74	285.60	1.36
			352.40	75.22	336.05	1.38
8-4	25	8-4	358.47	74.50	315.24	1.65
			355.05	73.93	267.82	1.64
			352.36	73.95	336.27	1.61
			357.18	76.06	285.34	1.56
8-4	21	8-4	358.47	74.50	315.24	2.45
			356.47	74.77	265.03	2.42
			352.47	74.31	336.37	2.45
			357.23	76.24	285.16	2.34

Despite numerous different approaches, such as fixing the positions of different solutions after the first translation function, the set of three peaks highlighted in Table

5.4 always gave the best statistics and free  $R$ -factor values after rigid-body refinement (Table 5.6).

### 5.3.2.2 *X-PLOR*

To clarify the situation, the MR process was repeated using the real-space methods implemented as part of the *X-PLOR* package of programs. Cross-rotation peaks were systematically identified using a variety of resolution ranges, Patterson vector lengths and *PC* refinement strategies. Several different orientations were identified and four of these consistently appeared regardless of the resolution range used (Table 5.7).

An initial translation function (15-3.5Å) for orientation A (356°, 74°, 267°) identified a clear peak at  $t_x^A=0.14$ ,  $t_y^A=0.022$  and  $t_z^A=0.424$  in fractional coordinates that was  $7.1\sigma$  above mean and  $1.4\sigma$  above the next highest peak. Subsequent translation functions (15-4Å, 8-4Å) for orientation B (358°, 74°, 316°), using partial contributions from the correctly-positioned molecule A, yielded the position of the second molecule ( $t_x^B=0.699$ ,  $t_y^B=0.237$  and  $t_z^B=0.098$ ). Attempts to locate the position of a third molecule, using the other rotation function orientations, were unsuccessful. Rigid-body refinement (15-3.5Å) of the two correctly positioned molecules improved both the conventional and free  $R$ -factors by about 1.5% ( $R_{out}$ : 48.2,  $R_{free}^{out}$ : 49.8). However, subsequent positional refinement at 2Å failed to improve the free  $R$ . At this stage a  $\sigma_A$ -weighted  $2F_o-F_c$  electron density map was calculated from the positions of the two molecules in the AU. Visual inspection of the map showed a large area of weak density that was unaccounted for and indicated that an additional molecule may be present.

The positions of the two molecules determined using *X-PLOR* were identical to those obtained with *AMoRe* (peaks 1 and 3 in Table 5.4) differing only by an origin shift of

$z+1/2$ . More surprisingly, the third molecule from the *AMoRe*'s best set of solutions could account remarkably well for the additional density seen in the  $\sigma_A$ -weighted  $2F_o-F_c$  map after the appropriate superpositions (using the two common molecules) had been performed. This independent verification that the three molecular orientations found using the *AMoRe* procedure were likely to be correct, was very encouraging. However, it provided no explanation as to why this 'correct' MR solution could not be refined.

As described in section 5.3.1.2 failure of the MR method in this case is most probably caused by the high symmetry of the target Patterson function (*P3m*). Although there was also some doubt as to the number of molecules in each crystallographic asymmetric unit, the results from the rotation searches clearly indicate the presence of three molecules. The strength of diffraction observed with the trigonal crystals, compared to other forms of LA, is also consistent with a low solvent content. The inability to refine the *AMoRe* three-molecule solution suggested that there was a major problem. However, independent identification of the same solution with *X-PLOR* MR was greeted by surprise and consternation. Clearly there must be some degree of truth in the solution. Although the  $R_{\text{free}}$  does not decrease significantly on refinement, a value of 40–44% (8–2Å) seems remarkably low. These observations suggest that perhaps only one of the six rotational and translational parameters may be in error. Translational errors of this kind are not uncommon and can result in the correct phasing of a considerable number of reflections. If initially undetected, such errors typically manifest themselves in the later stages of refinement in the form of poor quality electron density maps and uncorrelated temperature factors for bonded atoms (Derewenda *et al.*, 1990; Dodson, 1992). An unambiguous solution to the problem would be to obtain suitable, additional phasing

information from a single heavy atom derivative and perform a phased translation function.

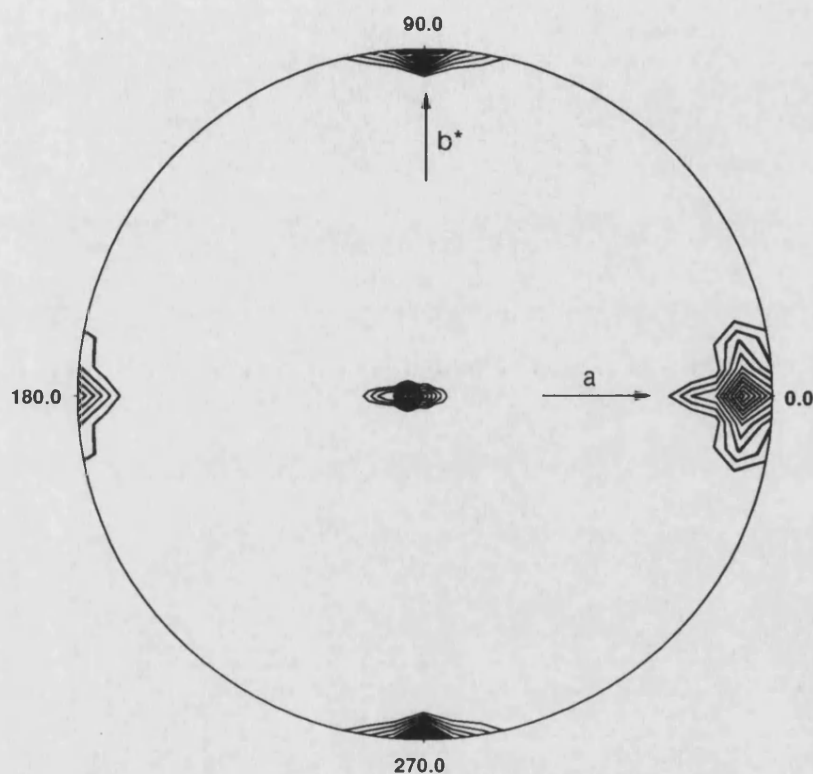


Figure 5.4 — Self-rotation function for GOLA monoclinic crystals.

A stereographic projection of the  $\kappa = 180^\circ$  section of the Patterson self-rotation function for monoclinic data of GOLA. The self-rotation function was calculated using all data between 8 and 4 Å with an integration radius of 21 Å. The data was sharpened with a temperature factor of  $-20 \text{ \AA}^2$ . The plot is contoured from 30% to 100% (origin peak) in steps of 5%.

### 5.3.3 Goat Milk LA

#### 5.3.3.1 Self-Rotation Function

The GOLA monoclinic crystals contain two molecules per asymmetric unit based on theoretical calculations. A self-rotation search was performed using the program

*POLARRFN* (CCP4) to locate the non-crystallographic symmetry operator that relates the two molecules. The self-rotation function was carried out by varying  $\phi$  and  $\omega$  in steps of  $5^\circ$  in the range  $\phi=0$  to  $180^\circ$ ,  $\omega=0$  to  $180^\circ$  while  $\kappa$  was held fixed at  $180^\circ$ . Using observed diffraction data between 8 and  $4\text{\AA}$ , a clear peak (81.6% of the origin), representing a strong non-crystallographic twofold symmetry operator, is seen at  $\phi=0$ ,  $\omega=85.5^\circ$  (and by symmetry  $\phi=0$ ,  $\omega=175.5^\circ$ ) (Figure 5.4). This peak was strikingly similar to that observed in the self-rotation function for the monoclinic mLA (Figure 5.2) and suggested that the direction of the non-crystallographic axis was essentially the same in both cases.

#### 5.3.3.2 Cross-Rotation and Translation Functions

A search model was constructed for GOLLA as described in Section 5.3. The final model consisted of all GOLLA's amino acid sequence apart from the C-terminal tripeptide which, for reasons discussed previously, was omitted. In the cross rotation search data between 8 and  $4\text{\AA}$  were used and model Patterson vectors were selected according to length (between  $21\text{\AA}$  and  $4\text{\AA}$ ). The two correct orientations were easily identifiable at the top of the rotation function output. The two solutions had correlation coefficients approximately 6% higher than the next best solution (Table 5.8). The positions of the molecules within the Cheshire cell were determined by an one-body TF using data between 8 and  $4\text{\AA}$ . A subsequent two-body TF, with the position of one of the molecules kept fixed, allowed the relative positions between the two molecules to be determined.

Table 5.8 — Summary of *AMoRe* solution for GOLA.

The columns correspond to the orientation Euler angles ( $\alpha, \beta, \gamma$ ), the positional parameters  $T_A, T_B, T_C$  (fractions of the unit cell) and the correlation coefficient (CC) and *R*-factor (Rf) respectively. In the two-body translation function the *R*-factor and correlation coefficient are only given for the 'free' peak. Abbreviations used: RF, cross-rotation function; TF-1, translation function; TF-2, two-body translation function; RBR, rigid-body refinement.

	$\alpha$	$\beta$	$\gamma$	$T_A$	$T_B$	$T_C$	CC	Rf
RF	262.34	84.62	15.52	—	—	—	21.1	—
	100.29	96.96	186.23	—	—	—	20.3	—
TF-1	262.34	84.62	15.52	0.33333	0.00000	0.17308	22.5	52.5
	100.29	96.96	186.23	0.30556	0.00000	0.30769	22.5	53.2
TF-2	262.34	84.62	15.52	0.33333	0.00000	0.17308	—	—
	100.29	96.96	186.23	0.30600	0.25796	0.30043	42.6	46.2
RBR	262.48	84.85	14.76	0.33264	-0.00034	0.17661	54.1	44.2
	100.50	95.97	185.74	0.30478	0.25840	0.29834	54.1	44.2

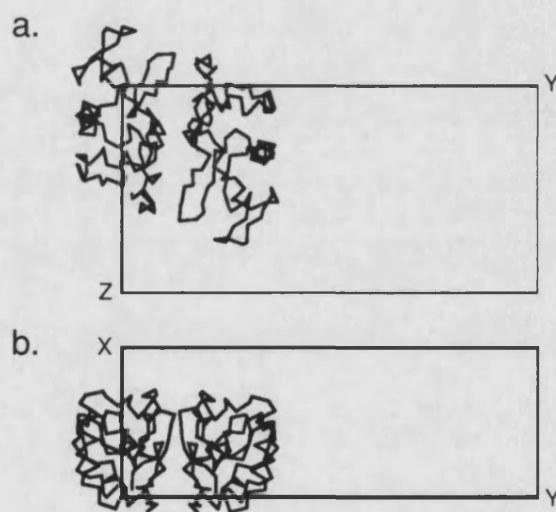


Figure 5.5 — Arrangement of the GOLA molecules in the asymmetric unit.

The two molecules are shown viewed **a.** along the local two-fold and **b.** perpendicular to the local symmetry axis.

### 5.3.3.3 The Solution

The two MR solutions were applied to the initial search model and the positions of the resultant molecules were visually examined for bad contacts using *O*. The

correctness of the solution was confirmed by the presence of a large positive peak of electron density, in an initial  $\sigma_A$ -weighted  $2F_o - F_c$  map, in a position corresponding to the calcium binding site of each monomer.

The two molecules are arranged as a dimer in the asymmetric unit as predicted by the self-rotation function (Figure 5.5). The non-crystallographic axis, that relates the two molecules, is oriented parallel to the  $x$  axis. Although the direction of this local two-fold axis is the same in both the GOLA and mLA monoclinic crystals, the relative orientation of LA molecules in the dimers is different. Refinement of this structure at 2.3Å resolution is described in Chapter 7.

## 5.4 Summary

As expected, obtaining a MR solution for some of the LA structures, the GOLA and mLA monoclinic forms, was relatively straightforward. However, the problems encountered with the hexagonal and trigonal forms were not anticipated. These difficulties primarily reflect the high symmetry of the target Pattersons. These forms are also unusual for a number of other reasons. Firstly, the two forms are obtained under identical growth conditions. They do not however appear in the same drop and the environmental factors that favour a particular form are not clear. Secondly, both the hexagonal and trigonal crystals have virtually identical unit cell dimensions (see Tables 3.1 and 3.2). This similarity is also highlighted in the MR studies where the trigonal and hexagonal forms share a common rotation function peak ( $P321$ :  $0.65^\circ$ ,  $108.3^\circ$ ,  $316.92^\circ$ ;  $P622$ :  $0.57^\circ$ ,  $71.54^\circ$ ,  $137.92^\circ$ ). This peak may be an artefact or ghost peak but raises the intriguing possibility that the internal packing within the two forms is very similar. A

similar conclusion was reached by Fenna based on the similarity in intensity distribution in certain zones of the diffraction patterns of the trigonal and hexagonal crystals (Fenna, 1982a).

Clearly, the difficulties encountered with the MR procedure can probably only be overcome by MIR methods. In both cases, a good single derivative should enable the molecules to be correctly positioned in the AU. Certainly, the observed similarity between the hexagonal and trigonal forms, combined with the information derived from the rotation function, should facilitate the identification of heavy atom sites in any potential derivatives.



## The Quest for Derivatives

### 6.1 Introduction

A preliminary screening of potential heavy atom derivatives has been carried out for the mLA hexagonal crystals due to the failure of the molecular replacement method. As described in the previous chapter, a clear peak was obtained in the rotation function but subsequent translation functions failed to yield a suitable positional solution. Additional phasing information from a reasonably good, single derivative should facilitate the determination of the position of the molecule in the unit cell through a phased translation function (Bentley, 1992).

The isomorphous replacement method can be used to estimate initial protein phase angles. Ideally, a small number of additional 'heavy' atoms must be incorporated at distinct sites in all the unit cells of the native crystal. These bound atoms will give rise to changes in the relative reflection intensities that, when compared to the native reflection intensities, can be used to determine their positions. Location of the bound atoms is the starting point for estimating the protein phase angles. A full treatise on the determination

of protein phases from heavy atom derivatives is given elsewhere (Blundell and Johnson, 1976; Drenth, 1994).

Suitable heavy atom derivatives must yield measurable changes in at least a modest number of reflection intensities. These intensity differences must also be large enough to measure accurately. In addition, the bound heavy atoms must not disturb either the packing or the conformation of the protein to any great extent. In other words, the native and derivative crystals must be relatively isomorphous so that the observed intensity differences are solely due to the attached heavy atoms. Although, perfect isomorphism between the native and derivative crystals is rare, slight changes in the protein's structure can often be tolerated. The presence of non-isomorphism is often signalled by a change in the unit cell dimensions. Theoretical calculations have shown that a 0.5% change in all unit cell dimensions will give rise to about a 15% change in intensity in the 3Å shell (Crick and Magdoff, 1956). Consequently, more substantial changes in cell parameters will seriously limit the effective phasing power of the derivative at high resolution.

The preparation of heavy atom derivatives is essentially an empirical method but the choice of reagents can be directed by the protein's chemical and physico-chemical characteristics. Information about the protein's amino acid sequence, its biological substrates and cofactors and the composition of the crystallisation medium can be useful when selecting heavy atom reagents. Further details on the preparation of heavy atom derivatives are given in several excellent texts (Chapter 8 in Blundell and Johnson, 1976; Petsko, 1985; Leslie, 1991).

## 6.2 Experimental Methods

Hexagonal crystals (*P622*) were grown for the recombinant LA as described in Chapter 3. The CNBr-treated material was used as bigger and more ordered crystals could be grown after the removal of additional N-terminal methionine. Heavy atom compounds were dissolved in a solution consisting of 0.1M PIPES pH 6.5, 1.9M ammonium sulphate and 10mM calcium chloride at the desired concentration. Soaking studies involved transferring native crystals into sitting drops containing 15-20 $\mu$ l of heavy atom solution. Although the hexagonal crystals can only be grown around 37°C, they are quite stable at room temperature. The soaking studies were therefore carried out at between 18-20°C.

Table 6.1 — Station Statistics

Station	7.2	9.5
Wavelength (Å)	1.488	1.0
Distance (cm)	199.8	490.0
Resolution at edge (Å)	3.5	3.4
Oscillation per image (°)	2.5	2.5
Recording device	18cm MAR	30cm MAR
Exposure (sec)	60	70
Derivative datasets collected	8	7

After the desired soaking time had elapsed, the crystals were mounted in thin-walled quartz capillaries and the majority of the heavy atom solution was removed. Medium resolution diffraction data were collected at the SRS, Daresbury Laboratory on stations 7.2 and 9.5. Various station parameters including crystal-to-detector distance, resolution at the edge and oscillation steps are given in Table 6.1.

*DENZO* and *SCALEPACK* were used to process and reduce the diffraction data (Otwinowski, 1993). Structure factor amplitudes were calculated in *TRUNCATE* (CCP4) and the programs *LOCAL* and *SCALEIT* (CCP4) were used to obtain an initial estimate of the quality of each potential derivative ( $\Delta_{\text{iso}}$ ,  $R_{\text{deriv}}$ ). Subsequent analysis was carried out using the *PHASES* package of programs (Furey and Swaminathan, 1990).

In addition to the soaking trials, attempts were made to replace LA's tightly-bound calcium ion with a heavier atom prior to crystallisation. A calcium-free (apo-) form of bovine milk LA was prepared using the acid displacement method (Kronman and Bratcher, 1983). A sample of LA was run down a G-15 Superdex column that had been equilibrated with 10mM HCl (pH 2). The eluted apo-LA was dialysed against distilled water (metal-free) and lyophilised. For the crystallisation trials, the apo-LA was dissolved in 0.1M PIPES pH 6.5, containing the desired concentration (2.5 and 5mM) of heavy metal, and equilibrated for 15mins. The final concentration of protein was 20mg/ml. Crystallisation trials were then carried out as described in Chapter 3 except that no  $\text{CaCl}_2$  was added to the crystallisation medium. The rare earth metals lanthanum ( $\text{La}(\text{Ac})_3$ ), samarium ( $\text{SmCl}_3$ ) and gadolinium ( $\text{Gd}(\text{NO}_3)_3$ ) were chosen to replace the calcium due to their similar atomic radii (Drenth, 1991).

Table 6.2 — Derivative Statistics

Compound	Conc (mM)	Soak (hrs)	$R_{\text{sym}}(I)$	Nref	Comp <sup>b</sup> (%)	$\Delta_{\text{iso}}^c$ (%)	$R_{\text{deriv}}^d$ (%)	Cell Dimensions (Å)		
								<i>a</i>	<i>b</i>	<i>c</i>
Native	—	—	6.0	2589	97.8	—	—	93.68	93.68	67.04
Gd(NO <sub>3</sub> ) <sub>3</sub>	5.0	1.5	5.4	2246 <sup>a</sup>	68.7	13.5	5.9	93.78	93.78	67.24
	2.5	24	5.5	2309 <sup>a</sup>	88.6	21.2	9.6	93.81	93.81	67.23
	1.0	24	7.3	1981	74.7	24.2	10.2	93.66	93.66	67.14
La(Ac) <sub>3</sub>	2.5	24	8.7	2311 <sup>a</sup>	92.8	25.8	11.3	93.22	93.22	66.44
	1.0	24	7.0	2468	92.6	12.6	5.1	93.66	93.66	67.17
K <sub>2</sub> PtCl <sub>4</sub>	5.0	1.5	10.2	2414 <sup>a</sup>	95.0	25.8	10.3	93.75	93.75	67.36
	2.5	24	12.2	2466 <sup>a</sup>	96.7	22.9	8.2	93.88	93.88	67.50
	2.5	384	9.8	1593	60.1	27.1	11.6	93.40	93.40	67.14
	1.0	24	5.4	2296	86.2	28.4	7.4	93.60	93.60	67.28
UO <sub>2</sub> (NO <sub>3</sub> ) <sub>2</sub>	2.5	24	7.5	1740 <sup>a</sup>	68.7	22.8	7.9	93.80	93.80	67.19
pCMBS	1.25	24	4.2	1837 <sup>a</sup>	72.3	8.7	3.6	93.89	93.89	67.34
Na <sub>2</sub> IrCl <sub>6</sub>	1.0	24	19.5	1353	51.0	27.4	10.0	93.40	93.40	66.70
KI	2.0	24	4.7	2506	94.7	11.5	4.5	93.47	93.47	66.90
K <sub>2</sub> HgI <sub>4</sub>	1.0	24	3.9	2209	82.9	9.4	3.7	93.70	93.70	67.17

<sup>a</sup> Data collected to 3.5Å. In all other cases diffraction data was collected to 3.4Å.

<sup>b</sup> Completeness  $\infty$  – Resmax ( $I > 0\sigma$ )

<sup>c</sup> RMS fractional isomorphous difference ( $\Delta_{\text{iso}}$ ) calculated by LOCAL

<sup>d</sup>  $R_{\text{deriv}}$  calculated by SCALE/IT. Defined as  $\sum_{hkl} |F_{\text{PH}} - F_{\text{P}}| / \sum_{hkl} |F_{\text{P}}|$

### 6.3 Results

Attempts to obtain derivatives using the calcium replacement strategy were relatively unsuccessful. Although cocrystals were obtained for  $\text{Sm}^{3+}$  and  $\text{Gd}^{3+}$ -treated apo-LA, initial diffraction analysis showed that they were non-isomorphous with the native hexagonal crystals and had extremely large cell dimensions ( $b \approx 200 \text{ \AA}$ ). Further analysis showed that these crystals are similar to a monoclinic form that is sometimes observed at  $37^\circ\text{C}$  (see Table 3.1).

In the soaking studies, a number of different heavy metal compounds were evaluated. Several could not be dissolved in the mother liquor due to the high concentration of ammonium sulphate (e.g.  $\text{NaAuCl}_4$ ). A summary of the derivative statistics is given in Table 6.2. In general, the crystals were fairly tolerant to a wide variety of compounds. Crystals cracked in only two cases: in the presence of pCMBS at concentrations higher than 1.25mM and at all concentrations of  $\text{SmCl}_3$  tested. Heavy metal soakings caused no appreciable changes in the hexagonal cell (Table 6.2). The maximum change in cell dimensions was observed for a  $\text{La}(\text{Ac})_3$  derivative (2.5mM, 24hrs) where  $a$  and  $b$  decreased by 0.5% and  $c$  decreased by 0.9%. Pt and Ir-soaked crystals exhibited a slight colour change indicating that heavy atom binding may have occurred. However, a marked deterioration in the diffraction quality of the  $\text{IrCl}_6$ -soaked crystals was observed during data collection and this is reflected by an extremely large  $R_{\text{sym}}(I)$ . Although, future soaking studies with  $\text{IrCl}_6$  should involve lower concentrations of this reagent, the rapid deterioration of the diffraction quality with time suggests a substantial increase in the crystal's susceptibility to radiation damage in the presence of this reagent.

KI,  $K_2HgI_4$ , pCMBS and  $La(Ac)_3$  soaked crystals were essentially 'native' based on their relatively low mean fractional isomorphous differences and  $R_{deriv}$  values (Table 6.2). It is unlikely that any heavy atom binding occurred in these cases. Of the remaining derivative datasets, only the  $Gd^{3+}$  and  $PtCl_4^{2-}$  appeared to be more promising. Subsequent examination of all the isomorphous difference Pattersons failed to yield any strong, clear peaks on the Harker sections ( $u, v$  and  $w = 0$ ). In some cases, several self-consistent peaks were identified but initial phases, calculated from these sites, had very poor statistics (*PHASIT*).  $R_{centric}$  values were used as a primary indicator of potential heavy atom sites and none of the potential sites examined had values below 0.70.

#### 6.4 Summary

The replacement of the calcium ion in LA with a suitable heavy atom was an extremely attractive way to obtain the additional phasing information required to solve the structure. The calcium binding site is well defined in the LA molecule and the calcium ion can be removed without denaturing the protein. The failure to obtain hexagonal crystals could be due to a number of reasons. The first possibility is that none of the heavy atom occupied the vacated binding site. However, the observation that protein crystals could be obtained at all suggests that the calcium binding site was occupied by the desired heavy atom. Previous studies have shown that apo-LA cannot be crystallised in the absence calcium (unpublished observations). Calcium rebinding due to contamination of the reagents was considered unlikely due to the large excess of rare earth metal used in the initial equilibration.

A second possible explanation for the failure to obtain hexagonal crystals is that the rare earth, metal-loaded apo-LA has a different conformation from the native calcium-bound protein. LA is known to undergo a series of conformational changes in the presence of various metal ions (see Kronman, 1989). Although a slight conformational change may affect packing interactions between the LA molecules, it seems likely that such an effect does not take place as the monoclinic crystal form, that was obtained in the cocrystallisation trials, is sometimes observed with native LA under the hexagonal conditions. A more plausible explanation is that the excess heavy atoms directly influence the crystallisation process. Rare earth metals are known to bind to at least two other low affinity sites in LA that are distinct from the calcium binding loop (Kronman, 1989). Binding at these additional sites may influence the packing and final spacegroup of the crystals. Therefore, preparation of a lanthanide-loaded form of apo-LA in a 1:1 stoichiometry, targeting the high affinity site, might remedy the current situation.

The lack in finding any suitable derivatives in the soaking studies may be due to the high concentrations of ammonium sulphate (1.9-2.0M) present in the mother liquor. Ammonium sulphate is known to be a poor mother liquor due to the production of  $\text{NH}_3$  nucleophiles which can compete with heavy metals for binding sites on the protein (Sigler and Blow, 1965). A possible solution would be to replace the ammonium sulphate with another salt, such as sodium or lithium sulphate, or polyethylene glycol (PEG). Ammonium sulphate also has a high ionic strength. This tends to weaken electrostatic interactions and may prevent heavy atom binding. Results from the preliminary screening suggest that some heavy atom binding might have occurred but it



is associated with a multitude of poorly occupied sites. Therefore, switching to a PEG-based mother liquor, may improve the quality of heavy atom binding.

The present work represents a preliminary screening of a subset of potential heavy atom reagents. Lack of success in a more exhaustive search may prompt the use of alternative strategies. A more radical approach would be to try to insert a mercury atom into the hyper-reactive 6–120 disulphide bond (Ikeguchi *et al.*, 1992). This technique was critical in the structure determination of LA from baboon milk (Smith *et al.*, 1987). As long as the disulphide bond is accessible in the hexagonal crystals, derivatisation can be accomplished in the crystalline state (Ely *et al.*, 1973). If all else fails, the recombinant LA could be genetically altered with relative ease so that it contains additional heavy atom ligands. Potential modifications include the introduction of a free cysteine residue in an accessible region of the protein or the incorporation of selenomethionine in the place of methionine (Tucker *et al.*, 1989; Nagai *et al.*, 1991). The structure of the latter could then be solved by the multiwavelength anomalous diffraction (MAD) technique. However, such genetic alteration is not always straightforward as there is no guarantee that crystals of the variant LA will be isomorphous with the 'native' hexagonal form.

# Refinement and Model Building

## 7.1 Refinement Methods

### 7.1.1 Introduction

In order to obtain an accurate description of a protein's atomic positional and thermal parameters ( $x, y, z, B$ ), the initial model must be altered in such a way as to maximise its agreement with the experimental observations. This can be achieved by minimising the sum of the squares of the differences between the observed and calculated structure factors,

$$\phi = \sum_{\mathbf{h}} w_{\mathbf{h}} (|F_{\text{obs}}(\mathbf{h})| - |F_{\text{calc}}(\mathbf{h})|)^2 \quad [\text{Eq. 7.1}]$$

where  $\mathbf{h}$  represents the reflection  $hkl$  and  $w_{\mathbf{h}}$  are the weights associated with  $F_{\text{obs}}(\mathbf{h})$ . In protein crystallography, this is achieved by methods involving least squares minimisation and empirical energy minimisation. Although true least squares minimisation can in principle be performed if the number of observations exceeds the number of parameters, the degree of experimental error in protein crystallography often dictates that the ratio of observables ( $F_{\text{obs}}$ ) to unknowns (atomic positions,  $B$ -factors etc.) is considerably higher

than is often available. In a few cases, true least-squares refinement methods can be used if atomic resolution data is available (Dauter *et al.*, 1992; Wilson, 1994). However, most protein structures can only be determined at less than atomic resolution (1.5–3.0Å) and consequently the observation to parameter ratios are relatively low. As a result, gross distortions of stereochemistry tend to be introduced into the structure due to the atomic shifts calculated from the least squares equations. Furthermore, the inaccuracy of initial protein structures, combined with the non-linear nature of the refinement process, typically results in convergence to a false minimum if true least squares methods are used. Therefore, in order to accurately refine protein structures at less than atomic resolution, the least squares equations are modified to incorporate physical and chemical information relevant to macromolecular structures (Hendrickson, 1985). Stereochemical restraints, such as bond distances and angles, the chirality of amino acids and the planarity of peptide bonds and aromatic rings, serve to increase the number of observations and augment the diffraction data.

Additional procedures that have been developed to improve the refinement of protein structures include combination of the restrained, least squares procedure with energy minimisation and the use of molecular dynamics to escape 'local' minima during the refinement process (Jack and Levitt, 1978; Brünger *et al.*, 1987). The rest of this section will briefly describe the refinement process implemented in the *X-PLOR* package of programs (Brünger, 1990a; 1992a).

## 7.1.2 X-PLOR

### 7.1.2.1 The Energy Function

*X-PLOR* provides all the tools required to carry out the crystallographic refinement of proteins. The aim of the refinement is to find the global minimum of the target function,

$$E = E_{\text{CHEMICAL}} + w_{\text{XRAY}} E_{\text{XRAY}} \quad [\text{Eq. 7.2}]$$

where  $E_{\text{CHEMICAL}}$  comprises information about known chemical interactions,  $E_{\text{XRAY}}$  is the crystallographic residual and  $w_{\text{XRAY}}$  is an appropriate weight (Jack and Levitt, 1978). The rationale for refining against an empirical energy function is the assumption that the true protein structure is at an energy minimum as well as being a best fit to the experimental observations. In the *X-PLOR* energy function, the total energy ( $E_{\text{TOTAL}}$ ) consists of two separate terms.

$$E_{\text{TOTAL}} = E_{\text{EMPIRICAL}} + E_{\text{EFFECTIVE}} \quad [\text{Eq. 7.3}]$$

$E_{\text{EMPIRICAL}}$  encompasses the geometric and stereochemical restraints used in the refinement. It is analogous to  $E_{\text{CHEMICAL}}$  (Eq. 7.2) and describes the energy of the molecules within the AU through a geometric energy function consisting of covalent interactions (bond lengths, bond angles, dihedral torsion angles, chiral centres and planarity of aromatic rings) and non-bonded (Van der Waals, hydrogen bonds and electrostatic) interactions. The force and geometric constants used in  $E_{\text{EMPIRICAL}}$  are derived from small molecule crystallography and those used in *X-PLOR3.1* are taken from a recent comprehensive survey of the Cambridge Crystallographic Database (Engh and Huber, 1991). The non-bonded energy terms in  $E_{\text{EMPIRICAL}}$  (electrostatic and Van der

Waals) are subdivided into terms that describe intramolecular, crystallographic and non-crystallographic interactions (Brünger, 1992a).

The  $E_{EFFECTIVE}$  term in Eq. 7.3 consists of restraining energy terms that use experimental information such as structure factor amplitudes ( $E_{XREF}$ ) and non-crystallographic symmetry ( $E_{NCS}$ ). The most commonly used form of  $E_{XREF}$  is the crystallographic residual (similar to Eq. 7.1) with suitable weights. A scale factor ( $k$ ) is applied to the  $F_{calc}$  terms to minimise the residual. Computation of the structure factors ( $F_{calc}$ ) is accomplished by Fast Fourier transformation of the electron density (Brünger, 1989). If local symmetry is present, NCS-related molecules can either be treated as being strictly identical or atomic positions within the related molecules can be restrained to their average position. Similarly, NCS-related temperature factors can be restrained to their average value in the refinement of individual  $B$ -factors. The use of NCS restraints considerably improves the effective observation to parameter ratio and also speeds up convergence of the refinement procedure. More detailed descriptions of the various effective and empirical energy terms are given elsewhere (Brünger, 1992a).

### 7.1.2.2 Monitoring the Refinement

Progress in the refinement process is traditionally measured by the  $R$ -factor, defined as

$$R = \frac{\sum_{\mathbf{h}} \|F_{obs}(\mathbf{h}) - k|F_{calc}(\mathbf{h})\|}{\sum_{\mathbf{h}} F_{obs}(\mathbf{h})} \quad [\text{Eq. 7.4}]$$

$R$ -factors vary from 59% for a random acentric structure to values of between 10 and 20% for final refined protein structures. These values contrast the noise level in the

diffraction data which is estimated to be about 5%. The discrepancy between the final protein  $R$ -factor and the noise most likely results from an inability to correctly model thermal atomic motion, disordered regions of the protein and the disordered nature of the bulk solvent.

The  $R$ -factor is closely related to the crystallographic residual. By increasing the number of model parameters and refining against the crystallographic residual, the  $R$ -factor can be reduced to a relatively small value. In other words, during refinement the diffraction data can be significantly ‘overfitted’ (misfitted) even though there is no concomitant improvement in the ‘correctness’ of the atomic model. As a result, an incorrect model can often be refined to a respectable  $R$ -factor (Brändén and Jones, 1990). The advent of sophisticated refinement methods has compounded this problem. A common concern during refinement is whether the alterations made to the model actually improve the accuracy of the phase angles or merely overfit the structure factor amplitudes. To overcome the impressionable nature of the conventional  $R$ -factor, Brünger proposed a second reliability index, the free  $R$ -factor ( $R_{\text{free}}$ ) (Brünger, 1992c).

$R_{\text{free}}$  is a statistical quantity that measures the agreement between observed and calculated structure factors for a set of reflections ( $T$ ) that are omitted from the refinement.  $R_{\text{free}}$  is defined as,

$$R_{\text{free}} = \frac{\sum_{\mathbf{h} \in T} \left| |F_{\text{obs}}(\mathbf{h})| - k|F_{\text{calc}}(\mathbf{h})| \right|}{\sum_{\mathbf{h} \in T} |F_{\text{obs}}(\mathbf{h})|} \quad [\text{Eq. 7.5}]$$

where  $\mathbf{h} \in T$  represents all the reflections ( $hkl$ ) belonging to the test set ( $T$ ) of reflections.  $R_{\text{free}}$  is highly correlated to the phase accuracy of the atomic model and is

independent of the number of model parameters and restraints used (Brünger, 1993). It is particularly useful for distinguishing correct and incorrect molecular replacement solutions (Brünger, 1992b), checking for over-fitting and generally guiding the refinement protocol. It can also be used to optimise the overall weighting between the diffraction data and the chemical restraints ( $w_{\text{xray}}$  in Eq. 7.2) and to choose an appropriate temperature-factor model (Brünger, 1993). Free  $R$ -factors tend to be higher than conventional  $R$ -factor values and reported values for  $R_{\text{free}}$  range from the low 20s to the mid thirties for macromolecular structures. As a rule of thumb, the final free  $R$  value is generally equal to the sum of the final conventional  $R$ -factor and the merging  $R$ -factor of the diffraction data (Kleywegt and Jones, 1995b).

### 7.1.2.3 Generalised Refinement Protocol

A standard procedure has been used to refine the initial structures determined by molecular replacement. Today's refinement methods are open to considerable abuse and this has prompted considerable attention recently (Brändén and Jones, 1990; Kleywegt and Jones, 1995a,b). The use of powerful minimisation techniques such as molecular dynamics, a judicious choice of weights ( $w_{\text{XRAY}}$ ) and over-enthusiastic refinement of atomic temperature factors, especially in cases where there are insufficient observations, can reduce the  $R$ -factor to an unrealistic value. Therefore, considerable care needs to be taken during the refinement procedure. In most cases, initial refinement procedures should restrict the number of degrees of freedom of the model so as to limit erroneous adjustments in the model. In this case, these considerations have not been rigorously applied as the high resolution structures of several LAs, on which the MR search model

was based, have been determined independently. Nevertheless, NCS restraints and  $R_{\text{free}}$  have been used in the present work wherever possible.

X-ray diffraction data were converted into *X-PLOR* format using *MTZ2VARIOUS* (CCP4) and subsequently divided into the two reflection sets required for the  $R_{\text{free}}$  calculation. The test set (*T*) typically consisted of either 5 or 10% of randomly selected reflections. Molecular topologies, describing the model's atomic and covalent properties, were created using *GENERATE*. The first refinement step after molecular replacement was a rigid-body refinement. In this method, groups of atoms, typically the molecules in the AU, are treated as rigid bodies and three rotational ( $\theta_1, \theta_2, \theta_3$ ) and three translational ( $x, y, z$ ) degrees of freedom are minimised against  $E_{\text{TOTAL}}$ .

After rigid-body refinement, the models were refined using a generalised refinement and rebuilding protocol (Figure 7.1). Prior to each refinement cycle, the ideal weight ( $W_A$ ) between  $E_{\text{XREF}}$  and  $E_{\text{EMPIRICAL}}$  was calculated (*CHECK*). This empirical method typically over-determines the value of  $W_A$  by a factor of 2 or 3. A better estimate of the weight can be obtained by minimising  $R_{\text{free}}$  using simulated annealing (Brünger, 1993). The molecule was subjected to both conventional and simulated annealing positional refinement. At all stages, the refinement process was monitored by  $R_{\text{free}}$  and the number of cycles of minimisation were adjusted to minimise  $R_{\text{free}}$ . In *PREPSTAGE* and *POS*, the model was subjected to conventional positional refinement using the conjugate gradient method of Powell (1977). In this method, the coordinates of all the free atoms are used as variables in the minimisation of  $E_{\text{TOTAL}}$ . Strain and bad contacts introduced by manual



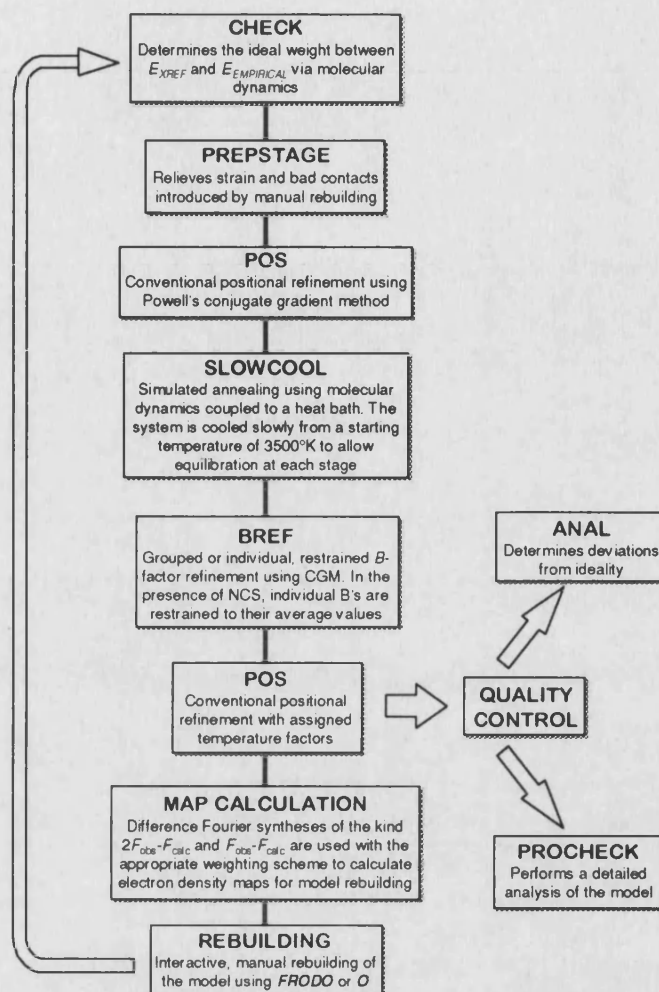


Figure 7.1 — Generalised refinement protocol.

rebuilding were relieved in *PREPSTAGE* by initially using a non-bonded ‘repel’ function without any electrostatic energy terms. After between 5 and 40 cycles of the repel step, depending on the state of the refined model, the repel function was turned off and the model was subjected to 100–200 cycles of conventional Powell minimisation. If further Powell minimisation was deemed beneficial, additional cycles were performed in *POS*.

The next step in the refinement protocol utilised simulated annealing (SA) through molecular dynamics (*SLOWCOOL*). This method is particularly useful in the early stages of refinement as it allows both uphill and downhill search directions to overcome barriers in the energy function. SA methods allow the sampling of a large area of conformational space and consequently have a large radius of convergence. Considerable improvements in the model's *R*-factor and geometry can be achieved using this technique (Brünger *et al.*, 1987; Kuriyan *et al.*, 1987). SA is also beneficial in cases where initial phase estimates have been obtained through MR, as it can remove most of the inherent bias towards the phasing (search) model. The 'slow-cooling' protocol used here typically involved heating the system up to 3500°K and then cooling to 300°K at a rate of 1°K per femtosecond (Brünger *et al.*, 1990c). The slow-cooling step was rounded off with up to 150 cycles of Powell energy minimisation. This relieves any geometric distortions that might have been introduced into the structure during the annealing process.

In *BREF*, the temperature factors of the model were refined. Temperature factor refinement is accomplished by tightly restraining the variance of the interatomic distances so that the *B*-factors of bonded atoms are correlated. The target function is minimised using the Powell conjugate gradient method. Although there was sufficient data to refine individual atomic *B*-factors in all cases, during the early stages only two temperature factors were refined per residue (grouped *B*-factor refinement). In the grouped refinement, the mainchain and sidechain atoms were allocated one parameter each. Once *B*-factors had been assigned to the atoms, the final step in the refinement protocol involved 50-100 cycles of positional refinement (*POS*).

### 7.1.2.4 Map Calculation

To minimise model bias in the calculated electron density maps, the program **SIGMAA** (CCP4) was used to obtain weighted map coefficients for the use in the difference Fourier syntheses. Structure factors were calculated for the refined model using **SFALL** (CCP4) and subsequently scaled with the observed structure factors using **RSTATS** (CCP4). Weighted Fourier coefficients were calculated from the calculated phases of the model using the program **SIGMAA** (CCP4) (Read, 1986; 1990). Initially **SIGMAA** estimates a value for  $\sigma_A$  over a series of resolution shells using normalised structure factors. A figure of merit for each reflection and the coordinate error in the model are then calculated. The output map coefficients were subsequently used in **FFT** (CCP4) to obtain difference maps of the type:

$$m|F_{\text{obs}}| - D|F_{\text{calc}}|e^{2\pi i\alpha_{\text{calc}}} \quad \text{and} \quad 2m|F_{\text{obs}}| - D|F_{\text{calc}}|e^{2\pi i\alpha_{\text{calc}}}$$

where  $m$  is the figure of merit and  $D$  is the coordinate error defined by Luzatti (1952). In the later stages of refinement and rebuilding, unweighted electron density maps were calculated using **X-PLOR**.

### 7.1.2.5 Manual Rebuilding

Both **FRODO** and **O** have been used for manual rebuilding (Jones, 1985; Jones *et al.*, 1991; Jones and Kjeldgaard, 1993). **FRODO** was run on an Evans and Sutherland PS390 and **O** (versions 5.9 and 5.10) was run on a Silicon Graphics *Indigo*<sup>2</sup> workstation. Due to the relatively high homology between the MR search model and the bovine, goat and guinea-pig LAs, rebuilding was fairly straightforward and mainly involved correcting sidechain torsion angles and adding tightly bound solvent molecules. Nevertheless, the

rotamer library in *O* was particularly useful for determining the conformation of problematic sidechains (Zou and Mowbray, 1994).

### 7.1.2.6 Assessing Model Quality

The refined model's stereochemical and geometric characteristics were routinely analysed at the end of each refinement cycle. Two methods were used; the first involved *X-PLOR*'s internal geometric analysis (*ANAL*) which produces a list of r.m.s deviations from ideality for bond lengths and angles, dihedral and improper angles and non-bonded and crystal packing contacts. The model's stereochemistry was also analysed using the program *PROCHECK* (Laskowski *et al.*, 1993). *PROCHECK* evaluates a series of parameters such as the quality of the Ramachandran plot, peptide bond geometry, bad contacts, hydrogen bond energies and sidechain geometry. *PROCHECK* was used to assess the overall quality of the model at different stages of refinement and also during manual rebuilding to identify potential problem areas in the structure. Other quality checks used during the refinement process included analysis of *B*-factors and, where appropriate, comparisons between NCS-related molecules.

## 7.2 Monoclinic mLA

### 7.2.1 Refinement

The *AMoRe* solution was initially refined against a 3Å in-house dataset (AD-LAB2) collected from the monoclinic III crystals (sodium form). Seventy cycles of rigid-body refinement (RBR), using data between 10 and 3.5Å (5737 reflections) with each of the four molecules in the AU as a separate entity, did not reduce the conventional *R*-factor (38.8%). This was not surprising as the MR solution had already been subjected to

RBR in *AMoRe* (*FITING*). Subsequent positional refinement (conventional and simulated annealing at 3500°K), using all data between 15 and 3.5Å (8199 reflections) and NCS restraints, reduced the conventional *R*-factor to 27.7%. The only rebuilding performed at this stage was the modification of the calcium binding ligands (Asp 82, 87 and 88) so that a calcium ion could be incorporated in each molecule.

High resolution refinement was begun after a relatively complete 2.3Å dataset had been collected from the monoclinic III crystals (magnesium form) at the SRS. A summary of the various cycles of refinement are given in Table 7.1.

*Cycle 1a* — Due to the slight unit cell differences between the sodium and magnesium forms of the monoclinic crystals (0.6Å in *a*, 0.4Å in *b*, 0.6Å in *c* and 0.15° in  $\beta$ ), RBR was initially used to adjust the position of the model. Five percent of the diffraction data (971 reflections) was partitioned into the test set (*T*) for the  $R_{\text{free}}$  calculation. Fifty cycles of four-body RBR, using data between 10 and 3.5Å reduced the conventional and free *R*-factors to the low 30s. Shifts produced by RBR were less than 1.8° for the angular parameters and less than 0.8Å for the translational parameters.

*Cycle 1b* — Immediately after the RBR, the model was subjected to conventional positional and individual *B*-factor refinement using all the data between 8 and 2.5Å. The atomic positions of the four molecules (A-D) in the AU were restrained using the available non-crystallographic symmetry. At the start of the refinement all the atomic *B*-factors were set to 15Å<sup>2</sup>. Molecule A was rebuilt manually using *O*. The additional N-terminal methionine residue, clearly visible in both the  $\sigma_A$ -weighted  $2F_o - F_c$  and  $F_o - F_c$  maps, was incorporated into the model. This residue was designated 1X so that it did not

interfere with the numbering of the wild type protein. The amino acid substitution at position 90 was also clearly visible, appearing as

Table 7.1 — Summary of the mLA refinement.

Cycle	Atoms		Protocol <sup>b</sup>	NCS <sup>c</sup>	Resolution Range (Å)	No. of <sup>d</sup> Reflection	<i>R</i> -factor		Free <i>R</i> -factor	
	Protein	Other <sup>a</sup>					IN	OUT	IN	OUT
1a	3864	4	RBR	4 body	10 – 3.5	5581	39.4	32.2	38.4	30.9
1b	3864	4	Pr/P/B/P	restrain	8 – 2.5	14534	46.8	25.7	46.9	35.4
2	3896	4	Pr/P/S/B/P	restrain	8 – 2.5	14534	35.7 <sup>e</sup>	24.0	37.4 <sup>e</sup>	31.3
3	3896	0	Pr/S/B/P	restrain	8 – 2.5	14534	26.7	20.0	31.3	30.5
4	3896	3	Pr/S/B/P	none	8 – 2.3	17711	25.1	21.7	31.4	30.9
5	3896	64 + 4	Pr/B/P	none	8 – 2.3	17711	25.9	21.3	31.1	29.5
6	3922	90 + 4	Pr/S/B	none	8 – 2.3	17711	25.7	21.5	31.2	32.0
7	3979	88 + 4	Pr/B/P	none	8 – 2.3	17711	23.1	20.4	32.1	31.9
8	3957	76 + 2	Pr/B/P	none	8 – 2.3	17711	23.3	21.4	30.0	30.3
9	3941	87 + 4	Pr/B/P/B/P	none	8 – 2.3	18605	25.0	20.9	–	–
10	3975	106 + 4	P/B/P/B/P/B/P	none	8 – 2.3	18605	31.9 <sup>e</sup>	20.8	–	–
11	3958	60 + 4	P	none	8 – 2.3	18605	20.8	20.8	–	–

<sup>a</sup> Non-protein atoms are water molecules and calcium ions (*italics*)

<sup>b</sup> Abbreviations used: RBR, rigid-body; Pr, *PREPSTAGE*; P, *POS*; S, *SLOWCOOL*; B, *BREF* (individual)

<sup>c</sup> NCS restraints equivalenced molecule A with the other three molecules using a 250kcal/mol-Å<sup>2</sup> effective force constant

<sup>d</sup> Number of reflections refers to those in the 'working set' (i.e. used for refinement)

<sup>e</sup> *B*-factors reset to 30Å<sup>2</sup> prior to refinement

a truncation of the sidechain density in the  $2F_o - F_c$  map, and the methionine in the search model was replaced with a valine. All the other alterations to the model involved the tweaking of sidechain torsion angles. Molecules B, C and D were generated from the rebuilt molecule A, using the least-squares superposition (*LSQ*) option in *O*, and the resultant model was re-refined.

*Cycle 2* — A similar refinement protocol to cycle 1b was used apart from an additional SA step (2500°K). This procedure improved both *R*-factors and the model's deviations from ideality. On viewing the  $2F_o - F_c$  and  $F_o - F_c$  maps, it became clear that all four molecules were slightly different and so each one was individually rebuilt. The electron density for the two dimer pairs (AB and CD) was radically different. Molecules A and B exhibited weak density for residues 43-47, 71-78 and 112-117 whereas molecules C and

D were poorly defined in the regions 1-4, 40-48 and 52-78. The disordered nature of these regions is reflected in their final  $B$ -values (Figure 7.4). Major alterations at this stage included fitting residues 41-48 into the electron density and adjusting the torsion angles of residues 45 and 46 in each protomer so that they fell into allowed regions of the Ramachandran plot and correcting the geometry of the calcium binding sites.

*Cycle 3* — The resultant model was refined using a similar protocol to cycle 2 whilst still maintaining NCS restraints. Although each of the protomers had been rebuilt separately, the extent of the data at 2.5Å resolution was deemed to be insufficient to permit a release of the restraints. At the end of the refinement, the conventional and free  $R$ -factors had improved by 4% and 0.8% on the values obtained after cycle 2. Residues 45 and 46 in each protomer required further torsional adjustments. In addition, electron density for the bound calciums was particularly weak, especially in molecules C and D where the  $B$ -factors were 50.33Å<sup>2</sup> and 51.67Å<sup>2</sup> respectively. The mainchain of calcium binding sites in molecules C and D was also distorted between residues 82 and 84. This scenario would become a recurrent theme during the rest of the refinement. Consequently, all four calcium ions were excluded from the model to see if they would return as peaks in the  $F_o - F_c$  difference density.

*Cycles 4 to 8* — During these cycles, the high resolution limit was increased to 2.3Å and the NCS restraints were removed. The addition of water molecules and further rebuilding had a minimal effect on  $R_{\text{free}}$  and it stagnated at around 30%. The excluded calcium ions reappeared in the  $F_o - F_c$  difference density and were incorporated into the model over the next two cycles (4 and 5). After cycle 5, the C-terminal tripeptide (Glu-121 to Leu-123) was included in molecule C as the electron density in this region was reasonably well-

defined; in the following cycle, the C-termini (poly-alanine) were tentatively added for molecules A, B and D. The occupancy and geometry of the calcium sites of molecules C and D continued to cause problems and so simulated annealing omit maps were calculated over this region. In each case, residues 79 to 89 were omitted along with a surrounding  $8\text{\AA}$  sphere and the remaining atoms were subjected to SA from a starting temperature of  $1000^\circ\text{K}$ . The calcium binding loops were rebuilt with a bound ion using the resultant omit maps. These maps indicated that there was no significant change in the calcium coordination in molecules C and D. The primary cause of the problems encountered in this part of the structure appears to be associated with its high temperature factors. A significant number of the water molecules that were added in each rebuilding step refined with  $B$ -values greater than  $50\text{\AA}^2$ . This can be attributed to the high overall  $B$ -factor applied to the data and water molecules were only included in the model if they fulfilled strict hydrogen-bonding criteria and had  $B$ -values that were less than  $60\text{\AA}^2$ .

*Cycles 9 to 11* — By this stage of the refinement, alterations to the model did not improve  $R_{\text{free}}$  and therefore all the data to  $2.3\text{\AA}$  was used in the last two refinement cycles. A series of positional and temperature factor refinements improved the conventional  $R$ -factor by 0.7% and gave a final value of 20.8%. Manual rebuilding was entirely focused on the positions of water molecules and the C-terminal tripeptides of each molecule. The large rise in the  $R$ -factor between the end of cycle 9 and 10 was due to resetting all the temperature factors to  $30\text{\AA}^2$ . The  $B$ -factors were subsequently reassigned using a series of short conventional positional refinements interspersed with individual  $B$ -factor refinements until the  $R$ -factor converged. Residues 122 and 123 in



molecule C had extremely poor density and were removed from the model. Solvent molecules with temperature factors above  $60\text{\AA}^2$  were removed from the final model before carrying out a final five cycles of positional refinement. The final structure had a *R*-factor of 20.8% for all data between 8 and  $2.3\text{\AA}$  (18605 reflections).

Table 7.2 — Statistics for the final mLA structure.

Molecule	All	A	B	C	D
Protein atoms	3958	992	983	983	1000
Water molecules (calcium ions)	60 (4)	26 (1)	20 (1)	8 (1)	6 (1)
Rmsd bond lengths <sup>a</sup> (Å)	0.011	0.012	0.012	0.010	0.011
Rmsd bond angles <sup>a</sup> (°)	1.73	1.81	1.80	1.63	1.66
Rmsd dihedrals <sup>a</sup> (°)	24.70	25.60	24.54	24.00	24.70
Rmsd impropers <sup>a</sup> (°)	1.55	1.56	1.59	1.55	1.52
Average overall <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	48.39	40.15	40.43	54.04	59.10
Rms $\Delta B$ bonded atoms <sup>b</sup> (Å <sup>2</sup> )	4.08	4.33	4.33	3.90	3.76
Average mainchain <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	47.60	38.98	39.46	53.18	58.67
Average sidechain <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	49.41	41.20	41.48	55.16	59.69
Average solvent <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	40.58	42.45	37.88	37.36	45.79
Average calcium <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	49.45	35.18	43.38	63.11	56.12
Ramachandran plot,					
Most favoured areas <sup>c</sup> (%)	81.6	83.2	80.4	83.9	78.9
Additional allowed areas <sup>c</sup> (%)	17.5	15.9	18.8	15.2	20.2
Generously allowed areas <sup>c</sup> (%)	0.9	0.9	0.9	0.9	0.9
Disallowed areas <sup>c</sup> (%)	0.0	0.0	0.0	0.0	0.0
Non-rotamer sidechain conformations <sup>c</sup> (%)	4.3	4.9	6.6	1.6	4.0
Overall <i>G</i> -factor <sup>c</sup>	0.12	0.09	0.13	0.15	0.11

<sup>a</sup> Calculated using *X-PLOR* geometric analysis tools

<sup>b</sup> Calculated using *4D MOLEMAN*

<sup>c</sup> Defined by *PROCHECK*

### 7.2.2 Assessment of the Quality of the Final Model

The final structure consists of 4022 atoms of which 3958 are protein, 60 are solvent and 4 are calcium ions. It proved impossible to completely define the positions of all 124 residues in each protomer due to the flexibility of the C-terminal tripeptide.

Molecule A comprises residues 1X-122<sup>1</sup>, molecules B and C of residues 1X-121 and molecule D of residues 1X-123. Despite the poor definition of some regions of the structure, the final model has acceptable stereochemical and geometric statistics (Table 7.2). The mean coordinate error is estimated by *SIGMAA* to be 0.46Å (5–2.3Å). This is a little higher than the range commonly observed for refined protein structures (0.2-0.3Å) and probably due to a combination of the effects of radiation damage on the crystals and the relatively poor quality of the high resolution diffraction data.

#### *A. Agreement with the Electron Density*

The electron density is continuous along the whole protein backbone apart from a few surface loops. A representative portion of the electron density is shown in Figure 7.2. Although the overall electron density is good, there are several regions that are extremely poorly defined. The exposed loop regions of molecules A and B, consisting of residues 45-47 and 63-65, have weak electron density and there are several breaks in the continuity of the mainchain electron density. Residues 112-117 and those after Leu-119 are also poorly defined. Mobility at the C-terminus is highlighted by the extremely poor density for the S<sup>γ</sup> atom of Cys-120.

In contrast, molecules C and D are poorly defined at the N-terminus (Glu-1 and Gln-2) and around residues 43-47, 56-78 and those after Cys-120. The majority of the mainchain electron density breaks occur in the second half of the β-domain. This region in molecules C and D is not involved in any crystal contacts and is particularly mobile (Figure 7.4).

---

<sup>1</sup> 1X refers to the additional N-terminal methionine residue present in the recombinant mLA structure.

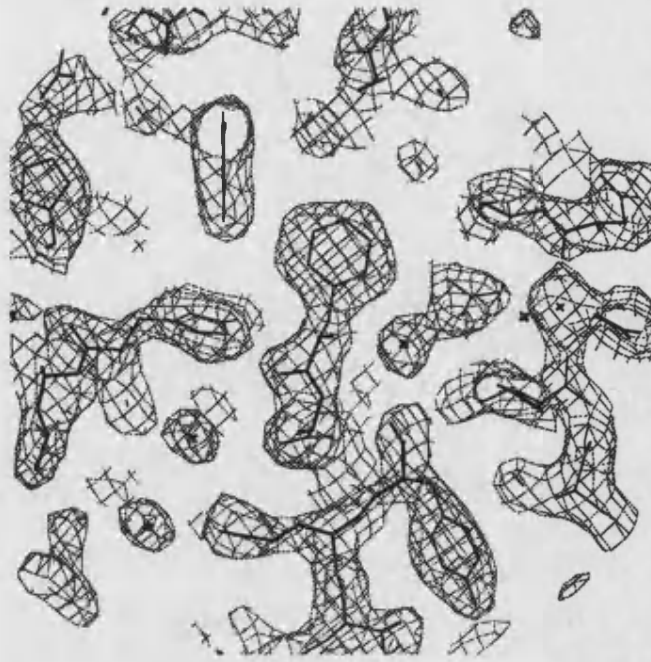


Figure 7.2 — Representative portion of the  $2F_o - F_c$  electron density with the final mLA structure.

The view shows the electron density in the core of the  $\alpha$ -domain of molecule A. Residues Phe-53 (centre), Tyr-50 (lower middle), Trp-26 and Trp-104 are shown. The calcium binding site is also apparent on the far right hand side of the plot. The two main chain calcium ligands (Lys-79 and Asp-84) are shown along with the carboxylate group of Asp-88. The resolution is  $2.3\text{\AA}$  and the map is contoured at  $1\sigma$  (where  $\sigma$  is the root mean square density throughout the unit cell). Crosses indicate the positions of water molecules. Figure produced with *4D\_OPLOTT*.

### B. Ramachandran Plot and General Stereochemistry

Analysis with the program *PROCHECK* shows that the final model has acceptable stereochemistry and geometry. All eleven parameters evaluated are either within (6), or better than (5) the bounds established from well refined structures of an equivalent resolution. None of the non-glycine residues are located in disallowed regions of the Ramachandran plot and only four residues (Asn-45A, Lys-114B, Ala-40C, Glu-121D), are located in a generously allowed region (Figure 7.3). All these residues are located in ill-defined and highly mobile regions of the structure.

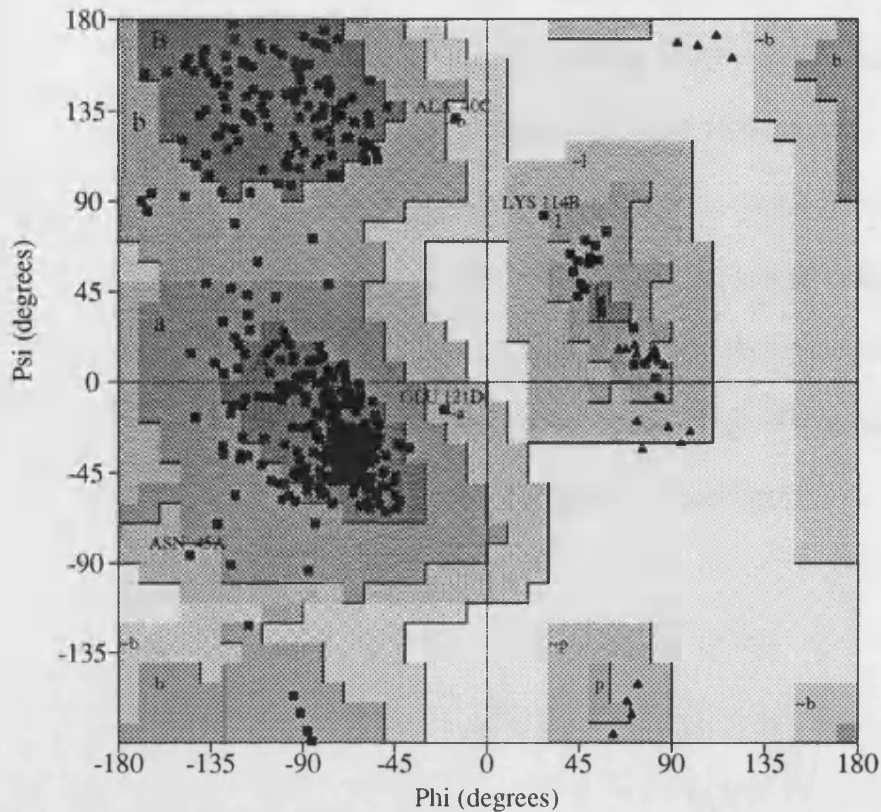


Figure 7.3 — Ramachandran plot for the mLA monoclinic structure.

All four molecules in the asymmetric unit are shown. The various shaded areas of the plot correspond to the most favoured (A,B,L), additional allowed (a,b,l,p) and generously allowed (~a,~b,~l,~p) regions. The glycine residues are shown as triangles. Those residues that fall in the generously allowed regions are highlighted. The plot was produced with *PROCHECK* (Laskowski *et al.*, 1993).

### C. Temperature Factors

The average overall *B*-factors for all four molecules are quite high. This can probably be attributed to the scaling process where data, from a relatively large number of crystals, had to be merged to obtain a reasonably complete set of data. The high overall *B*-factor applied in *TRUNCATE* ( $38.1 \text{ \AA}^2$ ) reflects the weak nature of the high resolution data. Nevertheless, the isotropic *B*-factor model used during refinement appears to have been appropriate as there is a good correlation between the *B*-values for

bonded atoms (Table 7.2). The molecules within the two dimers (AB and CD) also have similar overall, mainchain and sidechain average temperature factors even though, for most of the refinement, the  $B$ -values were not restrained between NCS-related molecules (Figure 7.4). The large differences in the thermal mobility of the two dimers that is apparent in Figure 7.4 probably reflects their different environments within the crystal. Nevertheless, the overall trend of the temperature factors, with respect to residue number, is similar for each dimer pair and parallels the variation seen for other LA structures (Acharya *et al.*, 1989; 1991; see also Figures 7.7 and 7.10).

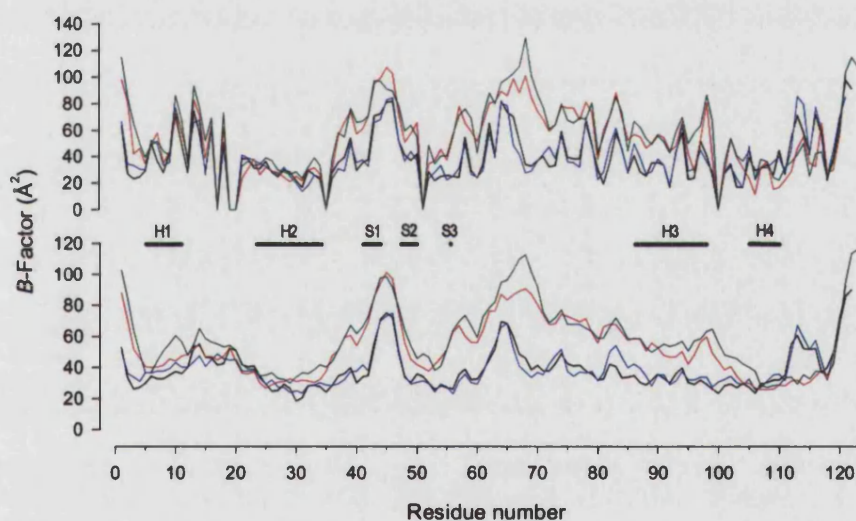


Figure 7.4 — Average temperature factors for the mLA structure.

The upper panel shows average temperature factors for sidechain atoms, and the lower panel shows average temperature for mainchain atoms, against residue number. All four molecules in the AU are shown – A (black), B (blue), C (red), D (green). The positions of the major secondary structural elements are indicated.

#### D. Similarity of NCS-Related Molecules

As with the temperature factors, the mainchain conformation of the molecules within each dimer is similar but the conformation of molecules from separate dimers

differs significantly (Table 7.3). This breakdown in NCS probably reflects the different packing forces that the flexible regions in the LA are subjected to in the crystal. The major conformational differences between the four molecules occur in the solvent exposed regions Arg10-Asp14, Gln43-Asp46, Lys62-Gln65 and Ser112-Asp116 (only in the AB pair). These regions are poorly defined and exhibit relatively high thermal displacements (Figure 7.4). Furthermore, it appears that the conformational differences between molecules from each dimer (i.e. AC, AD, BC, BD) are compounded by relative mobilities of each dimer pair. The average mainchain B-factor for the AB pair is  $39.2\text{\AA}^2$  compared to  $56\text{\AA}^2$  for the less well-defined CD dimer.

Table 7.3 — Comparison of NCS-related molecules.

The dimer pairs (AB and CD) are shaded. The molecules were superposed using the *LSQ\_EXP* command in *O* and the various root mean square deviations were calculated with the program *4D\_RMSPDB*.

Molecule	AB	AC	AD	BC	BD	CD
Rmsd NCS-related C $\alpha$ atoms <sup>a</sup> (Å)	0.28	0.60	0.66	0.59	0.64	0.41
Rmsd all NCS atoms <sup>a</sup> (Å)	0.96	1.12	1.35	1.26	1.35	1.15
Rmsd core C $\alpha$ atoms <sup>b</sup> (Å)	0.20	0.31	0.36	0.33	0.37	0.3
Residues with $\Delta\phi > 10^\circ$ (%)	29.8	43.0	40.5	50.4	45.5	43.0
Residues with $\Delta\psi > 10^\circ$ (%)	32.2	44.6	49.6	47.1	49.6	47.1
Rms $\Delta B$ NCS-related C $\alpha$ atoms <sup>a</sup> (Å)	5.10	22.6	25.27	22.66	24.07	6.78
Rms $\Delta B$ all NCS atoms <sup>a</sup> (Å)	6.49	23.53	26.05	23.48	25.14	8.12

<sup>a</sup> Residues Met-1X to Cys-120

<sup>b</sup> Core defined as residues 5-11, 23-43, 48-61 and 73-111

What are the underlying causes of these differences? Clearly, the packing forces within the crystal are partly responsible. Although the local symmetry relates the AB and CD molecules, there are no intermolecular contacts either within or between the dimers in the crystallographic AU. Furthermore, the dimers are stabilised to different extents by the neighbouring molecules in the crystal (Table 7.4). Consequently, the relative variation in temperature factors with residue number in each molecule is highly

correlated with the crystal contacts. For example, virtually all the intermolecular contacts made by molecules C and D involve residues at the C-terminus (Leu105-Gln117) and it is understandable why this region is one of the most rigid parts of the structure (Figure 7.4). Conversely, no crystal contacts stabilise this region (Ser112-Gln-117) in molecules A and B and, as a result, it is poorly defined. Finally, the absence of any crystal contacts involving residues 34-105 in molecules C and D probably explains why this region exhibits considerably higher mobility than the corresponding regions in molecules A and B (see Figure 7.4).

In summary, the breakdown of local symmetry appears to be almost entirely due to differences in the conformation of the poorly defined regions of the structure. Clearly, it would be wrong to assume that the NCS-related molecules are identical. Realistic differences between molecules related by local symmetry are estimated to be between 0.2 and 0.3Å (Kleywegt and Jones, 1995b). In this case, the deviations are quite large and it appears that, for whatever reason, some overfitting may have occurred during the refinement. In hindsight, it may be more appropriate to maintain the NCS restraints throughout the refinement process given the poor quality and completeness of the high resolution data. Nonetheless, these considerations do not reduce the significance of the final model as long as one accepts that the differences between the four LA molecules are unrealistic. Better quality diffraction data will be required if the conformations of the various molecules are to be accurately modelled.



Table 7.4 — Intermolecular crystal contacts for mL.A.

Atom	Symm	Atom	Distance (Å)	Atom	Symm	Atom	Distance (Å)
<b>MOLECULE A</b>				<b>MOLECULE B</b>			
Gln2A N <sup>ε</sup> 2	1-A	Leu105C O	2.76	Gln2B N <sup>ε</sup> 2	1-A	Leu105D O	2.80
Gln2A O		Lys108C N <sup>ε</sup> 5	2.79	Lys5B N		Ala109D O	3.00
Lys5A N		Ala109C O	3.07	Lys5B N <sup>ε</sup> 5		Glu113D O <sup>δ</sup> 1	3.09
Lys5A N <sup>ε</sup> 5		Glu113C O <sup>ε</sup> 1	3.08	Gln43B N <sup>ε</sup> 2	2	Asp116A O <sup>δ</sup> 2	3.20
Asn56A N <sup>δ</sup> 2	2+A-B	Asp116C O <sup>δ</sup> 2	2.94	Asn56B N <sup>δ</sup> 2	2+A+C	Asp116D O <sup>δ</sup> 2	2.94
Trp60A N <sup>ε</sup> 1		Asp116C O <sup>δ</sup> 2	3.17	Trp60B N <sup>ε</sup> 1		Asp116D O <sup>δ</sup> 2	3.17
His68A N <sup>ε</sup> 2	2+A-B+C	Gln117D O <sup>ε</sup> 1	2.91	His68B N <sup>ε</sup> 1	2+A	Lys5C N <sup>ε</sup> 5	3.11
Val99A O	2+A-B	Ser22C O <sup>γ</sup>	2.60	Val99B O	2+A+C	Ser22D O <sup>γ</sup>	2.60
Asn102A N <sup>δ</sup> 2		Glu25C O <sup>ε</sup> 2	3.07	Lys108B N <sup>ε</sup> 5		Ser112D O <sup>γ</sup>	2.71
Lys108A N <sup>ε</sup> 5		Ser112C O <sup>γ</sup>	2.93				
Asp116A O <sup>δ</sup> 2	2-B	Gln43B N <sup>ε</sup> 2	3.20				
Glu121A O <sup>ε</sup> 2	2+A	Asn45C N <sup>δ</sup> 2	2.89				
<b>MOLECULE C</b>				<b>MOLECULE D</b>			
Lys5C N <sup>ε</sup> 5	2+A-B	His68B N <sup>δ</sup> 1	3.11	Ser22D O <sup>γ</sup>	2+A-B+C	Val99B O	2.84
Ser22C O <sup>γ</sup>	2+A	Val99A O	2.60	Leu105D O	1+A	Gln2B N <sup>ε</sup> 2	2.80
Glu25C O <sup>ε</sup> 2		Asn102A N <sup>δ</sup> 2	3.07	Ala109D O		Lys5B N	3.00
Asn45C N <sup>δ</sup> 2	2+A-B	Glu121A O <sup>ε</sup> 1	2.89	Ser112D O <sup>γ</sup>	2+A-B+C	Lys108B N <sup>ε</sup> 5	2.71
Leu105C O	1+A	Gln2A N <sup>ε</sup> 2	2.76	Glu113D O <sup>ε</sup> 1	1+A	Lys5B N <sup>ε</sup> 5	3.09
Lys108C N <sup>ε</sup> 5		Gln2A O	2.79	Asp116D O <sup>δ</sup> 2		Trp60B N <sup>ε</sup> 1	3.05
Ala109C O		Lys5A N	3.07	Asp116D O <sup>δ</sup> 2		Asn56B N <sup>δ</sup> 2	2.75
Ser112C O <sup>γ</sup>	2+A	Lys108A N <sup>ε</sup> 5	2.93	Gln117D O <sup>ε</sup> 1	2+A+C	His68B N <sup>ε</sup> 2	2.91
Glu113C O <sup>ε</sup> 1	1+A	Lys5A N <sup>ε</sup> 5	3.08				
Asp116C O <sup>δ</sup> 2	2+A	Trp60A N <sup>ε</sup> 1	3.17				
Asp116C O <sup>δ</sup> 2		Asn56A N <sup>δ</sup> 2	2.94				

Symmetry operators: (1) x, y, z; (2) -x, 1/2+y, -z

## 7.3 Goat LA

### 7.3.1 Refinement

The *AMoRe* solution was refined at 2.3Å resolution using all the data between 8 and 2.3Å. A summary of the various cycles of refinement are given in Table 7.5.

*Cycle 1a* — RBR was initially used to adjust the positions of the MR solution. Ten percent of the diffraction data (1045 reflections) was partitioned into the test set (*T*) for the  $R_{\text{free}}$  calculation. Fifty cycles of two-body RBR, using data between 15 and 3.0Å reduced the conventional and free *R*-factors to the mid forties. Shifts produced by RBR were less than 1° for the angular parameters and less than 0.5Å for the translational parameters.



*Cycle 1b* — Immediately after the RBR, the model was subjected to conventional positional and grouped *B*-factor refinement using all the data between 8 and 2.3Å. The atomic positions of the two molecules (A and B) in the AU were restrained using the available non-crystallographic symmetry. At the start of the refinement all the atomic *B*-factors and occupancies were set to 15Å<sup>2</sup> and 1.0 respectively. Molecule A was rebuilt manually using *O*. The bound calcium ion was clearly visible in both the  $\sigma_A$ -weighted  $2F_o - F_c$  and  $F_o - F_c$  maps and was incorporated into the model. The other major alterations made to the model involved regions 66–69 and 105–110. Residues 105–110 in particular needed considerable rebuilding due to significant conformational differences between the MR model and the goat structure. In GOLLA, the electron density for this region appeared to be ‘looped-out’ away from the body of the protein. The C-terminal tripeptide (121–123) was also included (poly-alanine). Molecule B was generated from the rebuilt molecule A using the least-squares superposition (*LSQ*) option in *O* and the resultant model was re-refined.

Table 7.5 — Summary of the GOLLA refinement.

Cycle	Atoms		Protocol <sup>b</sup>	NCS <sup>c</sup>	Resolution Range (Å)	No. of <sup>d</sup> Reflection	<i>R</i> -factor		Free <i>R</i> -factor	
	Protein	Other <sup>a</sup>					IN	OUT	IN	OUT
1a	1938	0	RBR	2 body	15 – 3.0	4460	47.0	45.6	46.8	46.6
1b	1938	0	Pr/P/Bg/P	restrain	8 – 2.3	9107	47.1 <sup>e</sup>	28.3	48.3 <sup>e</sup>	34.3
2	1968	0 + 2	Pr/P/S/B/P	restrain	8 – 2.3	9107	32.6	23.4	32.3	29.8
3	1968	0 + 2	Pr/P/S/B/P	restrain	8 – 2.3	9107	32.2	21.7	31.5	28.7
4	1979	41 + 2	Pr/P/S/B/P	none	8 – 2.3	9107	29.3	19.4	27.8	27.5
5	1938	42 + 2	Pr/P/B/P	none	8 – 2.3	9107	28.7 <sup>e</sup>	20.6	28.4 <sup>e</sup>	28.1
6	1938	23 + 2	Pr/P/S/B/P	none	8 – 2.3	9107	21.3	19.3	26.7	27.0
7	1938	60 + 2	Pr/P/B	none	8 – 2.3	9107	20.2	19.8	27.1	27.0
8	1938	70 + 2	Pr/P/B/P	none	8 – 2.3	10138	20.9	20.2	–	–
9	1938	75 + 2	Pr/P/B/P	none	8 – 2.3	10138	20.6	19.8	–	–
10	1938	68 + 2	Pr/P/B/P	none	8 – 2.3	10138	20.5	19.5	–	–

<sup>a</sup> Non-protein atoms are water molecules and calcium ions (*italics*)

<sup>b</sup> Abbreviations used: RBR, rigid-body; Pr, *PREPSTAGE*; P, *POS*; S, *SLOWCOOL*; Bg, *BREF* (grouped); B, *BREF* (individual)

<sup>c</sup> NCS restraints equivalenced molecule A with the other three molecules using a 300kcal/mol-Å<sup>2</sup> effective force constant

<sup>d</sup> Number of reflections refers to those in the ‘working set’ (i.e. used for refinement)

<sup>e</sup> *B*-factors reset to 15Å<sup>2</sup> prior to refinement

*Cycle 2 and 3* — A similar refinement protocol to cycle 1b was used apart from an additional SA step (3000°K). The atomic positions of the two molecules in the AU were restrained to their average positions. This procedure improved both the conventional and free  $R$ -factors. The conformation of residues 44-47, 66-69 and 105-110 was improved by manual rebuilding so that all the residues fell in allowed regions of the Ramachandran plot. The  $2F_o - F_c$  electron density for the C-terminal tripeptide of molecule A was extremely ill-defined and this region was removed from the model. After the third refinement cycle, several water molecules were assigned to strong peaks ( $3\sigma$ ) in the  $F_o - F_c$  electron density map.

*Cycle 4 to 7* — At this stage in the refinement, virtually all the residues had acceptable  $\phi$  and  $\psi$  angles and so the NCS restraints were removed. Omit maps were used to unambiguously define the conformations of residues 44-47, 66-69 and 105-110. In contrast, the conformation of the three residues at the C-terminus continued to cause problems. In a final attempt to determine whether these residues should be included in the model, a  $2F_o - F_c$  electron density map was calculated with residues 121-123 omitted. These omit maps failed to show any significant continuous electron density and so no further attempts were made to model this region in either molecule. Additional water molecules were also included in the model. Both the conventional and free  $R$ -factors converged by cycle 7.

*Cycle 8 to 10* — In the last three refinement cycles all the data between 8 and 2.3Å was used. The bound water molecules were tidied up and those that had  $B$ -values greater than 50Å<sup>2</sup> were removed. The final model, consisting of residues 1-120 in molecules A and B, has an a  $R$ -factor of 19.5% for all data between 8 and 2.3Å (10138 reflections).

### 7.3.2 Assessment of the Quality of the Final Model

The final model consists of 2008 atoms of which 1938 are protein, 68 are solvent and 2 are calcium ions. A summary of the model's stereochemical and geometric statistics is given in Table 7.6. The mean coordinate error is estimated by *SIGMAA* to be 0.30 Å (5–2.3 Å).

Table 7.6 — Statistics for the final GOLLA structure.

Molecule	All	A	B
Rmsd bond lengths <sup>a</sup> (Å)	0.011	0.011	0.012
Rmsd bond angles <sup>a</sup> (°)	1.64	1.63	1.65
Rmsd dihedrals <sup>a</sup> (°)	23.83	23.07	23.74
Rmsd impropers <sup>a</sup> (°)	1.41	1.45	1.36
Average overall <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	19.62	20.20	19.05
Rms $\Delta B$ bonded atoms <sup>b</sup> (Å <sup>2</sup> )	3.57	3.57	3.58
Average mainchain <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	16.85	17.34	16.35
Average sidechain <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	21.02	21.46	20.57
Average solvent <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	39.0	41.39	36.28
Average calcium <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	18.23	17.53	18.93
Ramachandran plot,			
Most favoured areas <sup>c</sup> (%)	87.4	86.5	88.3
Additional allowed areas <sup>c</sup> (%)	12.2	13.5	10.8
Generously allowed areas <sup>c</sup> (%)	0.45	0	0.9
Disallowed areas <sup>c</sup> (%)	0	0	0
Non-rotamer sidechain conformations <sup>c</sup> (%)	0.83	1.7	0
Overall G-factor <sup>c</sup>	0.23	0.24	0.22

<sup>a</sup> Calculated using *X-PLOR* geometric analysis tools

<sup>b</sup> Calculated using *4D\_MOLEMAN*

<sup>c</sup> Defined by *PROCHECK*

#### A. Agreement with the Electron Density

The electron density is essentially continuous along the whole protein backbone. The fit of atoms to the final  $2F_o - F_c$  electron density map is shown in Figure 7.5. The density for residues 63A–68A is particularly ill-defined and a break in the mainchain density occurs at Asn-66A. Several of the long, hydrophilic sidechains that are on the exterior of the protein are also partially disordered (Gln-10A/B, Lys-11A, Lys-98A/B,

Lys108A/B, Lys-114A, Lys-58B, Lys-94B). The lack of density in these regions correlates with their high atomic temperature factors (Figure 7.7).

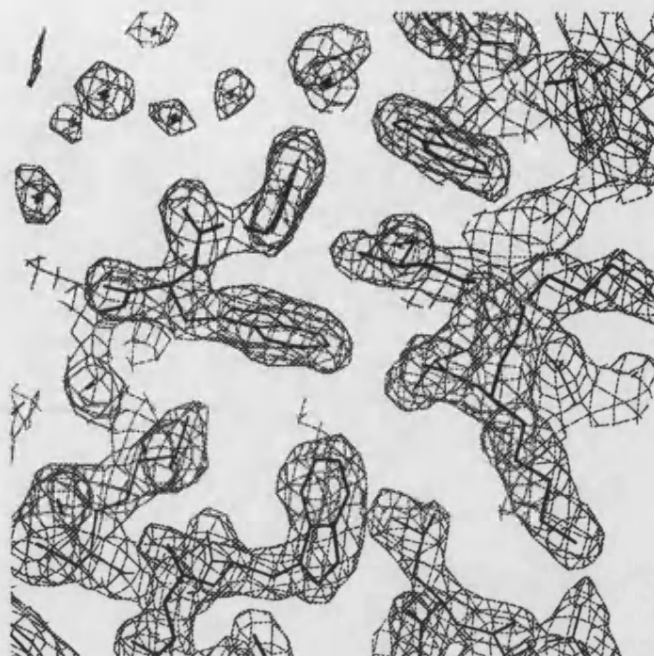


Figure 7.5 — Representative portion of the  $2F_o - F_c$  electron density with the final GOLA structure.

This view shows the electron density in the hydrophobic box region of molecule A. The four principal hydrophobic box residues Ile-95, Trp-104, Tyr-103 and Trp-60 (labelled in a clockwise direction) are apparent in the upper portion of the figure. In the lower portion, the view includes residues in the core of the  $\alpha$ -domain (Trp-26 and Leu-15). The resolution is  $2.3\text{\AA}$  and the map is contoured at  $1\sigma$  (where  $\sigma$  is the root mean square density throughout the unit cell). Crosses indicate the positions of water molecules. Figure produced with *4D\_OPLOTT*.

### B. Ramachandran Plot and General Stereochemistry

Analysis with the program *PROCHECK* shows that the final model has acceptable stereochemistry and geometry. All eleven parameters evaluated are either within (3), or better than (8) the bounds established from well refined structures of an equivalent resolution. None of the non-glycine residues are located in disallowed regions of the

Ramachandran plot and only one residue, Gln-54B, is located in a generously allowed region (Figure 7.6).

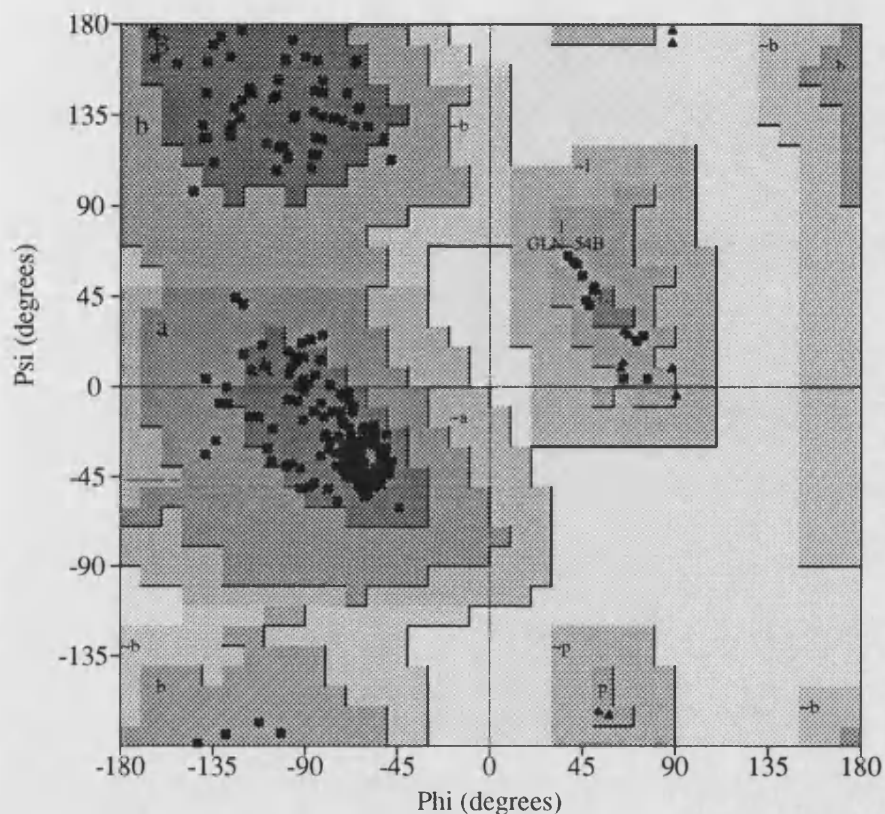


Figure 7.6 — **Ramachandran plot for the GOLA monoclinic structure.**

Both molecules in the asymmetric unit are shown. Those residues that fall in the generously allowed regions are highlighted. The shaded areas are the same as those defined in Figure 7.3. The plot was produced with *PROCHECK* (Laskowski *et al.*, 1993).

### C. Temperature Factors

Figure 7.7 shows a plot of the average mainchain and sidechain *B*-factors for both molecules. The most mobile regions in the structure correspond to the two solvent-exposed loops in the  $\beta$ -domain of LA (Gln43-Ser47 and Asp63-His68). Regions that are involved in the formation of the hydrophobic core of LA (part of helix H2 and residues

Tyr50-Phe53) and the calcium-binding site (residues in the Lys-79 to Asp-88 region) represent the most rigid parts of the structure. A similar variation in temperature factors is also observed for the mLA and GPLA structures (Figures 7.4 and 7.10) and is fairly typical for LA (Acharya *et al.*, 1989).

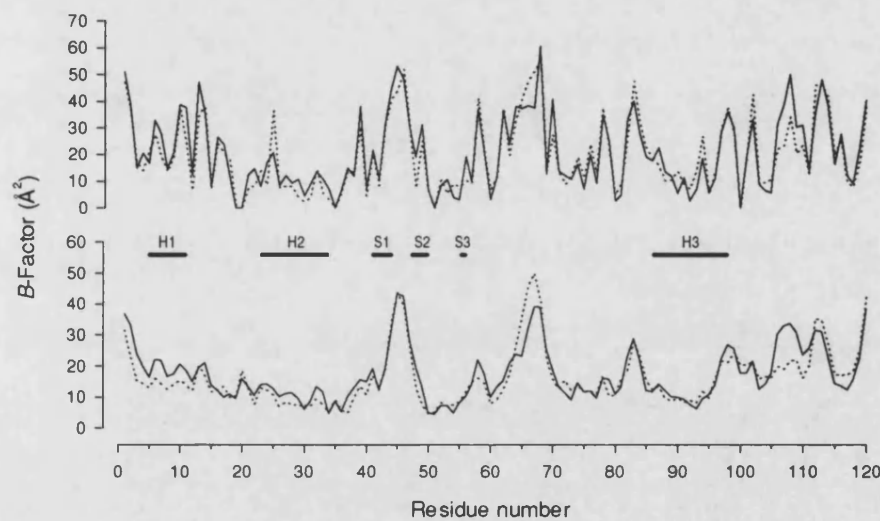


Figure 7.7 — Average temperature factors for the GOL structure.

The upper panel shows average temperature factors for sidechain atoms, and the lower panel shows average temperature for mainchain atoms, against residue number. *B*-factors for both molecule A (solid line) and molecule B (dotted) are shown. The positions of the major secondary structural elements are indicated.

#### D. Similarity of NCS-Related Molecules

The two molecules in the asymmetric unit of the crystal are very similar (Table 7.7). All the quality criteria evaluated, apart from the  $\Delta\phi$  and  $\Delta\psi$  values, are within the strict ranges suggested by a recent analysis of model quality (Kleywegt and Jones, 1995b). The differences between the two NCS-related molecules are localised in the

flexible regions of the structure. Almost all of the large deviations in backbone torsion angles ( $\Delta\phi$ ,  $\Delta\psi$ ) are confined to the Gln43-Ser47 and Asp63-Arg70 loops.

Table 7.7 — Comparison of NCS-related molecules.

The molecules were superposed using the *LSQ\_EXP* command in *O* and the various root mean square deviations were calculated with the program *4D\_RMSDDB*. The comparison included all 120 residues and the core region was defined as residues 5-11, 23-43, 48-61 and 73-111.

Rmsd NCS-related C $^{\alpha}$ atoms (Å)	0.28
Rmsd all NCS atoms (Å)	0.87
Rmsd core C $^{\alpha}$ atoms (Å)	0.23
Residues with $\Delta\phi > 10^{\circ}$ (%)	20
Residues with $\Delta\psi > 10^{\circ}$ (%)	16.7
Rms $\Delta B$ NCS-related C $^{\alpha}$ atoms (Å)	4.39
Rms $\Delta B$ all NCS atoms (Å)	5.72

The environments of molecules A and B are very similar within the crystal. The dimer is stabilised by a number of interactions between molecules A and B (Table 7.8). The major contact regions are the Gln43-Ser47 loop and the C-terminal tail (Glu113-Gln117). In contrast, there are relatively few intermolecular contacts between the dimers in the crystal. The majority of these involve residues Asp14-Gly19 in one molecule and

Table 7.8 — Intermolecular subunit contacts between molecules A and B.

Atom	Atom	Distance (Å)	Atom	Atom	Distance (Å)
Gln2A N $^{\delta}2$	Leu3B O	3.21	Leu3B O	Gln2A N $^{\delta}2$	3.21
His32A N $^{\delta}1$	Wat 138B	3.00	His32B N $^{\delta}1$	Wat 130A	3.44
Gln43A N	Gln117B O $^{\delta}1$	2.97	Gln43B N	Gln117A O $^{\delta}1$	3.01
Gln43A O	Gln117B N $^{\delta}2$	3.20	Gln43B O	Gln117A N $^{\delta}2$	2.99
Asn45A N	Glu113B O $^{\delta}1$	3.19	Asn45B N	Glu113A O $^{\delta}1$	3.13
Ash45A O $^{\delta}1$	Glu113B O $^{\delta}2$	2.87	Glu113B O $^{\delta}1$	Asn45A N	3.19
Glu113A O $^{\delta}1$	Asn45B N	3.13	Gln117B O $^{\delta}1$	Gln43A N	2.97
Gln117A O $^{\delta}1$	Gln43B N	3.01	Gln117B N $^{\delta}2$	Gln43A O	3.00
Gln117A N $^{\delta}2$	Gln43B O	2.99	Wat 138B	Wat 130A	3.35

Arg-70 in a neighbouring molecule. In only one case do the crystal contacts affect the mobility protein. Residues Leu-105–Leu-110 in molecule B appear to be stabilised by a intermolecular hydrogen bond between the N $^{\delta}1$  of His-107 and the carboxylate of Glu-

7B in a neighbouring molecule. Consequently this region has lower overall temperature factors than the corresponding region in molecule A despite having a virtually identical conformation (Figure 7.7).

## 7.4 Guinea-pig LA

### 7.4.1 Refinement

The structure of an orthorhombic form ( $P2_12_12_1$ ) of guinea-pig LA (GPLA), previously solved using MR with the program *MERLOT* (Fitzgerald, 1988; K.R. Acharya, unpublished results), was refined at high resolution. The cell dimensions of the crystals are  $a=32.59\text{\AA}$   $b=64.87\text{\AA}$   $c=49.01\text{\AA}$  and there is one molecule per asymmetric unit.

Although the diffraction data extended to  $1.8\text{\AA}$ , it was fairly incomplete and only data to  $1.9\text{\AA}$  was used in the refinement (88% complete  $\infty - 1.9\text{\AA}$ ). The structure was initially refined using *X-PLOR2.1* but was subsequently briefly re-refined using *X-PLOR3.1* and the Engh and Huber geometric parameters. A summary of the various refinement cycles is given in Table 7.9.

Table 7.9 — Summary of the GPLA refinement.

Cycle	Atoms <sup>a</sup>		Protocol <sup>b</sup>	Resolution Range ( $\text{\AA}$ )	No. of Reflection	<i>R</i> -factor	
	Protein	Other <sup>a</sup>				IN	OUT
1	995	56 + <i>l</i>	P/Pr/S/B/P	8 – 1.9	7694	35.4 <sup>c</sup>	19.7
2	995	58 + <i>l</i>	Pr/P/S/B/P	8 – 1.9	7694	27.9 <sup>c</sup>	18.4
3	995	66 + <i>l</i>	Pr/P/S/B/P	8 – 1.9	7694	19.3	17.5
4	995	71 + <i>l</i>	Pr/P/S/B/P	8 – 1.9	7694	24.9	17.0
5	995	77 + <i>l</i>	Pr/P/B/P/S	8 – 1.9	7694	17.5	15.9
6	995	70 + <i>l</i>	Pr/P/B/P	8 – 1.9	7694	17.4	17.8
7	995	86 + <i>l</i>	Pr/P/B/P	8 – 1.9	7694	19.1	18.2
8	995	73 + <i>l</i>	Pr/P/B/P	8 – 1.9	7694	18.2	17.9

<sup>a</sup> Non-protein atoms are water molecules and calcium ions (*italics*)

<sup>b</sup> Abbreviations used: Pr, *PREPSTAGE*; P, *POS*; S, *SLOWCOOL*; B, *BREF* (individual *B*'s)

<sup>c</sup> *B*-factors reset to  $15\text{\AA}^2$  prior to refinement.



*Cycle 1* — Conventional positional refinement, simulated annealing (4000°K) and restrained, individual  $B$ -factor refinement reduced the  $R$ -factor to below 20%. The majority of the model rebuilding was carried out using *FRODO*. Only minor alterations were required, such as the adjustment of some sidechain torsion angles and the removal of poor water molecules. Attempts were made to define the conformation of the C-terminal tripeptide even though the  $2F_o - F_c$  and  $F_o - F_c$  maps indicated that this region was flexible and adopted at least two distinct conformations.

*Cycles 2 to 5* — Subsequent refinement and rebuilding improved the structure and  $R$ -factor. Two loops in the  $\beta$ -domain (Asn44-His47 and Ser64-Thr66) were poorly defined and required substantial torsional adjustments to locate all the residues in allowed regions of the Ramachandran plot. After cycle 5, the model exhibited reasonable stereochemistry with r.m.s deviations for ideal bond lengths and angles of 0.022Å and 3.34° respectively. These deviations from ideality were acceptable given the nature of the *CHARMM* force and geometric constants used in this version of *X-PLOR*. (Brünger, 1990a).

*Cycles 6 to 8* — The model was refined with a newer version of *X-PLOR3.1* using the Engh and Huber force and geometric constants. As the major part of the refinement was carried out with *X-PLOR2.1*, the free  $R$ -factor was not used to monitor the progress of the refinement. The conventional  $R$ -factor increased slightly but the overall geometry and stereochemistry of the model was considerably improved (Table 7.10). The final model, consisting of all 123 residues, 73 water molecules and one calcium ion, has a  $R$ -factor of 17.9% for all the data between 8 and 1.9Å (7694 reflections).

### 7.4.2 Assessment of the Quality of the Final Model

A summary of the model's stereochemical and geometric statistics is given in Table 7.10. The mean coordinate error is estimated by *SIGMAA* to be 0.19Å (5–1.9Å).

Table 7.10 — Statistics for the final GPLA structure.

Rmsd bond lengths <sup>a</sup> (Å)	0.009
Rmsd bond angles <sup>a</sup> (°)	1.517
Rmsd dihedrals <sup>a</sup> (°)	23.427
Rmsd impropers <sup>a</sup> (°)	1.321
Average overall <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	20.08
Rms $\Delta B$ bonded atoms <sup>b</sup> (Å <sup>2</sup> )	5.34
Average mainchain <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	16.54
Average sidechain <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	21.12
Average solvent <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	36.99
Average calcium <i>B</i> -factor <sup>b</sup> (Å <sup>2</sup> )	8.31
Ramachandran plot,	
Most favoured areas <sup>c</sup> (%)	84.3
Additional allowed areas <sup>c</sup> (%)	14.8
Generously allowed areas <sup>c</sup> (%)	0.9
Disallowed areas <sup>c</sup> (%)	0.0
Non-rotamer sidechain conformations <sup>c</sup> (%)	2.4
Overall G-factor <sup>c</sup>	0.27

<sup>a</sup> Calculated using *X-PLOR* geometric analysis tools

<sup>b</sup> Calculated using *4D\_MOLEMAN*

<sup>c</sup> Defined by *PROCHECK*

#### A. Agreement with the Electron Density

The electron density is continuous along the whole protein backbone apart from a few surface loops. Residues 43 to 47 and 66 to 68 are poorly defined and two breaks in the mainchain electron density occur in these regions. Electron density for the C-terminal tripeptide (121-123) is also particularly weak. Some of the long, hydrophilic sidechains that are exposed on the surface of the protein are also partially disordered (Lys-1, Arg-19, Lys-58, Lys-79, Lys-108, Lys-114 and Gln-117). Nevertheless, the remainder of the protein is extremely well defined (Figure 7.8).

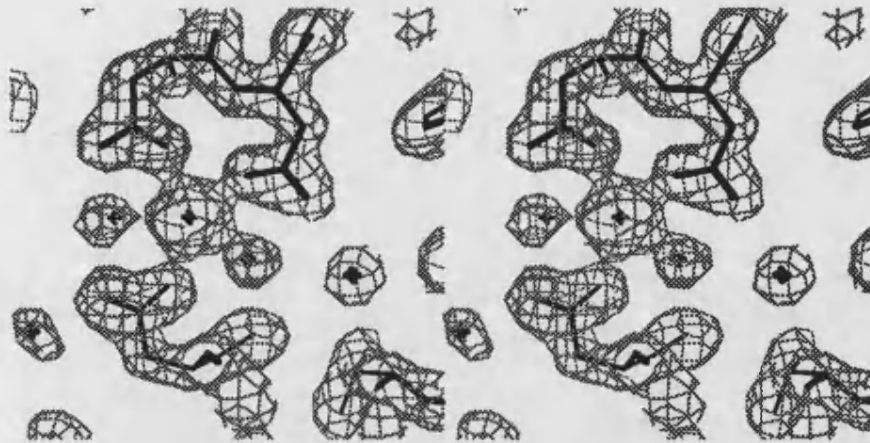


Figure 7.8 — Representative portion of the  $2F_o - F_c$  electron density with the final GPLA structure.

The portion shown corresponds to the calcium binding site. The site is viewed perpendicular to the pentagonal plane and shows the three carboxylate (Asp-82, 87, 88) and the two water ligands. The calcium is in the centre of the plot. The resolution is  $1.9\text{\AA}$  and the map is contoured at  $1\sigma$  (where  $\sigma$  is the root mean square density throughout the unit cell). Crosses indicate the positions of water molecules. Figure produced with *4D\_OPLOTT*.

### B. Ramachandran Plot and General Stereochemistry

Analysis with the program *PROCHECK* shows that the final model has acceptable stereochemistry and geometry. All eleven parameters evaluated are either within (3), or better than (8) the bounds established from well refined structures of an equivalent resolution. None of the non-glycine residues are located in disallowed regions of the Ramachandran plot and only one residue, His-47, is located in a generously allowed region (Figure 7.9). This residue is located at the end of a flexible  $\beta$ -bend that connects strands S1 and S2.

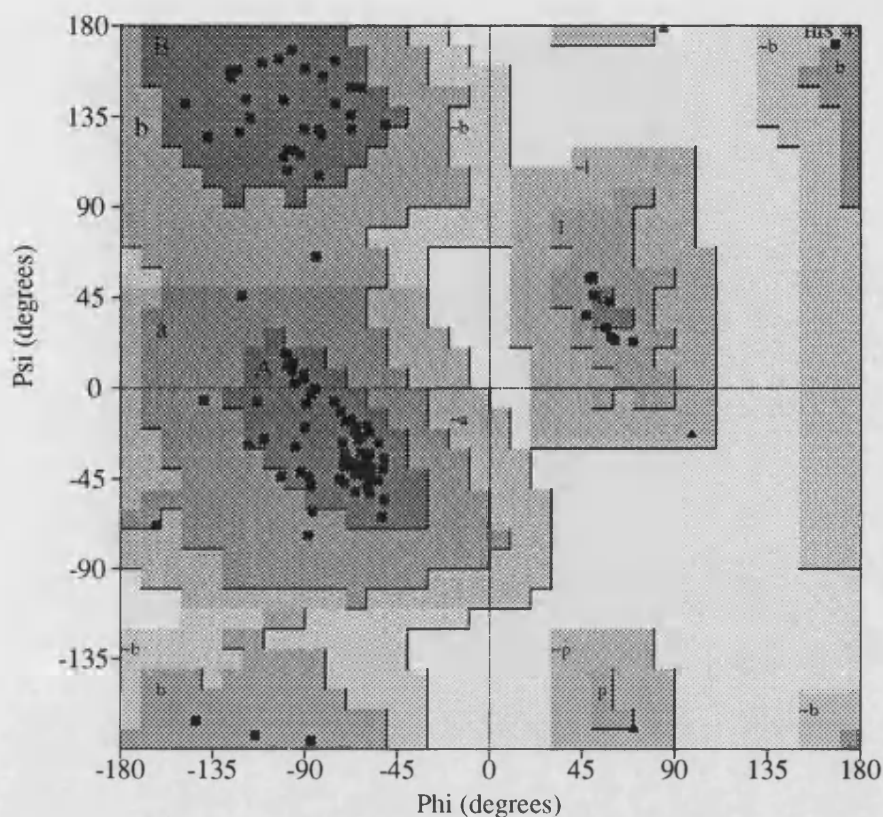


Figure 7.9 — Ramachandran plot for the GPLA structure.

Those residues that fall in the generously allowed regions of the plot are highlighted. The shaded areas are the same as those defined in Figure 7.3. The plot was produced with *PROCHECK* (Laskowski *et al.*, 1993).

### C. Temperature Factors

Figure 7.10 shows a plot of the average mainchain and sidechain *B*-factors for the GPLA structure. The majority of the above average mainchain temperature factors are located in two loop regions (43–47 and 64–69) and the C-terminus (120–123). These parts of the molecule are very exposed to solvent and have similarly high *B*-factors in other LA structures. The remainder of the structure is highly ordered. The most rigid

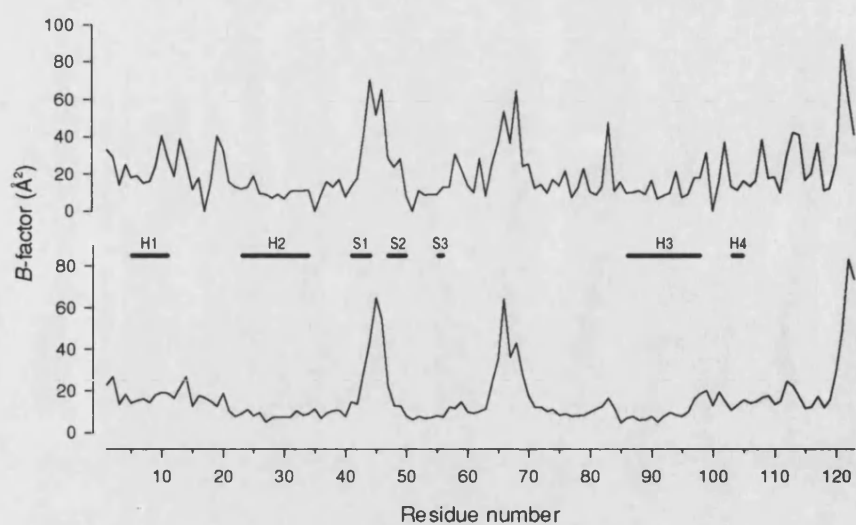


Figure 7.10 — Average temperature factors for the GPLA structure.

The upper panel shows average temperature factors for sidechain atoms, and the lower panel shows average temperature for mainchain atoms, against residue number. The positions of the major secondary structural elements are indicated.

parts of the molecule correspond to regions involved in the  $\alpha$ -domain's hydrophobic core (helix H2 and H3) and the calcium binding site (Lys-79–Asp-88).

## Description of the Structures

### 8.1 Introduction

In the absence of any structural information on the variant LAs, the following discussion is concerned with the structures of mLAs, goat and guinea-pig LA. The structures are examined both in terms of their similarity to those previously determined for LA and also with regard to the function of LA in the lactose synthase complex.

The sequence identity between the bovine (mLA), goat and guinea-pig LAs varies from 64% (GOLA vs GPLA) to 92% (mLA vs GOLA). This range is considerably broader than that exhibited by human and baboon LA (93% identical), for which the crystal structures have been previously determined (Acharya *et al.*, 1989, Acharya *et al.*, 1991), and provides an ideal opportunity to investigate the possible structural heterogeneity that exists between LAs from different species. The lack of immunological cross-reactivity between LAs from different species (Brodbeck *et al.*, 1967; McKenzie and White, 1991) and the interspecies variation in the rate of lactose synthesis (Ley and Jenness, 1970; Khatra *et al.*, 1974; Quarforth and Jenness, 1975) suggests that some structural differences may exist.

## 8.2 Similarity of the Structures

### 8.2.1 Overall Similarity

The overall features of the refined mLA, GOLLA and GPLA structures are similar to those reported for human (HLA) and baboon (BBLA) LA (Acharya *et al.*, 1989; 1991). In the following analysis, the structures of all the LAs determined in the present work are compared with that of HLA. In the case of GOLLA and mLA, where there are more than one molecule in the crystallographic asymmetric unit, the 'subunit' with the lowest overall average temperature factor (molecule A in both cases) was used in the analysis.

#### 8.2.1.1 Secondary Structure

As described in Chapter 1, the LA structure is divided into a large ( $\alpha$ -domain) and small domain ( $\beta$ -domain) by a deep cleft (Figure 8.1). The  $\alpha$ -domain comprises residues 1 to 34 and 86 to 123 and is composed of three major  $\alpha$ -helices (H1: 5–11; H2: 23–34; H3: 86–98). The  $\beta$ -domain comprises residues 35 to 85 and has little regular secondary structure. It is composed of a series of loops, a small antiparallel  $\beta$ -pleated sheet (S1, S2, S3) and a short  $3_{10}$  helix (h2: 77–80).

The major secondary structural elements (H1–3; S1–3; h1b, 2, 3c) are conserved in all the LA structures (Table 8.1). However, a few regular secondary structural

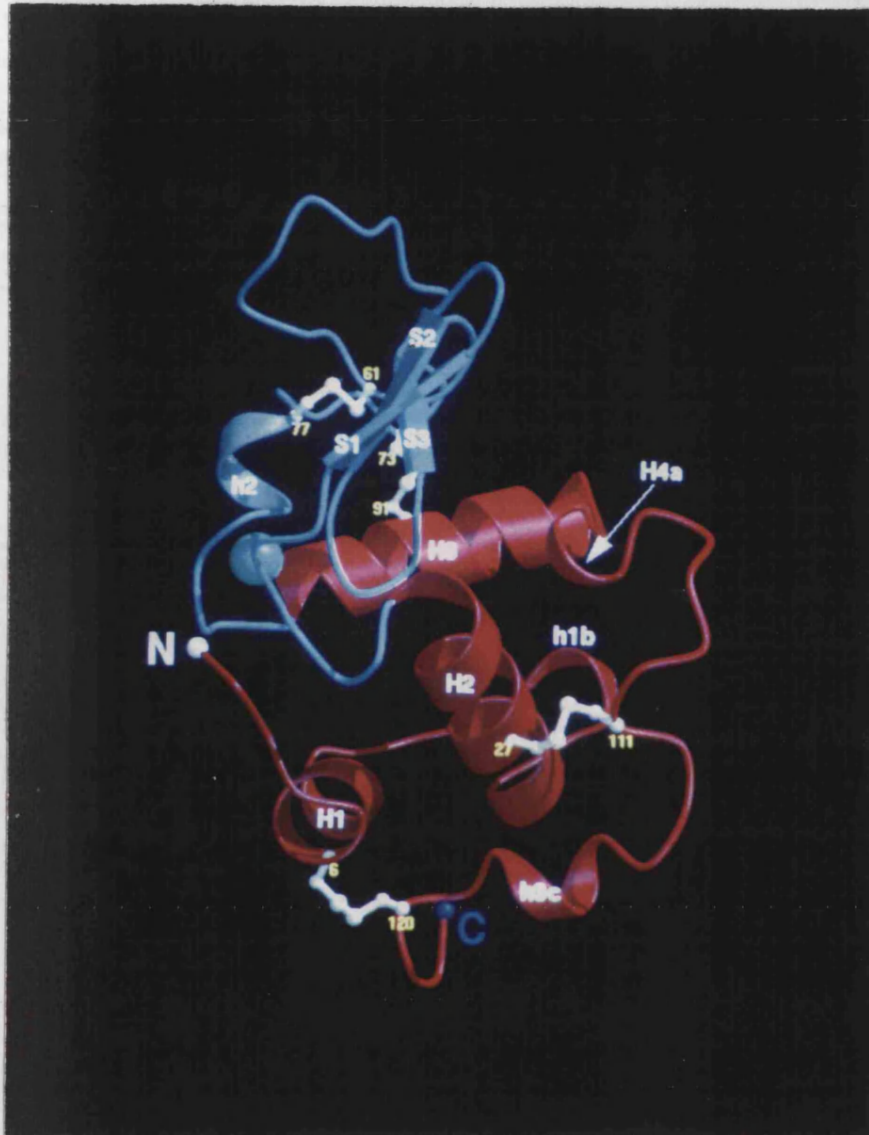


Figure 8.1 — Three-dimensional structure of LA.

The structure of GPLA is schematically represented. The all helical  $\alpha$ -domain is highlighted in red and the  $\beta$ -domain in light blue. The four disulphide bridges (white) and the bound calcium ion (light green sphere) are also shown. The secondary structural elements of GPLA are labelled (see Table 8.1). The figure was created with *MOLSCRIPT* (Kraulis, 1991) and rendered using *RASTER3D* (Merritt and Murphy, 1994).



Table 8.1 — Secondary structural elements in the LA structures.

The secondary structure of GPLA, mLA, GOLA and HLA is that defined by *PROCHECK* (Laskowski *et al.*, 1993) and *DSSP* (Kabsch and Sander, 1983).

	GPLA	mLA	GOLA	HLA
<b>HELIX (H)</b>				
1	5 – 11	5 – 11	5 – 11	5 – 11
2	23 – 34	23 – 34	23 – 34	23 – 34
3	86 – 98	86 – 98	86 – 99	86 – 98
4a	102 – 105	–	–	–
4b	–	105 – 110	–	106 – 110
<b>STRAND (S)</b>				
1	41 – 44	41 – 43	41 – 43	41 – 43
2	47 – 50	48 – 50	48 – 50	48 – 50
3	55 – 56	55 – 56	55 – 56	55 – 56
<b><math>3_{10}</math> HELIX (h)</b>				
1a	–	13 – 15	13 – 15	13 – 15
1b	18 – 20	18 – 20	18 – 20	18 – 20
2	77 – 80	77 – 80	77 – 80	77 – 81
3a	–	101 – 103	101 – 103	101 – 103
3b	–	–	108 – 111	–
3c	115 – 118	115 – 118	115 – 118	116 – 118

elements are not observed in all structures. In particular, the conformation of residues in the region 103–111, and hence the secondary structure, shows considerable variability.

### 8.2.1.2 Tertiary Structure

In general, the structures are very similar and reflect their high level of amino acid sequence identity. The r.m.s. deviation for the 120 equivalent C $^{\alpha}$  atoms ranges from 0.9Å to 1.6Å (Table 8.2). R.m.s. deviations for the core of the molecule, defined as all residues apart from the external loops and the C-terminal tail, are between 0.5Å and 0.9Å.

The differences in mainchain conformation are confined to several distinct regions of the structure. Figure 8.2 highlights the root mean square deviations between the mLA, GOLA, GPLA and HLA structures in terms of the overall conformation of the molecule.

Table 8.2 — Comparison of the various LA structures

The r.m.s deviations and average deviations (in parentheses) are given. In the case of mLA, the additional N-terminal methionine residue was not included in the comparison. Two sets of values have been calculated. Those indicated in **boldface** represent the r.m.s deviation for all 120 C $\alpha$  residues of LA after superposition with the program *ASH* (D.I. Stuart, unpublished program). The values in normal typeface are for the core C $\alpha$  residues calculated using *4D\_RMSPDB*. The core region is defined as residues 5-11, 23-40, 50-61 and 71-104.

	mLA	GOLA	GPLA	HLA
mLA	—	<b>1.14 (0.75)</b>	<b>1.42 (0.86)</b>	<b>0.92 (0.75)</b>
GOLA	0.60 (0.49)	—	<b>1.35 (1.0)</b>	<b>1.35 (1.04)</b>
GPLA	0.49 (0.44)	0.76 (0.67)	—	<b>1.58 (1.10)</b>
HLA	0.69 (0.62)	0.88 (0.79)	0.70 (0.64)	—

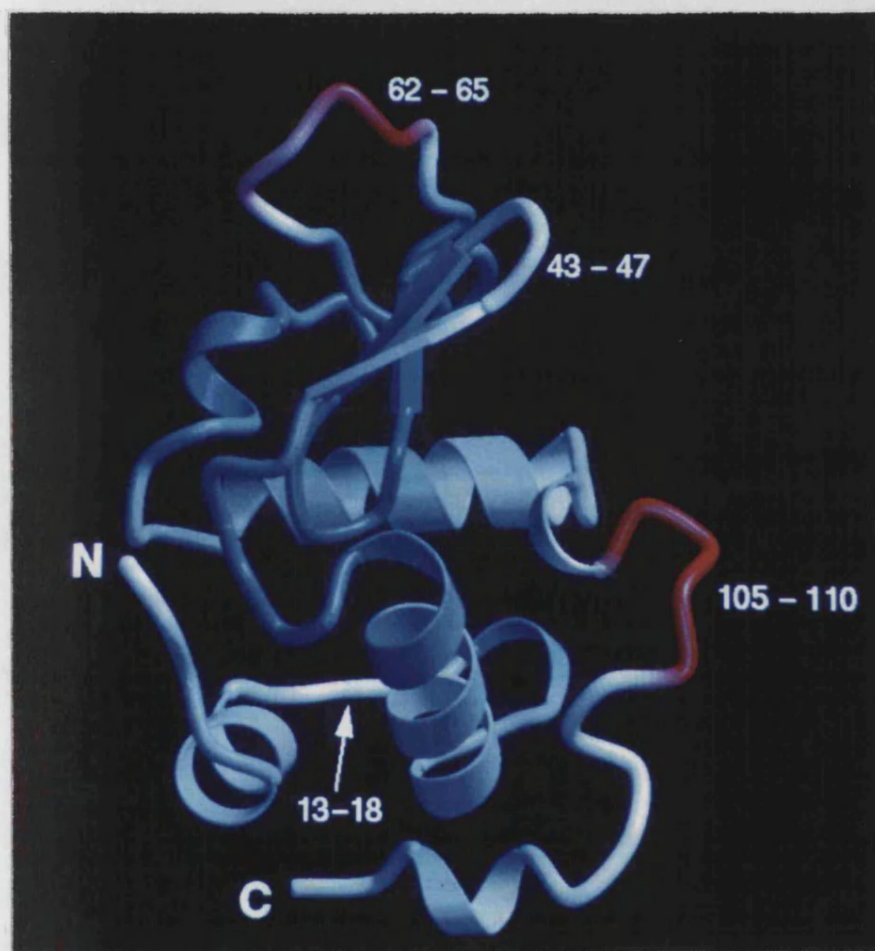


Figure 8.2 — Conformation differences between the LA structures.

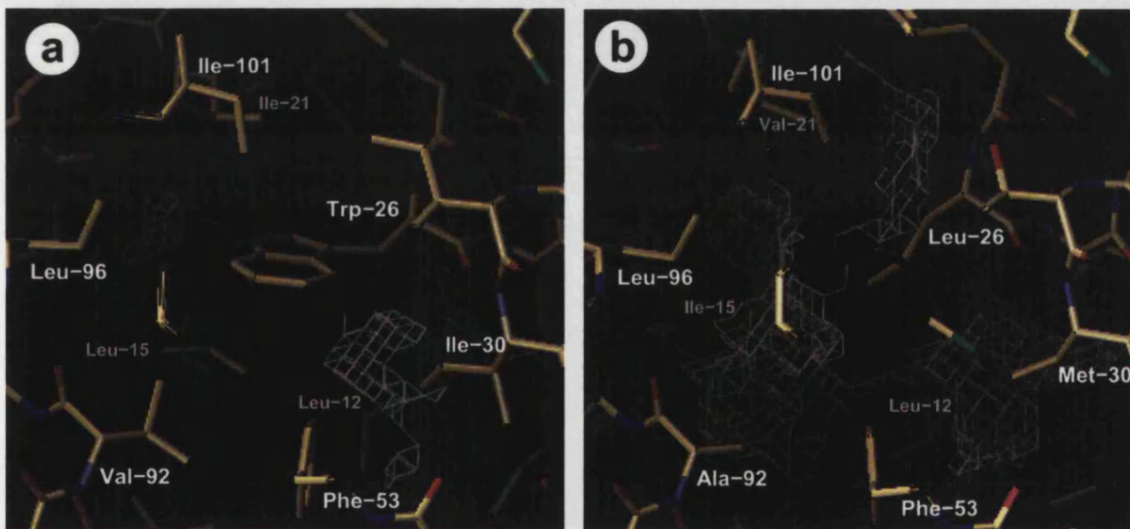
The r.m.s differences between HLA, mLA, GPLA and GOLA are schematically represented in relation to the three-dimensional structure of GPLA. GPLA's secondary structural elements are shaded (blue through red) according to the average C $\alpha$  r.m.s deviations between the four structures. Regions coloured in blue are identical whereas regions in red exhibit considerable heterogeneity in mainchain conformation (C $\alpha$  r.m.s.d > 2.8Å). Regions coloured in white are intermediate. The parts of the structure that exhibit the highest degree of variability are labelled. The figure was created with *BOBSRIPT* (R. Esnouf's modified version of *MOLSCRIPT* (Kraulis, 1991)) and rendered using *RASTER3D* (Merritt and Murphy, 1994).

Variation in the polypeptide backbone occurs in two loops in the  $\beta$ -domain (43-47 and 62-65), the region between helix H1 and H2 (residues 13-18) and in the 'flexible loop' adjacent to the lower end of the cleft (residues 105-111). There are also considerable differences in the conformation of the C-terminal tail, in particular the tripeptide after Cys-120. As mentioned in Chapter 7, the conformation of this region of LA is highly mobile in both crystal structures that have been previously determined (Acharya *et al.*, 1989; 1991). Due to the disordered nature of this part of the molecule, it has not been included in the analysis. The differences in the other regions of the structure are described briefly below.

#### *Residues 13-18*

This region lies between helix H1 and H2 in the  $\alpha$ -domain and contributes several sidechains to the  $\alpha$ -domain's hydrophobic core. The first three residues (13-15) form a  $3_{10}$  helix (h1a) in all the structures apart from GPLA (Table 8.1) and the remaining residues (15-18) adopt a standard type II turn motif. The mainchain conformations of mLA, GOLLA and GPLA are very similar but HLA exhibits a marked difference in backbone topology. This probably results, at least in part, from changes in the sidechain character of the buried residues of HLA in this region. The replacement of Trp-26 (GOLLA, GPLA, mLA) in HLA by a leucine results in a rearrangement of part of the  $\alpha$ -domain's hydrophobic core. Although the reduction in sidechain size at position 26 in HLA is compensated for by an increase in the sidechain size at position 30 (Met in HLA compared to Thr, Ala, Ile in mLA, GOLLA, GPLA respectively), the packing of the sidechains in the cluster is subtly different. In HLA, the  $3_{10}$  helix (h1a) does not pack as close to the protein as in the other LA structures. Consequently, the core of the  $\alpha$ -

-domain, involving a number of hydrophobic residues (including Leu-12, Leu/Ile-15, Val/Ile-21, Trp/Leu-26, Phe-53, Val/Ala-92, Leu-96 and Ile-101), is not as compact in HLA as it is in the mLA, GOLLA and GPLA structures (Figure 8.3). The observation that the BBLA structure has an identical conformation to HLA in this region suggests that this effect is entirely sequence dependent (BBLA exhibits identical amino acid sequence changes at positions 26 and 30). An identical double substitution (W26L and X30M) is present in the wallaby LA amino acid sequence (Shewale *et al.*, 1984) and a similar local conformation to that of HLA is predicted for this region.



**Figure 8.3 — Packing differences in the core of LA.**

The hydrophobic core packing around residue 26 is schematically represented for a) GPLA and b) HLA. The white chicken wire mesh denote cavities in the core of the proteins. The pseudo cavities were calculated with the program *VOIDOO* (Kleywegt and Jones, 1994) using a 0.5Å accessible probe radius. Residues in the  $\alpha$ -domain core are labelled.

*Residues 43-47 and 62-65*

Both these regions are located in the  $\beta$ -domain and form flexible, solvent exposed loops. Residues 43 to 47 forms a  $\beta$ -bend that links S1 and S2 while residues 62 to 65 form a type III turn at the top of the cleft (Acharya *et al.*, 1989). In all the structures, at least one of these loops is involved in crystal contacts and this accounts for the variation in mainchain conformation.

*Residues 105-111*

This region lies adjacent to the lower end of the cleft and interacts with the C-terminal end of helix H2. It adopts two distinct conformations in the LA structures determined to date. In mLA and HLA, the polypeptide backbone adopts a distorted  $\alpha$ -helical conformation (H4b). In contrast, the mainchain is looped out away from the body of the protein and exposed to solvent in the GPLA and GOLA crystal structures (Figure 8.4).

The observed structural heterogeneity in this region is associated with the conditions used to obtain crystals. At high pH (6.5–8.0), the helical conformation is observed (HLA and mLA) whereas at low pH (4.6), the looped out form predominates (GPLA and GOLA). This pH effect has been previously noted in HLA and is associated with His-107 (Harata and Muraki, 1992). In the low pH crystal forms, complete burial of a fully protonated histidine would be very unfavourable and therefore a helix-to-coil transition occurs to locate this residue in the bulk solvent. The ‘looped out’ conformations are remarkably well-defined in the electron density of the GPLA and GOLA structures. In GPLA, the hydrogen-bonded loop is stabilised by extensive crystal contacts involving His-107, Lys-108 and Leu-110 (see Table 8.3). However, the

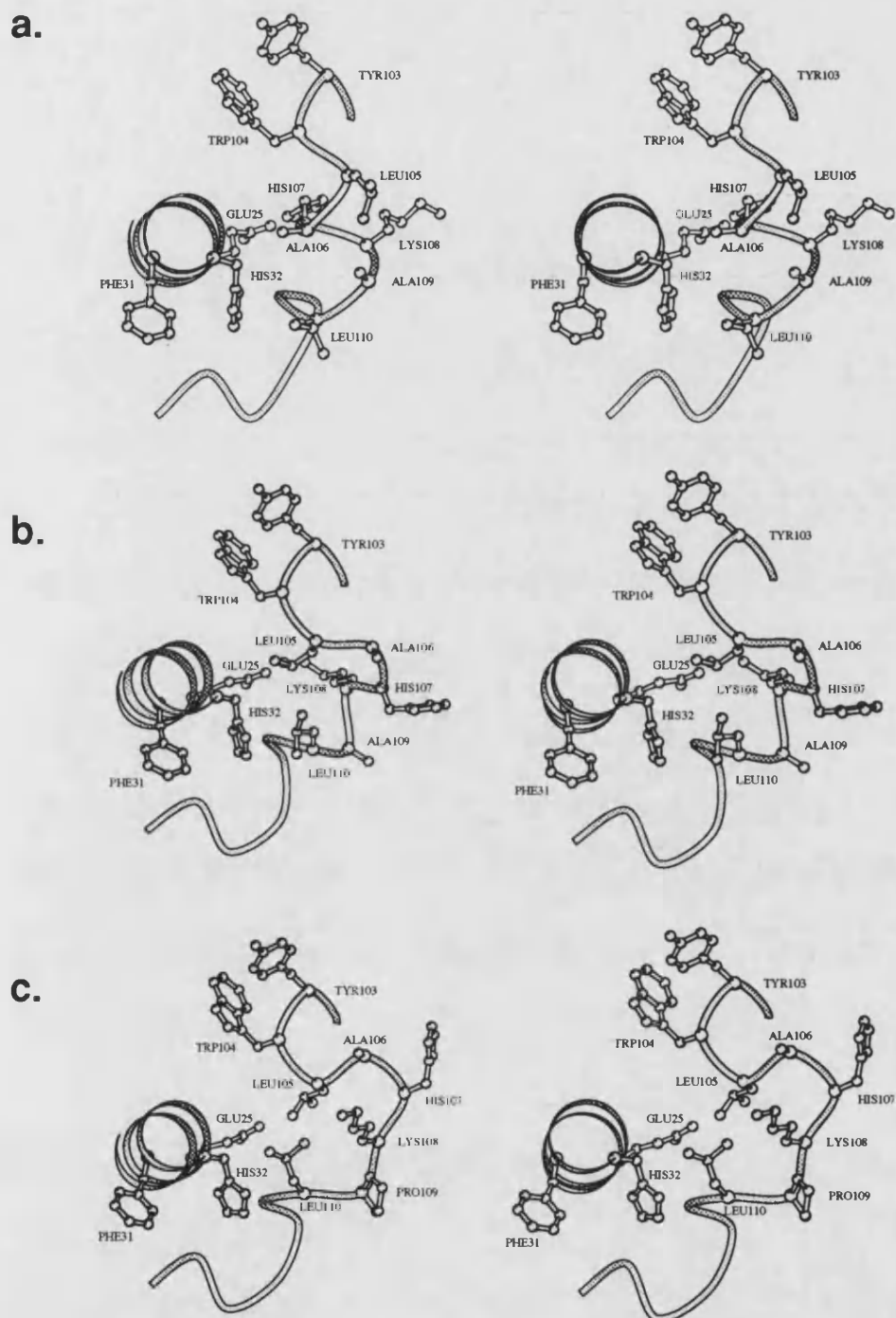


Figure 8.4 — Conformation of residues 105–110.

The stereo plots of conformation of the 105–110 region are schematically shown for **a.** HLA, **b.** GOL and **c.** GPLA. All three molecules are viewed in the same orientation. The figure was produced by *MOLSCRIPT* (Kraulis, 1991).

Table 8.3 — Protein-protein intermolecular contacts in the GPLA crystal.

Atom	Symm	Atom	Distance (Å)	Atom	Symm	Atom	Distance (Å)
Asn13 O <sup>δ</sup> 1	4-A+C	Glu62 O <sup>ε</sup> 1	2.96	Arg70 NH1	2+B	Leu110 O	2.62
Asp14 O <sup>δ</sup> 2		Asp78 O <sup>δ</sup> 1	2.73	Asn71 O		His107 N <sup>δ</sup> 1	2.56
Arg19 O	2+B-C	Gln68 N <sup>ε</sup> 2	3.06	Cys73 O	4-A+C	Gln39 N <sup>ε</sup> 2	3.09
Arg19 O		Gln68 O <sup>ε</sup> 1	2.99	Asp74 O		Gln39 N <sup>ε</sup> 2	2.75
His32 N <sup>ε</sup> 2	4	Ala 122 O	2.60	Asp78 O <sup>δ</sup> 1	4	Asp14 O <sup>δ</sup> 2	2.73
Gln39 N <sup>ε</sup> 2	4+C	Asp74 O	2.75	Asp78 O <sup>δ</sup> 2		Lys93 N <sup>ε</sup>	3.08
Gln39 N <sup>ε</sup> 2		Cys73 O	3.09	Asp84 O <sup>δ</sup> 2		Tyr50 OH	2.62
Tyr50 OH	4+C	Asp84 O <sup>δ</sup> 2	2.62	Leu85 N		Ser64 O	3.05
Glu62 O <sup>ε</sup> 1		Asn13 O <sup>δ</sup> 1	2.96	Thr86 O <sup>γ</sup> 1		Ser64 N	2.99
Ser64 N		Thr86 O <sup>γ</sup> 1	2.99	Lys93 N <sup>ε</sup>		Asp78 O <sup>δ</sup> 2	3.08
Ser64 O		Leu85 N	3.05	Asp102 O <sup>δ</sup> 1	2+B-C	Ser69 O	2.81
Glu68 O <sup>ε</sup> 1	2+B	Arg19 O	2.99	His107 N <sup>δ</sup> 1		Asn71 O	2.56
Glu68 N <sup>δ</sup> 2		Arg19 O	3.06	Lys108 O		Arg70 NH1	2.83
Ser69 O		Asp102 O <sup>δ</sup> 1	2.81	Leu110 O		Arg70 NH1	2.62
Arg70 NH1		Lys108 O	2.83	Ala122 O	4-A	His32 N <sup>ε</sup> 2	2.60

Symmetry operators: (1) x, y, z; (2) 1/2-x, -y, 1/2+z; (3) -x, 1/2+y, 1/2-z; (4) 1/2+x, 1/2-y, -z

observed helix-to-coil transition is not solely driven by crystal packing forces as the analogous loop in GOLA does not participate in any crystal contacts.

Although the helical conformations of mL A and HLA are identical, the two loop structures show considerable heterogeneity (Figure 8.4). This reflects both sequence differences and the degree to which the loop conformations are stabilised in GPLA and GOLA structures. In GPLA, the polypeptide backbone forms a single turn of  $\alpha$ -helix (H4a: Asp102–Leu105) and then loops out to form an exposed hydrogen-bonded turn (Ala106–Cys111). In GOLA, the polypeptide backbone adopts a similar initial helical conformation (h3a: Ile101–Tyr103), loops in a short bend (Leu105–His107) before returning to a helical path (Lys108–Cys111). The conformation of the 105-110 region in GOLA can therefore be considered to be intermediate between that of mL A (and HLA) and GPLA. The underlying reason for the subtle differences in mainchain conformation between the low pH forms of GPLA and GOLA is probably sequence related. Although the amino acid sequence of all LAs is highly conserved in the 104-111 region, some variability occurs at position 109. Bovine, ovine, caprine and human LAs have an alanine



at this position whereas all others sequenced to date have a proline. Pro-109 in GPLA probably restricts the flexibility of the polypeptide backbone, due to the constrained nature of its backbone dihedral angles, and accounts for the differences between the conformations of the C-terminal portion of the 105-110 region in GPLA and GOLA. However, the presence of a proline at position 109 is not expected to preclude helix formation in the 105–110 region of GPLA. Modelling studies demonstrate that the backbone dihedral angles in the helical conformation ( $\phi -65^\circ$ ,  $\psi -47^\circ$ ) observed in mLAs and HLAs are compatible with a proline residue.

Given the overall similarity of the structures, it is not clear why LAs from different species exhibit a lack of immunological cross-reactivity. However, the surface characteristics of the various LAs do show some variation, particularly in the small  $\beta$ -domain which contains the major antigenic determinants in the homologous LYZs (Benjamin *et al.*, 1984). The high degree of variability in this part of the LA sequence (residues 61 to 77) appears to be mainly responsible for generating the observed lack of antigenic response.

### 8.2.2 Conformation of the Calcium Binding Site

In general, the calcium binding site in the various LA structures is very similar both in terms of conformation and ligand coordination (Table 8.4). The calcium is cradled in a helix-turn-helix motif that spans the interface between the  $\alpha$ - and  $\beta$ -domains. The ion is coordinated by a constellation of seven oxygen groups that form a distorted pentagonal bipyramid. Five of the seven oxygen ligands are contributed by the protein mainchain and sidechains. One of these ligands originates from helix h2 (Lys-79), two



are from the Asp turn (Asp-82 and 84) and two from helix H3 (Asp-87 and 88). In addition to the five protein ligands, the calcium is also coordinated by two water

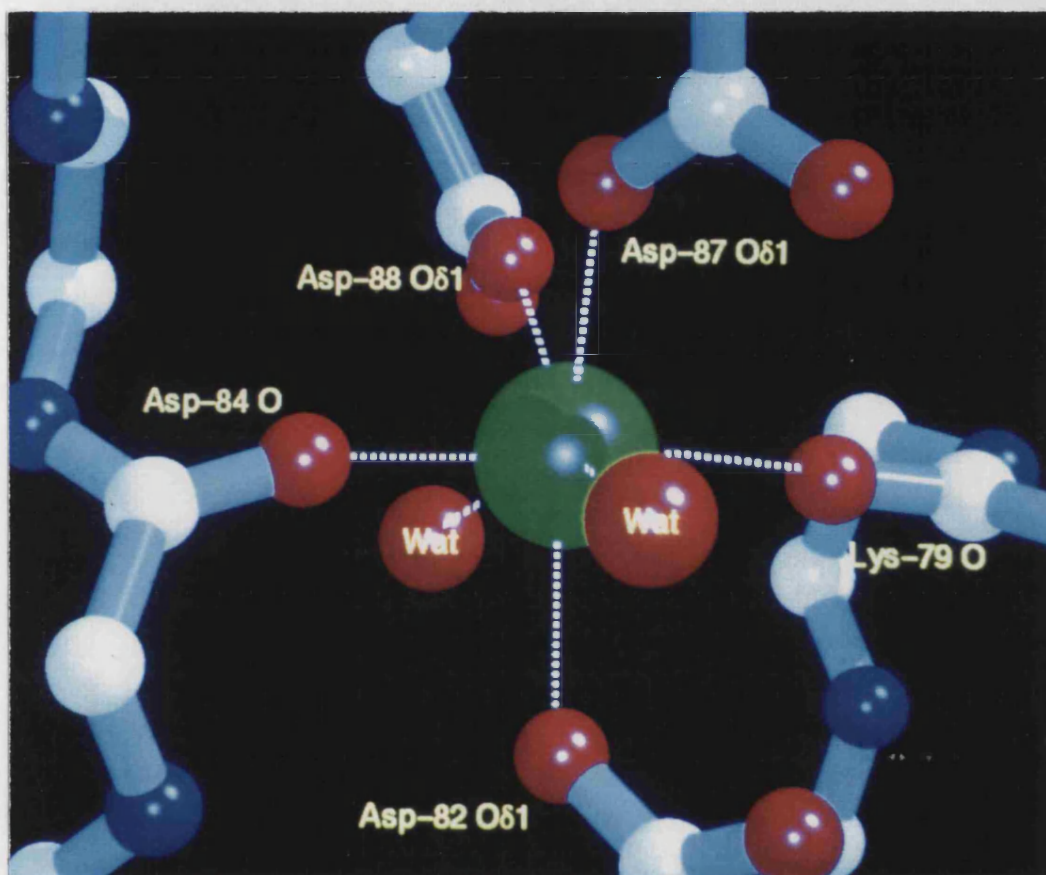


Figure 8.5 — Calcium binding site of LA.

The calcium binding site of GPLA is viewed along the pentagonal plane. The calcium ion (green sphere) is shown along with the protein and solvent ligands (light red spheres). The atoms that coordinate the ion are labelled. The figure was created with *MOLSCRIPT* (Kraulis, 1991) and rendered using *RASTER3D* (Merritt and Murphy, 1994).

molecules. The three carboxylate groups (Asp-82, 87 and 88) and the two water molecules form the pentagonal plane that includes the  $\text{Ca}^{2+}$  and the mainchain carbonyl groups of Lys-79 and Asp-84 are positioned at the apices of the bipyramidal arrangement above and below the central plane (Figure 8.5). All the protein ligand-calcium distances

are between 2.15–2.5 Å. This geometry is typical of that observed in other calcium-binding proteins (Strynadka and James, 1989).

The stabilisation of the three-dimensional structure of LA that is conferred by calcium binding is reflected by the temperature factors in the vicinity of the calcium binding loop. This region is the most rigid part of the HLA (Acharya *et al.*, 1991),

Table 8.4 — Calcium – ligand distances and temperature factors.

Residue	Ligand	Calcium – ligand distance Å (B-factor Å <sup>2</sup> )			
		mLA	HLA	GPLA	GOLA
Lys-79	O	2.15 (39.6)	2.34 (8.4)	2.24 (10.8)	2.30 (14.0)
Asp-82	Oδ1	2.40 (35.4)	2.35 (11.0)	2.42 (12.1)	2.50 (34.4)
Asp-84	O	2.24 (32.7)	2.24 (5.8)	2.16 (10.3)	2.28 (20.3)
Asp-87	Oδ1	2.49 (30.8)	2.40 (10.4)	2.44 (10.1)	2.41 (27.9)
Asp-88	Oδ1	2.59 (33.1)	2.42 (7.3)	2.33 (10.8)	2.47 (12.5)
WAT (int)	OW	— <sup>a</sup>	2.30 (7.9)	2.44 (8.6)	— <sup>a</sup>
WAT (ext)	OW	— <sup>a</sup>	2.50 (19.1)	2.33 (21.8)	— <sup>a</sup>
Calcium	—	(35.2)	(6.2)	(8.3)	(17.5)

<sup>a</sup> In mLA and GOLA, the solvent ligands are not well defined and were therefore not included in the final structures

GOLA and GPLA structures (see Figures 7.7 and 7.10). The differences in the relative atomic temperature factors of the calcium ligands in the HLA, GOLA and GPLA structures reflect the different resolutions at which the structures were determined.

In the mLA structure, the stabilising effects of the bound ion are less apparent and the calcium binding site in all four molecules is surprisingly mobile. The temperature factors for the calcium ion ranges from 35 to 63 Å<sup>2</sup> and the coordinating ligands also have relatively high *B*-values. These values are consistent with the overall disorder observed in the mLA structure. In all four molecules of monoclinic mLA structure, the calcium binding site is clearly occupied by an ion that is coordinated by five protein ligands. The

two additional solvent ligands are not apparent at the resolution (2.3Å) at which the structure was determined.

One concern during the initial stages of model refinement was to what extent the high concentrations of divalent cation ( $Mg^{2+}$ ) required to grow the monoclinic mLA crystals would influence the recombinant LA's structure. Magnesium, along with a large number of metal ions, binds to LA with an affinity in the millimolar range (Kronman, 1989). It is unlikely, however, that mLA's metal binding loop is occupied by magnesium rather than calcium ions. Firstly, LA's high affinity for calcium suggest that this ion has remained bound throughout the preparative and crystallisation procedure. Furthermore, although magnesium does bind to apo-LA, it does not appear able to either displace calcium from native holoprotein or alter the native conformation (Permyakov *et al.*, 1981b; Kronman, 1989). Secondly, there is no evidence from the coordination state of the bound cation that suggests it is magnesium rather than calcium. Magnesium has a smaller atomic radius than calcium and its coordination is characterised by shorter ligand-metal distances (about 2.1Å) and a coordination number of six rather than seven (Strynadka and James, 1989).

### 8.2.3 Solvent Structure

It is difficult to make a detailed comparison of the solvent structures of the four LAs as only two (HLA and GPLA) have been determined to high resolution. The positions of at least two internal waters are conserved in all the LA structures. One water (water 2, see Figure 8.6) hydrogen bonds to four protein ligands (O Thr-38, N Leu-52, O Phe/Leu-80, O<sup>δ</sup>2 Asp-88); the other (water 3) hydrogen bonds to three

groups (O Tyr-50, N Phe-53, O<sup>δ2</sup> Asp-88). This high degree of interaction with the protein is reflected by the relatively low temperature factors exhibited by these two water molecules. Both these water molecules are located in a solvent channel that runs along the domain interface. This channel originates at the calcium binding site and runs along helix H3 for a distance of between 10Å and 14Å. The channel also includes one of the water molecules that is involved in calcium ligation (water 1 in Figure 8.6). The three water molecules seem to be trapped inside the protein and are probably not free to exchange with the bulk solvent. Given that an identical solvent channel is present in the crystal structure of HEWL, it seems likely that this feature fulfils a general structural role in LA and LYZ.

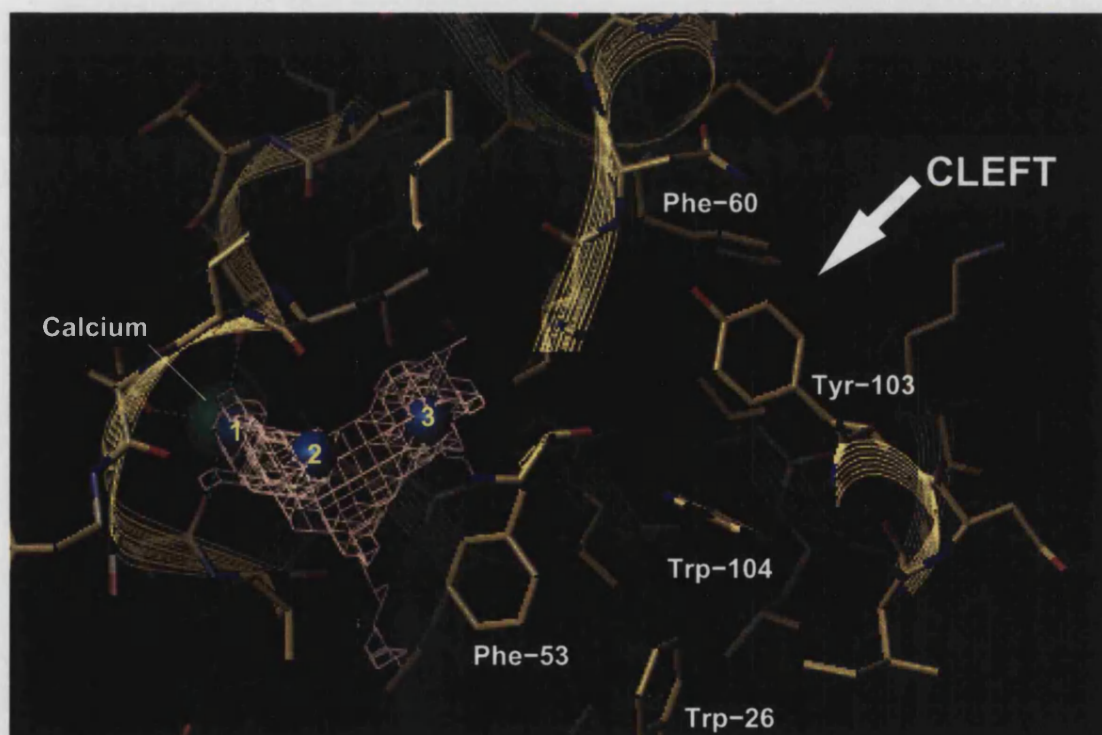


Figure 8.6 — Structural solvent channel.

The solvent channel that runs from the calcium binding site towards the cleft region in GPLA is shown. The calcium ion is shown as a green sphere and the three water molecules as blue spheres. The water molecules are numbered from one to three (see text for details). The cream-coloured chicken wire mesh denotes the extent of the channel. This was calculated with the program *VOIDOO* (Kleywegt and Jones, 1994) using a 0.5Å accessible probe radius. The backbone conformation of GPLA is schematically represented by the yellow ribbon.

## 8.3 Functional Regions in LA

### 8.3.1 Aromatic Clusters I and II

Aromatic clusters I and II have a similar conformation in all the LA structures.

The site-directed mutagenesis studies on mL<sub>A</sub> (Chapter 2) have highlighted the importance of the residues in aromatic cluster I in LA function. This cluster comprises of the invariant residues Phe-31, His-32, Gln-117 and Trp-118. Phe-31 and His-32 are located at the end of helix H2 immediately adjacent to the lower reaches of the cleft and Gln-117 and Trp-118 are located in the flexible C-terminal tail of LA. The cluster is flanked by Tyr-36 and lies on top of the Cys28–Cys111 disulphide bridge. Consequently, the cluster is one of the most rigid parts of the LA structure and all four residues exhibit low temperature factors. The Cys28–Cys111 disulphide bridge appears to be required for maintaining the structural integrity of the cluster as its removal eliminates LS activity (K. Brew, personal communication).

The environments of the four residues in the cluster are quite different. Phe-31 and Gln-117 are quite exposed to solvent (40-70% of surface area accessible). In contrast, His-32 and Trp-118 are somewhat shielded from solvent by the surrounding residues. The low  $pK_a$  values (5.18–6.06) reported for His-32 in solution (Alexandrescu *et al.*, 1992) result from its close contact with Phe-31 and Leu-110. In the four structures, there are only small differences between the orientations of the sidechains despite their location on the surface of LA (Figure 8.7). Phe-31 exhibits some rigid-body movements in the direction of the ring plane. The orientation of imidazole ring of His-32 appears to be primarily influenced by the conformation of the flexible loop region (see section 8.3.2).





Figure 8.7 — Conformation of Aromatic cluster I.

The relative orientation of the sidechains in aromatic cluster I are shown for HLA (blue), mLA (red), GPLA (green) and GOLA (yellow). The structures were superposed using all 120 equivalent  $C^{\alpha}$  atoms using the program *ASH* (D.I. Stuart, unpublished program). The flexible loop region (upper right) and Leu-110 are also shown. The figure was produced by *MOLSCRIPT* (Kraulis, 1991).

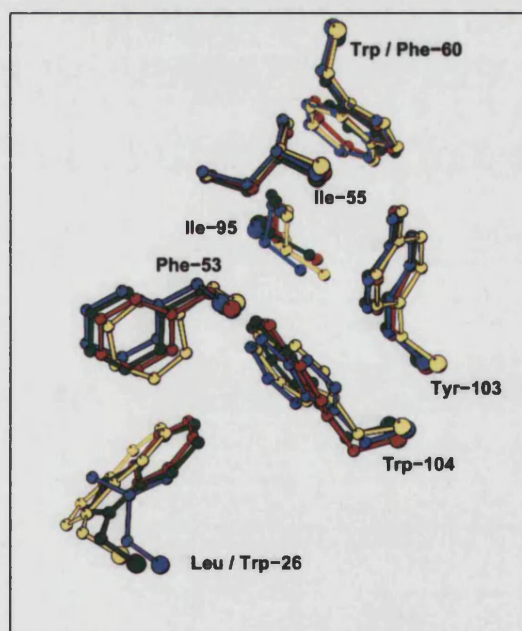


Figure 8.8 — Sidechain orientation in Aromatic cluster II.

The relative orientation of the sidechains in aromatic cluster II are shown for HLA (blue), mLA (red), GPLA (green) and GOLA (yellow). Only the sidechain atoms are shown. The structures were superposed using all 120 equivalent  $C^{\alpha}$  atoms using the program *ASH* (D.I. Stuart, unpublished program). The figure was produced by *MOLSCRIPT* (Kraulis, 1991).

The sidechains in aromatic cluster II (hydrophobic box) adopt identical orientations (Figure 8.8). This is consistent with the cluster's role in the folding and stabilisation of the tertiary fold of LA (Smith *et al.*, 1994). This buried cluster forms part of a more extensive hydrophobic core that stretches from the middle of the cleft into the centre of the  $\alpha$ -domain. This closely packed structural unit involves residues from both the helical  $\alpha$ -domain and the loops of the  $\beta$ -domain. It comprises residues Phe-53 ( $\beta$ ), Ile-55 ( $\beta$ ), Phe/Trp-60 ( $\beta$ ), Ile-95 ( $\alpha$ ), Tyr-103 ( $\alpha$ ) and Trp-104 ( $\alpha$ ). The sidechain of Tyr-103 projects into the cleft and is sandwiched between Trp-104 and the position 60 aromatic (either Trp or Phe). The sidechains of the cluster make hydrogen bonds in addition to their Van der Waals contacts. The phenolic OH of Tyr-103 is hydrogen-bonded to the mainchain nitrogen of Asn/Ser-56 and the ring nitrogen of Trp-104 is hydrogen-bonded to the carbonyl oxygen of Phe-53.

### 8.3.2 Flexible Loop Region

The conformation in this part of LA (residues 105–110) is particularly interesting in terms of LA function, given that it comprises part of the GT interaction site. Amino acid substitutions at positions 106, 107 and 110 primarily influence LA's ability to bind to GT (see Chapter 2). The crystal structures demonstrate that this region is particularly mobile and can adopt a number of different conformations depending on the conditions used to obtain crystals. Although this region exhibits a few well-defined conformations in the crystalline state, it seems likely that it is highly fluctuating in solution. This is

highlighted by the fact that both the helical and loop conformer crystal forms of HLA are significantly populated at pH 6.5 and room temperature (Harata and Muraki, 1992).

The two distinct conformations exhibited by the LA crystal structures have dramatic effects on the sidechain environments of the residues in this region. In the helical conformation, Leu-105 and Leu-110 are relatively exposed to solvent (60–90% surface area accessible to solvent). In contrast, these positions are completely buried in the GPLA structure (0–17% surface area accessible to solvent). Due to the intermediate nature of the GOLLA conformation in the ‘flexible loop’, Leu-105 is buried and Leu-110 is relatively exposed (50% accessible). The environment of His-107, the residue implicated in the helix-to-coil transition, is also radically different in the two conformations. In the helical conformer, the histidine is totally buried and forms a hydrogen bond (N<sup>ε</sup>2) with the carboxylate of Glu-25 whereas in the GPLA and GOLLA structures this sidechain is very exposed. Another consequence of the observed helix-to-coil transition, is that it leads to a subtle change in the environment of His-32. In GPLA, about 55% of the surface area of His-32 is solvent accessible; this compares to values of 43, 40 and 37% for His-32 in GOLLA, mLHA and HLA respectively. This difference is due to a change in the relative orientation of Leu-110 in the flexible loop region (see Figure 8.4 and 8.7). In the HLA, mLHA and GOLLA crystal structures the exposed sidechain of Leu-110 makes significant Van der Waals contacts with the imidazole ring of His-32. In the GPLA structure, however, Leu-110 is buried and as a result His-32 is considerably more accessible to solvent.

The slow exchange rates of His-107 indicate that this residue is partially buried in solution (Bradbury and Norton, 1975; Alexandrescu *et al.*, 1992). However, the pK<sub>a</sub>



value measured for this residue suggests that its environment in LA structure does not influence its innate electrostatic properties (Alexandrescu *et al.*, 1992). These results are difficult to reconcile with the known conformations of the 105–110 region in the various LA structures. His-107 is buried in the high pH conformer but its interaction with Glu-25 would undoubtedly affect its  $pK_a$ . Further studies are clearly required to define the conformation of this part of the LA molecule in solution. It is possible that the conformation of this region in the LS complex is different from that observed in any of the LA crystal structures. The conformation of this region is constrained to some extent by the tertiary fold of LA. Nevertheless, the subtle differences exhibited by the GOLLA and GPLLA crystal structures suggest that there is some scope for conformational adjustment in the 105-110 region. The inherent flexibility of this region may have some functional importance. Certainly the proximity of this region to residues implicated in glucose binding (Phe-31 and His-32) and its influence on the conformation of the cleft could be important for either the formation of the LS complex or the subsequent molecular processes.

### 8.3.3 The Cleft Region

Certain substitutions at Phe-31, His-32 and Leu-110 perturb LA's ability to promote monosaccharide binding in the LS complex. These residues are located in the subsite F region of HEWL and are homologous to residues that have been implicated in binding one of the saccharide units.

The active-site cleft region of LYZ consists of series of subsites (A–F) that are each capable of binding one saccharide unit of the hexasaccharide substrate (Blake *et al.*,

1967a,b). Crystallographic studies on the binding of GlcNAc, (GlcNAc)<sub>3</sub>, N-acetylmuramic acid (NAM) and NAM-GlcNAc polysaccharides to HEWL has provided detailed information about sugar binding in subsites A, B, C and D (Blake *et al.*, 1967b; Perkins *et al.*, 1978; Strynadka and James, 1991; Hadfield *et al.*, 1994). Until recently, only limited structural information has been obtained about saccharide binding in the E and F subsites. The nature of the protein-monosaccharide interactions in these sites have mainly been inferred from molecular modelling studies and the results of transglycosylation reactions (Blake *et al.*, 1967b; Imoto *et al.*, 1972; Pincus and Scheraga, 1981; Post *et al.*, 1986). The binding of saccharides in the E and F subsites is believed to provide the energy required for converting non-productive substrate binding into the productive binding required for catalysis (Holler *et al.*, 1975; Perkins *et al.*, 1981; Weaver *et al.*, 1995).

The structural nature of saccharide binding in the E and F subsites was highlighted in a recent crystallographic study of human lysozyme (HLYZ) co-crystallised with hexa-*N*-acetyl-chitohexaose (Song *et al.*, 1994). Although the hexasaccharide had been cleaved, (GlcNAc)<sub>2</sub> was observed to bind in a region close to the E and F subsites proposed on the basis of model building. Modelling and energy minimisation studies on saccharide binding mode in subsite F indicate that the protein-saccharide interactions involve the guanidino group of Arg-114 (with sugar ring O and β1 OH), the mainchain oxygen of Phe-34 and the sidechain nitrogen of Asn-37 (with primary alcohol O6) (Blake *et al.*, 1967b; Pincus and Scheraga, 1981). The role of Asn-37 is indirectly supported by site-directed mutagenesis studies as replacement of this residue alters HEWL's substrate-binding mode (Kumagai *et al.*, 1993). The subsite F' ligands in the human LYZ /

hexasaccharide crystal structure are similar and involve the mainchain carbonyls of Ser-36 (with primary alcohol O6) and Trp-34 (with O3 hydroxyl) and the N<sup>ε</sup> of Lys-33 (with O7 of the *N*-acetyl group) (Song *et al.*, 1994). These different binding modes exhibited by the model and crystal structures may reflect the intermediate state of the HLYZ / product complex. When the LA and LYZ crystal structures are superimposed with *SHP* (Stuart *et al.*, 1979), Phe-31 (LA) is structurally equivalent to Lys-33 (HEWL / HLYZ), His-32 (LA) to Phe/Trp-34 (HEWL / HLYZ) and Leu-110 (LA) to Arg-114 (HEWL). Although there is some uncertainty about the exact location of subsite F, this functional correlation with LYZ, combined with the fact that substitutions at any of these residues in LA perturbs its ability to promote monosaccharide binding in the LS complex, provides strong evidence that a region analogous to the F subsite participates in the binding of monosaccharides in the LS complex.

The overall conformation of the cleft around subsite F is similar in the LA and HEWL structures. The subsite is located at the lower end of the cleft where it broadens out on the protein's surface and, unlike the C and D subsites, is relatively shallow and accessible to solvent. In LA, the shape of cleft around subsite F is influenced, to some extent, by the conformation of the flexible loop (105-110) which flanks the pocket (Figure 8.9). In the GPLA structure, the sidechains of Lys-108 and Pro-109 in the loop conformer form a slight 'wall' on the external (right-hand) side of the subsite (Figure 8.9). In contrast, the subsite is more accessible when residues 105-110 adopt a helical conformation in HLA and mLA (inset in Figure 8.9). The functional significance of this difference, however, is difficult to quantify due to the unknown nature of the monosaccharide binding site in the LS complex. It is unlikely that the monosaccharide

binding mode in the F subsite of LYZ and the LS complex are the same and it would be naive to assume that this region functions in an identical fashion in both proteins. Given that GT is capable of binding glucose, albeit weakly, it seems likely that the sidechains of Phe-31 and His-32 provide stabilising interactions for the bound monosaccharide in the LS complex as suggested by the monosaccharide bridge model.

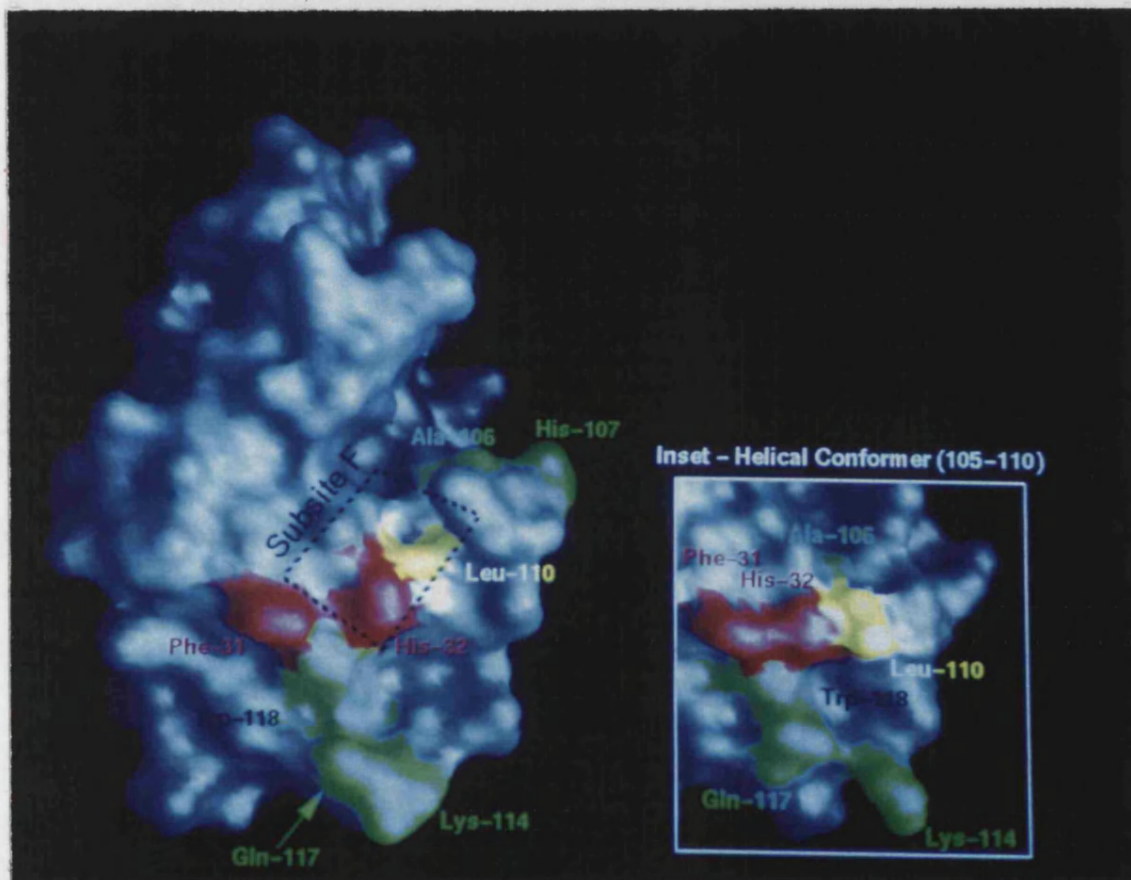


Figure 8.9 — Relationship between the functional residues in LA and the F subsite region.

The relationship between residues that have been implicated in the function of LA and the overall structure of LA are shown. The molecular surface of GPLA is shown in blue. Residues that appear to be important for LA's ability to promote glucose binding in the LS complex are shown in red and those important for affinity are shown in green. Leu-110 is coloured in yellow as changes at this position have differential effects on LA's function depending on the nature of the sidechain substituent. The approximate location of the analogous F subsite region of LA's cleft is indicated. **Inset:** The inset depicts the same region in mLA and shows the effect of the helical conformation of residues 105-110 on the structure of the F subsite. The residue colour scheme is the same as the main picture. Figure produced with *GRASP* (Nicholls *et al.*, 1991).

## 8.4 General Discussion

It has long been suspected that LA utilises residues in the cleft region to stabilise monosaccharide binding in the LS complex given its relationship with the C-type LYZs. The results from the site-directed mutagenesis of Phe-31, His-32 and Leu-110 demonstrate for the first time that LA and LYZ utilise a common subsite in their respective activities. Furthermore, amino acid changes at positions analogous to components of subsites C, D and E in LYZ have no specific effect on glucose binding in the LS complex (K. Brew, personal communication). Taken together, these results support the hypothesis that, in the LS complex, a region analogous to the F subsite participates in the binding of monosaccharides.

Monosaccharide binding sites are usually found in a cleft between either two domains or two proteins. Binding is mediated by hydrogen bonds between the sugar hydroxyls and polar groups in the protein as well as non-polar interactions between aromatic sidechains and the sugar rings (Quioco, 1986; Johnson *et al.*, 1988). Due to the lack of information about the nature of the monosaccharide binding site in GT, it is difficult to predict to what extent Phe-31 and His-32 participate in the binding of glucose in the LS complex. Phe-31 may make Van der Waals contacts with the glucose ring as substitutions that remove the aromatic ring (e.g. F31S) have major effects on LA's ability to promote glucose binding. In the case of F31Y-mLA, the introduction of an additional phenolic hydroxyl causes a smaller reduction in LA's ability to promote glucose binding in the LS complex (7-fold) while having a minimal effect (2-fold) on affinity. The reduced ability to promote glucose binding in this variant may be due to an unfavourable electrostatic effect between the phenolic hydroxyl and the bound

monosaccharide. The functional consequences of substitutions at His-32 demonstrate that this residue is also critical for monosaccharide binding, probably more so than Phe-31. Mutations at this position result in a drastic reduction of glucose binding in the LS complex combined with a lower affinity for GT. The ionisable sidechain of histidine has important sugar binding properties and it is involved in substrate binding in a number of glycosyl hydrolases (Qian *et al.*, 1994; Barrett *et al.*, 1995; Ducros *et al.*, 1995) and other oligosaccharide-binding proteins (Johnson *et al.*, 1988). Given the specific requirement of His-32 N<sup>δ</sup>1 atom for LA activity (Prieels *et al.*, 1979), it is tempting to suggest that this group makes direct contact with one of the polar hydroxyls of the galactosyl acceptor.

The results from these studies cannot completely rule out alternative mechanisms that might explain LA's modulatory effects. It is possible that LA binding to GT causes a conformational change that results in a complex that has higher affinity for glucose. However, this explanation seems unnecessary given that GT is capable of carrying out the galactosylation of glucose in the absence of LA (Khatra *et al.*, 1974). Spectroscopic measurements suggest that, although some sort of conformational change involving aromatic residues occurs on the addition of either UDP-galactose or UDP-galactose and GlcNAc to a GT·Mn<sup>2+</sup> complex, no significant changes occur when LA is added (Geren *et al.*, 1975; Takase and Ebner, 1981). Furthermore, large scale conformational changes on complexation are probably inconsistent with LA's rapid equilibrium binding to GT (Khatra *et al.*, 1974; Bell *et al.*, 1976; Powell and Brew, 1976b). Nevertheless, the flexibility exhibited by part of the GT interaction site indicates that more subtle

conformational changes may be important in the formation and function of the LS complex.

The interaction site for GT appears to involve a region on the surface of LA that is directly adjacent to the lower end of the cleft. Affinity for GT is selectively perturbed by substitutions for Ala-106 (Lys), His-107 (Ala, Tyr and Trp), Leu-110 (His), Lys-114 (Asn), Gln-117 (Ala) and Trp-118 (Tyr). All these residues are located in the flexible C-terminal portion of the molecule. The differential effects of amino acid substitutions at Leu-110, depending on the nature of the sidechain substituent, may result from local conformational differences in the 105-110 region. The unusual degree of conformational adaptability exhibited by LA in solution, depending on its interaction with metal ions (Kronman, 1989), and the observation that the A-state of LA retains the modulatory properties, despite its lack of definable tertiary structure, argues that LA's inherent ability to adjust its conformational state may be of paramount importance in the regulation mechanisms mediated by this unusual protein. Certainly, the various conformational forms of LA, especially those produced by physiologically relevant metal ions such as zinc, merit further investigation.

On a less speculative note, what are the future prospects of this work? Clearly, the mutagenesis studies have given us a greater insight into the functional regions of LA. The structural studies on the variant LAs should yield a plethora of information regarding the effects of particular substitutions. The current work has demonstrated that crystals, suitable for high resolution X-ray analysis, can be obtained for a number of the variants. Determination of these structures, in particular those involving the His-107 and Leu-110 variants, will undoubtedly facilitate a better understanding of the functional

effects of amino acid substitutions at these positions. At present, it is unclear whether these amino acid changes elicit any local conformational changes. However, it is worth emphasising that although determination of the variant LA crystal structures may provide a vast amount of useful data, a full understanding of the molecular mechanisms that underlie LA's biological function will require the elucidation of the three-dimensional structure of GT.

To the author's knowledge, there are no structures available for any of the glycosyltransferases despite their importance in a multitude of biological processes. Structural information about the nature of GT's active site and the LA interaction site is urgently needed before any significant advances can be made. Recent successes in cloning GT (Aoki *et al.*, 1990; Krezdorn *et al.*, 1993; Boeggeman *et al.*, 1993) and probing its functional residues by site directed mutagenesis (Zu *et al.*, 1995) are encouraging. Significant progress has also been made in producing large quantities of a soluble recombinant form of bovine GT (K. Brew, personal communication) and this should eventually facilitate crystallographic studies. Hopefully, these ongoing studies will provide invaluable information about the functional properties of GT. Our present knowledge of the enzyme lactose synthase has increased dramatically over the past 30 years. Nevertheless, many questions remain unanswered and a considerable amount of experimental work will be needed to fully understand the molecular basis of lactose biosynthesis.



# Appendix

## A1. DNA manipulation Protocols

### A1.1 Materials

The pET3a vector, T7 promoter and terminator primers and lysogenic *E. coli* expression host (BL21 DE3s) were purchased from Novagen (Madison, WI.). *Taq* polymerase, DNA ligase, *NdeI*, *BamHI* and assorted reaction buffers were from Pharmacia LKB. *BamHI* and *Sall* used in the restriction mapping were from New England Biolabs. Magic™ PCR Preps Purification System was purchased from Promega Corp. (Madison, WI.). Subcloning efficiency cells DH5α™ and NuSieve™ agarose gel were purchased from Gibco BRL (Gaithersburg, MD.). Other reagents were of analytical grade.

### A1.2 PCR Primers

The oligonucleotide primers were synthesised by Dr. Rudolf Werner, Department of Biochemistry and Molecular Biology, University of Miami. The sequences of the primers used for the construction of p-MLA and vector-encoded T7 primers are summarised in Table A1.1.

Table A1.1 – Primers used for construction of pMLA and PCR mutagenesis

Primer	Sequence	Orientation
NF-N	5'-CCATGCCACCCATATGGAACAGTTAAC-3'	Coding
NF-C	5'-GCAAAGACAGCGGATCCTCACAACCTTCTCAC-3'	Complementary
T7 Promoter	5'-TAATACGACTCACTATAGGG-3'	Coding
T7 Terminator	5'-GCTAGTTATTGCTCAGCGG-3'	Complementary

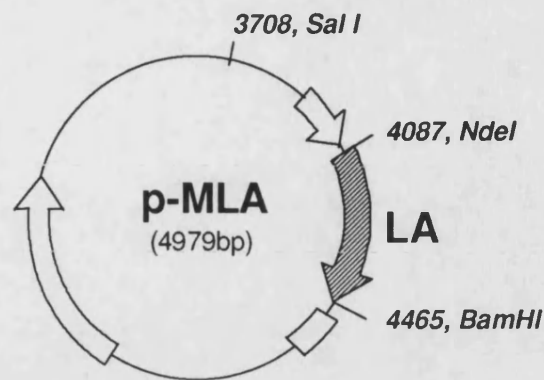
### A1.3 PCR Mutagenesis

The PCR reaction mixture consisted of 3µl 10X *Taq* reaction buffer (500mM KCl, 15mM MgCl<sub>2</sub>, 100mM Tris-HCl pH 9.0), 3µl 0.5mM dNTPs stock solution, 3µl 1/1000 dilution pMLA stock (4µg/µl), 0.5µl T7 promoter or terminator primer (30ng/µl), 0.25µl *Taq* DNA polymerase (5000 units/ml) and milliQ water to a final volume of 30µl. For each amplification, 1.5µl 1/100 dilution of the desired mutagenic primer was added to the reaction mixture. The PCR protocol involved an initial DNA melting step for 2min at 94°C followed by 25 cycles of a melting, primer annealing, elongation regime (1min at 94°C, 2min at 50°C, 1min at 74°C respectively) and a final cycle of elongation at 74°C for 10min. The product of the first reaction (megaprimer) was purified using Magic™ PCR Prep columns (as documented in column's Technical Bulletin). The purified megaprimer was eluted from the columns in 50µl milliQ, phenol-chloroform extracted, precipitated with ethanol and resuspended in 22µl milliQ. The second PCR amplification, using the cognate T7 primer, was carried as described above except that the megaprimer DNA replaced the milliQ in the reaction mixture (20µl). The full length PCR product was isolated by gel-purification on a 2% NuSieve™ agarose gel and purified directly from the excised band using the Magic™ PCR minicolumns (eluted in 45µl milliQ).

The isolated amplified DNA was double digested with *Bam*HI and *Nde*I at 37°C for 2.5 hours. The reaction mixture (200µl) consisted of 30µl amplified DNA, 20µl 10X One Phor All Buffer *PLUS*™ (100mM MgAc, 500mM KAc, 100mM Tris-acetate pH 7.5), 150µl milliQ, 1.5µl *Bam*HI (20000 units/ml) and 1.5µl *Nde*I (10000 units/ml). The digested PCR product was purified using the PCR minicolumns and ligated into *Nde*I/*Bam*HI digested pET3a. The ligation reaction (40µl) consisted of 20µl cut amplified DNA, 4µl 10X One Phor All Buffer *PLUS*™, 4µl 10mM ATP, 2µl cut phosphatased pET3a, 10µl milliQ and 5µl T4 ligase (8000 units/ml). The ligated vector was transformed into competent subcloning efficiency DH5α cells according to the shipped protocol.

Figure A1.1 – Plasmid map of p-MLA.

The relevant restriction sites and the expected fragment sizes are shown



Restriction Enzymes	Expected Fragment Size (b.p.)
<i>NdeI</i> / <i>BamHI</i>	378
<i>SalI</i> / <i>BamHI</i>	757

## A1.4 Restriction Mapping

Ampicillin-resistant transformants were characterised for correctly sized inserts using double digestions with *SalI/BamHI* and *NdeI/BamHI*. Digestions consisted of 10µl of mini-prepped DNA from each transformant and 10µl reaction mix (2µl 10X New England Biolabs buffer III, 0.2µl 100X bovine serum albumin, 10 units of *BamHI* and *SalI*). Digests were incubated at 37°C for 90min and then run on a 2% agarose gel. A fragment size of around 750 base pairs for the *SalI/BamHI* digestion indicated that the plasmids contained the correct insert (Figure A1.1).

## A2. Purification Methods

### A2.1 Isolation of Inclusion Bodies

Typically, the cells from a 6 litre expression run were harvested by centrifugation (30min @ 4100rpm). The cell pellet was resuspended in 40mls of lysis buffer (50mM Tris-HCl pH 8.0, 1mM EDTA, 100mM NaCl). The inclusion bodies were isolated by the lysozyme-lysis method (Sambrook *et al.*, 1989). The bacterial cell wall was digested by

addition of 900µl 10mg/ml lysozyme solution (Sigma) and incubation at RT for 20min with occasional stirring. The cell membranes were solubilised by addition of 250µl 200mg/ml deoxycholic acid (Sigma) and incubation at 37°C. The DNA was then solubilised by incubation at RT for 45min with 100µl 10mg/ml *DNAseI* (Boehringer). The inclusion bodies were harvested by centrifugation (15min @ 11400rpm), washed and then re-centrifuged. The inclusion body pellet was either stored at -70°C or solubilised in urea for the anion exchange separation.

## A2.2 HPLC Separations

Chromatography was carried out with a Hewlett-Packard 1090 liquid chromatograph fitted with a Hydrophase™ HP-PEI anion exchange column (Interaction Chromatography Inc., San Jose, CA.). Buffers used for the separation were : buffer A, 20mM Tris-HCl pH 7.8 containing 100µM CaCl<sub>2</sub> and buffer B, 20mM Tris-HCl pH 7.8 containing 100µM CaCl<sub>2</sub> and 1M NaCl. Typically, the column was equilibrated with 85% buffer A, 15% buffer B and elution was performed with a linear gradient from 15 to 60% buffer B over 1hr. The separation was monitored by absorbance at 280nm. The column was run at a flow rate of 1.0ml/min throughout. Native forms of bovine milk LA, MLA and mutants eluted at about 30min (0.375M NaCl).

## A2.3 Determination of Protein Concentration

Protein concentrations were estimated from absorbance at 280nm. A value of 20.1 was assumed for the  $E_{280}^{1\%}$  for mL A and corresponding values for the variant LAs were calculated from their tryptophan, tyrosine and cysteine content (Gill and von Hippel, 1989).

## A3. Kinetic Assay Protocols

### A3.1 Lactose Synthase Assays

The activity of MLA and the mutants in promoting glucose binding to bovine GT was determined by a radiochemical assay as described in previous studies (Brew *et al.*, 1968; Khatra *et al.*, 1974). All incubations were performed in 100µl reaction mixture

containing 10mM Tris-HCl pH 7.4, 10mM  $\alpha$ -D-glucose, 10mM  $MnCl_2$ , 0.1% bovine serum albumin, 10 $\mu$ l bovine GT stock solution (see below) and varying concentrations of mLA (0-10 $\mu$ M). Assay mixtures were prepared on ice in duplicate and the reaction was started by the addition of 10 $\mu$ l radiolabelled UDP-galactose. After incubation at 37°C for 10min, the reaction was stopped by the addition of 100 $\mu$ l 0.1M EDTA. Appropriate controls were included to correct for the non-specific hydrolysis of UDP-galactose. Neutral sugars were then separated from the unreacted UDP-galactose by anion exchange. The reaction mixture was transferred to columns (0.5 x 4cm) of AG1-X8 (chloride form; 200-400 mesh) strong anion exchange resin (BioRad) prepared in short glass Pasteur pipettes. The columns were pretreated with 0.1M lactose to minimise the non-specific binding of the radiolabelled product. The reaction tubes were washed twice with 0.5 and 1ml aliquots of milliQ water with both washings transferred to the columns. The column eluates were collected in vials, mixed with 5mls of EcoLume scintillation fluid and the radioactivity counted for 2min in a Packard Tri-Carb LKB scintillation counter. The radiolabelled UDP-galactose stock solution was prepared by adding tritiated UDP-galactose (UDP-D-[ $^3H$ ]galactose, purchased from Amersham) to a 3mM solution of unlabelled UDP-galactose (Sigma) so that the specific activity was approximately 1500cpm/ $\mu$ l. The GT stock solution used in the assays was prepared by dissolving lyophilised bovine milk GT (Fluka) in 0.1M Tris-HCl pH 7.4 to a concentration at which the amount used in the assay yielded a maximal activity of 15000cpm. The GT stock solution was stored in 200 $\mu$ l aliquots at -70°C.

### A3.2 Chitobiose Inhibition Assays

The ability of mLA and the mutants to inhibit galactosyl transfer to oligosaccharide substrates was measured in the same manner except that the reaction mixture contained 0.5mM N, N' diacetylchitobiose (Sigma) in place of glucose. A more concentrated range of mLA was also required (0-60 $\mu$ M). Due to the low activities of some of the mutants, much greater concentrations of mutant mLAs were required to accurately determine the kinetic parameters. The  $K_m$  for ChB was determined in a similar, but separate, reaction in the absence of LA. The concentration of ChB used to determine

the  $K_m^{ChB}$  ranged from 0 to 1.5mM. The kinetic parameters were determined by fitting the kinetic data to the appropriate rate equations using the *ENZFITTER* program (Biosoft).

#### A4. Crystallisation Methods

All reagents were of analytical grade. All chemicals were obtained from Sigma unless stated otherwise. A variety of different screening methods were employed:

a) **Crystal Screen I and II:** A commercially available reagent kit, designed to provide a highly effective and rapid screening method for the crystallisation of macromolecules (Hampton Research, Riverside, CA, USA). The screens are based on the sparse matrix method of Jancarik and Kim (1991). The screens were carried out at 16°C, room temperature (16-20°C) and 37°C. The various screen conditions are given in Tables A1.2 and A1.3.

b) **Grid screens:**

i) **Ammonium sulphate grid:** The matrix was designed to cover a pH range of 4.6 to 9.0 in 0.5 pH unit steps and an AS concentration range of 1.4 to 2.8M. 10mM calcium chloride was also included in the 8 x 10 grid to saturate the calcium binding site of LA. Separate trials were carried out at 16 and 37°C.

ii) **Polyethylene glycol 4000 (PEG 4K) grid:** The matrix was designed with PEG 4K as the sole precipitating agent over a range of concentrations from 3 to 24% (w/v) and a pH range of 4.6 to 9.0 in 0.5 pH unit steps. The resulting 8 x 12 grid, again including 10mM CaCl<sub>2</sub>, was screened at 16°C and 35°C.

Crystallisation trials were carried out using the vapour diffusion method with either hanging or sitting drops (McPherson, 1982). This method relies on the transport of either water or some volatile agent between a drop of mother liquor, usually 10µl in volume, and a much larger reservoir solution of 0.75–1ml volume. Due to the large volume of the reservoir, relative to the drop, the final equilibration conditions are essentially those of the initial reservoir state.

Table A1.2 — Hampton Research *Crystal Screen I* conditions.

Tube #	Salt	Buffer	Precipitant
1	0.02M Ca Chloride	0.1M Na Acetate pH4.6	30% (v/v) MPD
2	None	None	0.4M K, Na Tartrate
3	None	None	0.4M NH4 Phosphate
4	None	0.1M Tris HCl pH 8.5	2.0M NH4 Sulphate
5	0.2M Na Citrate	0.1M Na HEPES pH 7.5	30% (v/v) MPD
6	0.2M Mg Chloride	0.1M Tris HCl pH 8.5	30% (w/v) PEG 4000
7	None	0.1M Na Cacodylate pH 6.5	1.4M Na Acetate
8	0.2M Na Citrate	0.1M Na Cacodylate pH 6.5	30% (v/v) 2-propanol
9	0.2M NH4 Acetate	0.1M Na Citrate pH 5.6	30% (w/v) PEG 4000
10	0.2M NH4 Acetate	0.1M Na Acetate pH4.6	30% (w/v) PEG 4000
11	None	0.1M Na Citrate pH 5.6	1.0M NH4 Phosphate
12	0.2M Mg Chloride	0.1M Na HEPES pH 7.5	30% (v/v) 2-propanol
13	0.2M Na Citrate	0.1M Tris HCl pH 8.5	30% (v/v) PEG 400
14	0.2M Ca Chloride	0.1M Na HEPES pH 7.5	28% (v/v) PEG 400
15	0.2M NH4 Sulphate	0.1M Na Cacodylate pH 6.5	30% (w/v) PEG 8000
16	None	0.1M Na HEPES pH 7.5	1.5M Li Sulphate
17	0.2M Li Sulphate	0.1M Tris HCl pH 8.5	30% (w/v) PEG 4000
18	0.2M Mg Acetate	0.1M Na Cacodylate pH 6.5	20% (w/v) PEG 8000
19	0.2M NH4 Acetate	0.1M Tris HCl pH 8.5	30% (v/v) 2-propanol
20	0.2M NH4 Sulphate	0.1M Na Acetate pH4.6	25% (w/v) PEG 4000
21	0.2M Mg Acetate	0.1M Na Cacodylate pH 6.5	30% (v/v) MPD
22	0.2M Na Acetate	0.1M Tris HCl pH 8.5	30% (w/v) PEG 4000
23	0.2M Mg Chloride	0.1M Na HEPES pH 7.5	30% (v/v) PEG 400
24	0.2M Ca Chloride	0.1M Na Acetate pH4.6	20% (v/v) 2-propanol
25	None	0.1M Imidazole pH 6.5	1.0M Na Acetate
26	0.2M NH4 Acetate	0.1M Na Citrate pH 5.6	30% (v/v) MPD
27	0.2M Na Citrate	0.1M Na HEPES pH 7.5	20% (v/v) 2-propanol
28	0.2M Na Acetate	0.1M Na Cacodylate pH 6.5	30% (w/v) PEG 8000
29	None	0.1M Na HEPES pH 7.5	0.8M Na, K Tartrate
30	0.2M NH4 Sulphate	None	30% (w/v) PEG 8000
31	0.2M NH4 Sulphate	None	30% (w/v) PEG 4000
32	None	None	2.0M NH4 Sulphate
33	None	None	4.0M Na Formate
34	None	0.1M Na Acetate pH4.6	2.0M Na Formate
35	None	0.1M Na HEPES pH 7.5	1.6M Na, K Phosphate
36	None	0.1M Tris HCl pH 8.5	8% (w/v) PEG 8000
37	None	0.1M Na Acetate pH4.6	8% (w/v) PEG 4000
38	None	0.1M Na HEPES pH 7.5	1.4M Na Citrate
39	None	0.1M Na HEPES pH 7.5	2% (v/v) PEG 400 & 2.0M NH4 Sulphate
40	None	0.1M Na Citrate pH 5.6	20% (v/v) 2-propanol & 20% (w/v) PEG 4000
41	None	0.1M Na HEPES pH 7.5	10% (v/v) 2-propanol & 20% (w/v) PEG 4000
42	0.05M K Phosphate	None	20% (w/v) PEG 8000
43	None	None	30% (v/v) PEG 1500
44	None	None	0.2M Mg Formate
45	0.2M Zn Acetate	0.1M Na Cacodylate pH 6.5	18% (w/v) PEG 8000
46	0.2M Ca Acetate	0.1M Na Cacodylate pH 6.5	18% (w/v) PEG 8000
47	None	0.1M Na Acetate pH4.6	2.0M NH4 Sulphate
48	None	0.1M Tris HCl pH 8.5	2.0M NH4 Phosphate
49	1.0M Li Sulphate	None	2% (w/v) PEG 8000
50	0.5M Li Sulphate	None	15% (w/v) PEG 8000

Table A1.2 — Hampton Research *Crystal Screen II* conditions.

Tube #	Salt	Buffer	Precipitant
1	2.0M Na Chloride	None	10% (w/v) PEG 6000
2	0.01M Cetyl trimethyl ammonium bromide	None	0.5M Na Chloride
3	None	None	25% Ethylene Glycol
4	None	None	35% (v/v) Dioxane
5	2.0M NH4 Sulphate	None	5% (v/v) 2-propanol
6	None	None	1.0M Imidazole pH 7.0
7	None	None	10% (w/v) PEG 1000 10% (w/v) PEG 8000 10% (v/v) Ethanol
8	1.5M Na Chloride	None	2.0M Na Chloride
9	None	0.1M Na Acetate pH4.6	30% (v/v) MPD
10	0.2M Na Chloride	0.1M Na Acetate pH4.6	1.0M 1,6-Hexanediol
11	0.01M Co Chloride	0.1M Na Acetate pH4.6	30% (v/v) PEG 400
12	0.1M Cd Chloride	0.1M Na Acetate pH4.6	30% (v/v) PEG MME 2000
13	0.2M NH4 Sulphate	0.1M Na Acetate pH4.6	2.0M NH4 Sulphate
14	0.2M K, Na Tartrate	0.1M Na Citrate pH 5.6	1.0M Li Sulphate
15	0.5M NH4 Sulphate	0.1M Na Citrate pH 5.6	4% (v/v) Polyethyleneimine
16	0.5M Na Chloride	0.1M Na Citrate pH 5.6	35% (v/v) tert-butanol
17	None	0.1M Na Citrate pH 5.6	10% (w/v) Jeffamine M-600
18	0.01M Fe Chloride	0.1M Na Citrate pH 5.6	2.5M 1,6-Hexanediol
19	0.01M Mn Chloride	0.1M Na Citrate pH 5.6	1.6M Mg Sulphate
20	None	0.1M MES pH 6.5	2.0M Na Chloride
21	0.2M Na/K Phosphate	0.1M MES pH 6.5	12% (w/v) PEG 20000
22	None	0.1M MES pH 6.5	10% (v/v) Dioxane
23	1.6M NH4 Sulphate	0.1M MES pH 6.5	30% (w/v) Jeffamine M-600
24	0.05M Ce Chloride	0.1M MES pH 6.5	1.8M NH4 Sulphate
25	0.01M Co Chloride	0.1M MES pH 6.5	30% (w/v) PEG MME 5000
26	0.2M NH4 Sulphate	0.1M MES pH 6.5	25% (w/v) PEG MME 500
27	0.01M Zn Sulphate	0.1M MES pH 6.5	1.6M Na Citrate pH 6.5
28	None	None	30% (v/v) MPD
29	0.5M NH4 Sulphate	0.1M Na HEPES pH 7.5	10% (w/v) PEG 6000
30	None	0.1M Na HEPES pH 7.5	5% (v/v) MPD 20% (w/v) Jeffamine M-600
31	None	0.1M Na HEPES pH 7.5	1.6M NH4 Sulphate
32	0.1M Na Chloride	0.1M Na HEPES pH 7.5	2.0M NH4 Formate
33	None	0.1M Na HEPES pH 7.5	1.0M Na Acetate
34	0.05M Cd Sulphate	0.1M Na HEPES pH 7.5	70% (v/v) MPD
35	None	0.1M Na HEPES pH 7.5	4.3M Na Chloride
36	None	0.1M Na HEPES pH 7.5	10% (w/v) PEG 8000 8% (v/v) Ethylene Glycol 20% (w/v) PEG 10000
37	None	0.1M Na HEPES pH 7.5	3.4M 1,6-Hexanediol
38	None	0.1M Na HEPES pH 7.5	25% (v/v) tert-butanol
39	0.2M Mg Chloride	0.1M Tris HCl pH 8.5	1.0M Li Sulphate
40	0.1M Ca Chloride	0.1M Tris HCl pH 8.5	12% (v/v) Glycerol
41	0.01M Ni Chloride	0.1M Tris HCl pH 8.5	50% (v/v) MPD
42	1.5M NH4 Sulphate	0.1M Tris HCl pH 8.5	20% (w/v) Ethanol
43	0.2M NH4 Phosphate	0.1M Tris HCl pH 8.5	20% (w/v) PEG MME 2000
44	None	0.1M Tris HCl pH 8.5	30% (w/v) PEG MME 550
45	0.01M Ni Chloride	0.1M Tris HCl pH 8.5	2.0M Mg Chloride
46	0.1M Na Chloride	0.1M Bicine pH 9.0	10% (w/v) PEG 20000
47	None	0.1M Bicine pH 9.0	
48	2% (v/v) Dioxane	0.1M Bicine pH 9.0	



In the hanging drop procedure, a drop of mother liquor (3–5 $\mu$ l protein stock + 3–5 $\mu$ l precipitating solution) is suspended from the underside of a siliconized microscope coverslip and placed over a reservoir containing 0.75ml of precipitating solution. Disposable plastic tissue culture plates (Linbro model, Flow Laboratories) were used for all the trials. The chamber formed by the reservoir and inverted coverslip was sealed with silicone grease. For the ‘sitting drop’ procedure, 5 $\mu$ l of protein stock and 5 $\mu$ l of precipitating solution were mixed in plastic microbridges (Crystal Microsystems, Oxford, UK), a small device that allows the drop to be suspended over the reservoir solution. The well was then sealed with an ordinary coverslip and some silicone grease. This method can accommodate larger drop volumes than the hanging drop procedure and also minimises drop flattening and spreading. It was found to be particularly beneficial in the case of the monoclinic III mLA crystals.

## A5. Frequently Used Computer Software

### A5.1 Data Collection and Processing

<i>XDS</i> and <i>XSCALE</i>	Wolfgang Kabsch’s processing and reduction programs for the area detector diffraction data.
<i>DENZO</i> and <i>SCALEPACK</i>	Zbyszek Otwinowski’s processing and reduction/scaling programs for image plate diffraction data.
<i>X2L</i>	Garry Taylor’s in-house program to convert <i>XSCALE</i> output to LCF format.
<i>F2MTZ</i> , <i>LCF2MTZ</i> , <i>MTZ2VARIOUS</i>	Various CCP4 programs allowing the interchange of the format of reflection data files.
<i>TRUNCATE</i>	CCP4 program to apply Wilson statistics to diffraction data and conversion of reflection intensities to structure factor amplitudes.

### A5.2 Molecular Replacement

<i>AMoRe</i>	Jorge Navaza’s MR program (maintained by CCP4).
<i>POLARRFN</i>	Wolfgang Kabsch’s polar rotation function program used for calculating self-rotation function (CCP4).
<i>PDBUTILS</i>	In-house utility for manipulating PDB files. Primarily used here for applying rotation and translations to the MR search model.

<b>X-PLOR</b>	No need for an explanation here. Also used for refinement.
<b>TFFC</b>	Ian Tickle's full symmetry translation function (CCP4).

### A5.3 Model Building and Analysis

<b>FRODO</b>	Alwyn Jones' original program modified for use on an Evans and Sutherland PS390.
<b>O</b>	State of the art rebuilding program. Includes a whole number of utility programs (see below)
<b>PROCHECK</b>	Roman Laskowski's model validation program.
<b>DSSP</b>	Kabsch and Sanders secondary structure analysis program.
<b>CONTACT</b>	Calculates intra- and intermolecular contacts from a PDB input (CCP4).
<b>ASH</b>	David Stuart's (Oxford) program for the superposition of protein structures using specific residue equivalencies.
<b>SHP</b>	David Stuart's (Oxford) program for the dynamic superposition of protein structures. This does not require any prior knowledge of equivalences between the two structures.

### A5.4 Uppsala utilities

<b>4D_MAPMAN</b>	General program for manipulation of electron density maps from contouring to changing the map sectioning. Also used for creating connectivity files for <b>O</b> .
<b>4D_MOLEMAN</b>	Utility for analysing model <i>B</i> -factors etc.
<b>4D_OPLOT</b>	Rendering <b>O</b> plot files for postscript printer.
<b>4D_VOIDOO</b>	Program to calculate cavities in protein structures.
<b>4D_RMSPDB</b>	Performing RMS deviation calculations.

### A5.5 Heavy Atom Processing

<b>PHASES</b>	Complete package for interpreting heavy atom data.
<b>SCALEIT</b>	Scaling native and heavy atom data (CCP4).
<b>LOCAL</b>	Scaling native and heavy atom data (CCP4).

## A5.6 Map Calculation

<b>PDBSET</b>	CCP4 utility for manipulating PDB files.
<b>SFALL</b>	Structure factor calculation (CCP4).
<b>RSTATS</b>	Scaling of $F_{\text{obs}}$ and $F_{\text{calc}}$ (CCP4).
<b>SIGMAA</b>	Randy Read's phase combination program used to calculate minimum bias coefficients for difference Fourier syntheses (CCP4).
<b>FFT</b>	Fast Fourier transform algorithm (CCP4).

## A5.7 Graphical Display

<b>GRASP</b>	Graphical program to display a number of different types of molecular surfaces.
<b>MOLSCRIPT</b>	Per Kraulis' program for creating schematic plots of proteins.
<b>BOBSCRIPT</b>	Robert Esnouf's (Oxford) modified version of <b>MOLSCRIPT</b> .
<b>RASTER3D</b>	Ethan Merritt's program for photorealistic surface rendering. Used in conjunction with <b>MOLSCRIPT</b> / <b>BOBSCRIPT</b> .
<b>SHOWCASE</b>	SGI's general drawing program.
<b>ALSCRIPT</b>	Geoff Barton's program for producing figures from multiple sequence alignments.

This thesis was produced using Microsoft Word 6 for Windows™. All the graphs were drawn using SigmaPlot for Windows™ and some of the schematic figures were created using Microsoft PowerPoint 4.0.

## References

- Acharya, K.R., Stuart, D.I., Walker, N.P.C., Lewis, M. and Phillips, D.C. (1989) Refined structure of baboon  $\alpha$ -lactalbumin at 1.7Å resolution. *J. Mol. Biol.* **208**, 99-127.
- Acharya, K.R., Stuart, D.I., Phillips, D.C. and Scheraga, H.A. (1990) A critical evaluation of the predicted and X-ray structures of  $\alpha$ -lactalbumin. *J. Prot. Chem.* **9**, 549-563.
- Acharya, K.R., Ren, J., Stuart, D.I., Phillips, D.C. and Fenna, R.E. (1991) Crystal structure of human  $\alpha$ -lactalbumin at 1.7Å 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 three-dimensional structures of echidna, horse and pigeon lysozymes: Calcium binding lysozymes and their relationship with  $\alpha$ -lactalbumins. *J. Prot. Chem.* **13**, 569-584.
- Alexandrescu, A.T., Broadhurst, R.W., Wormald, C., Chyan, C-L., Baum, J. and Dobson, C.M. (1992) <sup>1</sup>H-NMR assignments and local environments of aromatic residues in bovine, human and guinea-pig variants of  $\alpha$ -lactalbumin. *Eur. J. Biochem.* **210**, 699-702.
- 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 two-dimensional NMR study. *Biochemistry* **32**, 1707-1718.
- Aoki, D., Appert, H.E., Johnson, D., Wong, S.S. and Fukuda, M.N. (1990) Analysis of the substrate binding sites of human galactosyltransferase by protein engineering. *EMBO J.* **9**, 3171-3178.
- Aoki, D, Lee, N., Yamaguchi, N., Dubois, C. and Fukuda, M.N. (1992) Golgi retention of a trans-golgi membrane-protein, galactosyltransferase, requires cysteine and histidine residues within the membrane-anchoring domain. *Proc. Natl. Acad. Sci. USA.* **89**, 4319-4323.
- Aschaffenburg, R., Fenna, R.E., Handford, B.O. and Phillips, D.C. (1972a) Crystallography of  $\alpha$ -lactalbumin. I. Low salt crystals of goat  $\alpha$ -lactalbumin. *J. Mol. Biol.* **67**, 525-528.
- Aschaffenburg, R., Fenna, R.E. and Phillips, D.C. (1972b) Crystallography of  $\alpha$ -lactalbumin. II. High salt crystals of goat  $\alpha$ -lactalbumin. *J. Mol. Biol.* **67**, 529-531.
- Aschaffenburg, R., Fenna, R.E., Phillips, D.C., Smith, S.G., Buss, D.H., Jenness, R. and Thompson, M.P. (1979) Crystallography of  $\alpha$ -lactalbumin. III. Crystals of baboon milk  $\alpha$ -lactalbumin. *J. Mol. Biol.* **127**, 135-137.

- Babad, H. and Hassid W.Z. (1964) A soluble lactose-synthesising enzyme from bovine milk. *J. Biol. Chem.* **239**, PC-946-948.
- Babad, H. and Hassid W.Z. (1966) Soluble uridine diphosphate: D-glucose  $\beta$ -4-O-galactosyl transferase from bovine milk. *J. Biol. Chem.* **241**, 2672-2678.
- Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J. and Hughes, M.A. (1995) The crystal structure of a cyanogenic  $\beta$ -glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* **3**, 951-960.
- Barton, G.J. (1993) *ALSCRIPT*— A tool to format multiple sequence alignments. *Prot. Eng.* **6**, 37-40.
- Baum, J., Dobson, C.M., Evans, P.A. and Hanley, C. (1989) Characterisation of a partly folded protein by NMR methods: Studies on the molten globule state of guinea-pig  $\alpha$ -lactalbumin. *Biochemistry* **28**, 7-13.
- Bauby, J.J. and Fauquant, J. (1989) Evidence for the presence of glycosylated alpha-lactalbumin in an industrial bovine alpha-lactalbumin preparation. *Lait* **69**, 315-322.
- Bell, J.E., Castellino, F.J., Trayer, I.P. and Hill, R.L. (1975) Modification of  $\alpha$ -lactalbumin with N-bromosuccinimide and 2-hydroxy-5-nitrobenzylbromide. *J. Biol. Chem.* **250**, 7579-7585.
- Bell, J.E., Beyer, T.A. and Hill, R.L. (1976) The kinetic mechanism of bovine milk galactosyltransferase. *J. Biol. Chem.* **251**, 3003-3013.
- Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R.N., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M., Reichlin, M., Sercarz, E.E., Smith-Gill, S.J., Todd, P.E. and Wilson, A.C. (1984) The antigenic structure of proteins. *Ann. Rev. Immunol.* **2**, 67-101.
- Bentley, G.A. (1992) Some applications of the phased translation function using calculated phases. In "Molecular Replacement" Proceedings of the CCP4 Study Weekend, 31 January-1 February, 1992, (eds. Dodson, E.J., Gover, S., Wolf, W.) pp. 20-32. SERC Daresbury Laboratory, Warrington, England.
- Berliner, L. and Koga, K. (1987)  $\alpha$ -Lactalbumin binding to membranes: evidence for a partially buried protein. *Biochemistry* **26**, 3006-3009.
- Blake, C.C.F., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1967a) On the conformation of the hen egg-white lysozyme molecule. *Proc. Roy. Soc. B.* **167**, 365-377.
- Blake, C.C.F., Johnson, L.N., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1967b) Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. Roy. Soc. B.* **167**, 378-388.
- Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press, London, UK.
- Boeggeman, E.E., Balaji, P.V., Sethi, N., Masibay, A.S. and Qasba, P.K. (1993) Expression of deletion constructs of bovine  $\beta$ -1,4-galactosyltransferase in *Escherichia coli*: importance of Cys134 for its activity. *Prot. Engineering.* **6**, 779-785.

- Bradbury, J.H. and Norton, R.S. (1975) Proton-magnetic-resonance spectroscopic study of histidine residues of bovine  $\alpha$ -lactalbumin. *Eur. J. Biochem.* **53**, 387-396.
- Brändén, C.I. and Jones, T.A. (1990) Between objectivity and subjectivity. *Nature* **343**, 687-689.
- Brew, K., Vanaman, T.C. and Hill, R.L. (1967) Comparison of the amino acid sequences of bovine alpha-lactalbumin and hen's egg-white lysozyme. *J. Biol. Chem.* **242**, 3747-3749.
- Brew, K., Vanaman, T.C. and Hill, R.L. (1968) The role of  $\alpha$ -lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction. *Proc. Natl. Acad. Sci. USA* **59**, 491-497.
- Brew, K. (1969) Secretion of  $\alpha$ -lactalbumin into milk and its relevance to the organisation and control of lactose synthetase. *Nature* **223**, 671-672.
- Brew, K. (1972) The complete amino-acid sequence of guinea-pig alpha-lactalbumin. *Eur. J. Biochem.* **27**, 341-353.
- Brew, K. and Hill, R.L. (1975) Lactose biosynthesis. *Rev. Physiol. Biochem. Pharmacol.* **72**, 105-158.
- 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.
- Brew, K., Richardson, R.H. and Sinha, S.K. (1979) Structural basis of the regulation of galactosyltransferase. In *From Gene to Protein: Information Transfer in Normal and Abnormal Cells* eds. T.R. Russell, K. Brew, H. Faber & J. Schultz. **16** pp. 433-447, Miami Winter Symposia, Academic Press, New York.
- Brew, K. and Grobler, J.A. (1992)  $\alpha$ -Lactalbumin in "Advances in Dairy Chemistry – I", (ed. Fox, P.), pp. 191-229, Elsevier, London.
- Brew, K., Grobler, J.A., Wang, M., Scott, W.A., Yadav, S.P. and Malinovskii, V. (1993) Molecular basis of specificity modulation in galactosyltransferase by  $\alpha$ -lactalbumin. In "Advances in Gene Technology: Protein Engineering and Beyond" (eds. K. Brew, G. Petsko, F. Ahmad, H. Bialy, S. Black, A. Fernandez, R. Fenna, E. Lee and W. Whelan). **3**, pp. 98, Miami Winter Symposia, IRL Press, Oxford.
- Brodbeck, U. and Ebner, K.E. (1966) Resolution of a soluble lactose synthetase into two protein components and solubilisation of microsomal lactose synthetase. *J. Biol. Chem.* **241**, 762-764.
- Brodbeck, U., Denton, W.L., Tanahashi, N. and Ebner, K.E. (1967) The isolation and identification of the B protein of lactose synthetase as  $\alpha$ -lactalbumin. *J. Biol. Chem.* **242**, 1391-1397.
- Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. and Hill, R.L. (1969) A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.* **42**, 65-86.

- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic *R*-factor refinement by molecular dynamics. *Science* **235**, 458-460.
- Brünger, A.T. (1989) A memory-efficient fast fourier transformation algorithm for crystallographic refinement on supercomputer. *Acta Cryst.* **A45**, 42-50.
- Brünger, A.T. (1990a) X-PLOR manual - Version 2.1 "A System for Crystallography and NMR" New Haven, Yale University Press.
- Brünger, A.T. (1990b) Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Cryst.* **A46**, 46-57.
- Brünger, A.T., Krukowski, A. and Erickson, J. (1990c) Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.* **A46**, 585-593.
- Brünger, A.T. (1992a) X-PLOR manual - Version 3.1 "A System for Crystallography and NMR" New Haven, Yale University Press.
- Brünger, A.T. (1992b) Molecular replacement with *XPLOR*: PC-refinement and free *R*-value. In "Molecular Replacement" Proceedings of the CCP4 Study Weekend, 31 January-1 February, 1992, pp.49-61. SERC Daresbury Laboratory, Warrington, England.
- Brünger, A.T. (1992c) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.
- Brünger, A.T. (1993) Assessment of phase accuracy by cross validation: the free *R* value. Methods and applications. *Acta Cryst.* **D49**, 24-36.
- Brünger, A.T. and Nilges (1993) Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Quart. Rev. Biophys.* **26**, 49-125.
- Carter, C.W. Jr. and Carter, C.W. (1979) Protein crystallisation using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12219-12223.
- Carter, C.W. Jr. (1992) Design of crystallisation experiments and protocols. In "Crystallization of Proteins and Nucleic acids: A Practical Approach." (eds. Ducruix, A. & Giegé, R.) pp. 47-71, IRL Press, Oxford University Press, U.K.
- Castellano, E., Oliva, G. and Navaza, J. (1992) Fast rigid-body refinement for molecular replacement techniques. *J. Appl. Cryst.* **25**, 281-284.
- Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
- Creighton, T.E. (1980) Role of the environment in the refolding of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* **144**, 521-550.
- Crick, F.H.C. and Magdoff, B. (1956) *Acta Cryst.* **9**, 901-908.
- Crowther, R. (1972) The fast rotation function. In "The Molecular Replacement Method" (ed. Rossmann, M.G.), pp. 173-178, Gordon & Breach, New York.
- Crowther, R. and Blow, D. (1967) A method for positioning a known molecule in an unknown crystal structure. *Acta Cryst.* **23**, 544-548.

- Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994) Screening and optimisation strategies for macromolecular crystal growth. *Acta Cryst.* **D50**, 414-423.
- Dauter, Z., Sieker, L.C. and Wilson, K.S. (1992) Refinement of rubredoxin from *Desulfovibrio vulgaris* at 1.0Å with and without restraints. *Acta Cryst.* **B48**, 42-59.
- Dautigny, A., Prager, E.M., Pham-Dinh, D., Jollès, J., Pakdel, F., Grinde, B., Jollès, P. (1991) cDNA and amino acid sequences of rainbow trout (*Oncorhynchus mykiss*) lysozymes and their implications for the evolution of lysozyme and  $\alpha$ -lactalbumin. *J. Mol. Evol.* **32**, 187-198.
- Derewenda, U., Dodson, E.J., Dodson, G.G., Hodgkin, D.C. and Swift, H.J. (1990) Uncertainty and bias in difference fourier maps. In "Accuracy and Reliability of Macromolecular Crystal Structures". Proceedings of the CCP4 Study Weekend, 26-27 January, 1990. (eds. Henrick, K., Moss, D.S., Tickle, I.J.) pp. 103-113, SERC Daresbury Laboratory, Warrington, England.
- Devereux, J. Haerberli, P. and Smithies, O. (1984) A comprehensive set of sequence-analysis programs for the VAX. *Nucleic Acid Res.* **12**, 387-395.
- Dodd, S.C., Forsyth, I.A., Buttle, H.L., Gurr, M.I. and Dils, R.R. (1994) Hormonal induction of alpha-lactalbumin and beta-lactoglobulin in cultured mammary explants from pregnant pigs. *J. Dairy Sci.* **61**, 35-45.
- Dodson, E.J. (1992) From molecular replacement solution to refined structure. In "Molecular Replacement" Proceedings of the CCP4 Study Weekend, 31 January-1 February, 1992. (eds. Dodson, E.J., Gover, S. and Wolf, W.) pp. 84-86, SERC Daresbury Laboratory, Warrington, England.
- Dodson, E.J., Gover, S. and Wolf, W. (1992) Editors. "Molecular Replacement" Proceedings of the CCP4 Study Weekend, 31 January-1 February, 1992. SERC Daresbury Laboratory, Warrington, England.
- Drenth J. (1991) The Chemistry of Heavy Atom Attachment. In "Isomorphous Replacement and Anomalous Scattering" Proceedings of the Daresbury Study Weekend, 25-26 January, 1991. (eds. Wolf, W., Evan, P.R., Leslie, A.G.W.) pp. 1-8, SERC Daresbury Laboratory, Warrington, England.
- Drenth, J. (1994) *Principles of Protein X-ray Crystallography*. Springer-Verlag, Berlin, Germany.
- Driessen, H.P.C., Bax, B., Slingsby, C., Lindley, P.F., Mahadevan, D., Moss, D.S. and Tickle, I.J. (1991) Structure of oligomeric  $\beta$ -B2-crystallin - an application of the T2 translation function to an asymmetric unit containing 2 dimers. *Acta Cryst.* **B47**, 987-997.
- Ducros, V., Czjzek, M., Belaich, A, Gaudin, C., Fierobe, H-P, Belaich, J-P, Davies, G.J. and Haser, R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* **3**, 939-949.



- Ebner, K.E., Denton, W.L. and Brodbeck, U. (1966) The substitution of alpha-lactalbumin for the B protein of lactose synthetase. *Biochem. Biophys. Res. Commun.* **24**, 232-236.
- Ely, K.R., Girling, R.L., Schiffer, M., Cunningham, D.E. and Edmundson, A.B. (1973) Preparation and properties of crystals of a Bence-Jones dimer with mercury inserted into the interchain disulphide bond. *Biochemistry* **12**, 4233-4237.
- Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray structure refinement. *Acta Cryst.* **A47**, 392-400.
- Evans, S.C., Youakim, A. and Shur, B.D. (1994) Biological consequences of targeting  $\beta$ 1,4-galactosyltransferase to two different subcellular compartments. *BioEssays* **17**, 261-268.
- Fenna, R.E. (1982a) Crystallization of cow  $\alpha$ -lactalbumin. *J. Mol. Biol.* **161**, 203-210.
- Fenna, R.E. (1982b) Crystallization of human  $\alpha$ -lactalbumin. *J. Mol. Biol.* **161**, 211-215.
- Fitzgerald, P.M.D. (1988) *MERLOT*, an integrated package of computer programs for determination of crystal structures by molecular replacement. *J. Appl. Cryst.* **21**, 274-278.
- Furey and Swaminathan (1990) *PHASES* – A program package for the processing and analysis of diffraction data from macromolecules. *American Crystallographic Association Meeting Abstracts* **18**, 73.
- Fujinaga, M. and Read, R.J. (1987) Experiences with a new translation-function program. *J. Appl. Cryst.* **20**, 517-521.
- Garman, E.F. (1995) Modern methods for rapid X-ray diffraction data collection from crystals of macromolecules. in *Methods in Molecular Biology* **7**, The HUMANA Press Inc., New Jersey, USA.
- Geren, C.R., Magee, S.C. and Ebner, K.E. (1975) Circular dichroism changes in galactosyltransferase upon substrate binding. *Biochemistry* **14**, 1461-1463.
- Gerken, T.A. (1984) Amino group environments and metal-binding properties of carbon-13 reductively methylated bovine alpha-lactalbumin. *Biochemistry* **23**, 4688-4697.
- Gewirth, D. (1995) *The HKL Manual: A description of the programs DENZO, XDisplayF and SCALEPACK*. Yale University, New Haven, Connecticut.
- Giegé, R. and Ducriux, A. (1992) Editors *Crystallization of Proteins and Nucleic acids: A Practical Approach.*, IRL Press, Oxford University Press, U.K.
- Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.
- Godovac-Zimmermann, J., Conti, A. and Napolitano, L. (1988) The primary structure of donkey (*Equus asinus*) lysozyme contains the Ca(II) binding site of alpha-lactalbumin. *Biol. Chem. Hoppe-Seyler* **269**, 1109-1115.

- Godovac-Zimmermann, J., Conti, A. and Napolitano, L. (1990) The complete primary structure of  $\alpha$ -lactalbumin isolated from pig (*Sus scrofa*) milk. *Biol. Chem. Hoppe-Seyler* **371**, 649-653.
- Grobler, J.A., Rao, R., Pervaiz, S. and Brew, K. (1994) Sequences of two highly divergent canine type c lysozymes: Implications for the evolutionary origins of the lysozyme /  $\alpha$ -lactalbumin superfamily. *Arch. Biochem. Biophys.* **313**, 360-366.
- Grütter, M.G., Weaver, L.H. and Matthews, B.W. (1983) Goose lysozyme structure: An evolutionary link between hen and bacteriophage lysozymes? *Nature* **303**, 828-831.
- Hadfield, A.T., Harvey, D.J., Archer, D.B., MacKenzie, D.A., Jeenes, D.J., Radford, S.E., Lowe, G., Dobson, C.M. and Johnson, L.N. (1994) Crystal structure of the mutant D52S Hen egg-white lysozyme with an oligosaccharide product. *J. Mol. Biol.* **243**, 856-872.
- Hall, L., Craig, R.K., Edbrooke, M.R. and Campbell, P.N. (1982) Comparison of the nucleotide sequence of cloned human and guinea-pig pre-alpha-lactalbumin cDNA with that of chick pre-lysozyme cDNA suggests evolution from a common ancestral gene. *Nucleic Acids Res.* **10**, 3503-3515.
- Hall, L. and Campbell, P.N. (1986)  $\alpha$ -Lactalbumin and related proteins: A versatile gene family with an interesting parentage. In "Essays in Biochemistry" **22**, 1-26.
- Hall, L., Emery, D.C., Davies, M.S., Parker, D. and Craig, R.K. (1987) Organisation and sequence of the human alpha-lactalbumin gene. *Biochem J.* **242**, 735-742.
- Harada, Y., Lifchitz, A., Berthou, J. and Jolles, P. (1981) A translation function combining packing and diffraction information - an application to lysozyme (high-temperature form). *Acta Cryst.* **A37**, 398-406.
- 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.
- Harushima, Y. and Sugai, S. (1989) Hydrogen exchange of the tryptophan residues in bovine, goat, guinea-pig and human  $\alpha$ -lactalbumin. *Biochemistry* **28**, 8568-8576.
- Heine, W.E., Klein, P.D. and Reeds, P.J. (1991) The importance of alpha-lactalbumin in infant nutrition. *J. Nutrition* **121**, 277-283.
- Hendrickson, W.A. (1985) Stereochemically restrained refinement of macromolecular structures. *Met. Enzymol.* **115**, 252-270.
- Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous synchrotron radiation. *Science* **254**, 51-58.
- Higuchi, R., Krummel, B. and Saiki, R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: Study of protein and DNA interactions. *Nucleic Acid Res.* **16**, 7351-7367.
- Hill, R.L. and Brew, K. (1975) Lactose synthetase. *Adv. Enzymol. Rel. Areas Mol. Biol.* **43**, 411-489.
- Hiroaka, Y., Segawa, Y., Kuwajima, K., Sugai, S. and Murai, N. (1980)  $\alpha$ -lactalbumin: a calcium metalloprotein. *Biochem. Biophys. Res. Commun.* **93**, 1098-1104.

- Holler, E., Rupley, J.A. and Hess, G.P. (1975) Productive and nonproductive lysozyme-chitosaccharide complexes. *Biochemistry* **14**, 2377-2385.
- Hopp T.P. and Woods K.R. (1979) The primary sequence of rabbit alpha-lactalbumin. *Biochemistry* **18**, 5182-5192.
- Hopper, K.E and McKenzie, H.A. (1974) Comparative studies of alpha-lactalbumin and lysozyme: echidna lysozyme. *Mol. Cell. Biochem.* **3**, 93-108.
- Huber, R. and Schneider, M. (1985) A group refinement procedure in protein crystallography using Fourier-transforms. *J. Appl. Cryst.* **18**, 165-169.
- Huber, R. (1985) Experience with the application of Patterson search techniques. In "Molecular Replacement" Proceedings of the Daresbury Study Weekend, 15-16 February, 1985, pp. 58-61. SERC Daresbury Laboratory, Warrington, England.
- Ikeguchi, M., Sugai, S., Fujino, M., Sugawara, T. and Kuwajima, K. (1992) Contribution of the 6-120 disulfide bond of  $\alpha$ -lactalbumin to the stabilities of its native and molten globule states. *Biochemistry* **31**, 12695-12700.
- Imoto, T., Johnson, L.N., North, A.C.T., Phillips, D.C. and Rupley, J.A. (1972) Vertebrate Lysozymes in "The Enzymes" Vol VII (ed. Boyer, P), pp 665-868, Academic Press, London.
- Inaka, K., Kuroki, R., Kikuchi, M. and Matsushima, M. (1991) Crystal structures of the apo- and holomutant human lysozymes with an introduced  $\text{Ca}^{2+}$  binding site. *J. Biol. Chem.* **266**, 20666-20671.
- Inman, J.K. and Bryan, R.F. (1966) The unit cell of crystalline  $\alpha$ -lactalbumin. *J. Mol. Biol.* **15**, 683-686.
- Jack, A. and Levitt, M. (1978) Refinement of large structures by simultaneous minimisation of energy and *R*-factor. *Acta Cryst.* **A34**, 931-955.
- Jancarik, J. and Kim, S. (1991) Sparse matrix sampling: a screening method for crystallisation of proteins. *J. Appl. Cryst.* **24**, 409-411.
- Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J. and Pavletich, N.P. (1995) Mechanism of CDK activation revealed by the structure of a cyclin A-CDK2 complex. *Nature* **376**, 313-320.
- Jenness, R. (1970) Protein Composition of Milk. In "Milk Proteins, Chemistry and Molecular Biology" (eds. McKenzie, H.A) pp. 17-43. Academic Press, New York.
- Johnson, L.N., Cheetham, J., McLaughlin, P.J., Acharya, K.R., Barford, D. and Phillips, D.C. (1988) Protein-oligosaccharide interactions: Lysozyme, phosphorylase, amylases. *Curr. Top. Microbiol. Immun.* **139**, 81-134.
- Jones, T.A. (1985) Interactive computer graphics: *FRODO Meth. Enzymol.* **115** 151-171.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.
- Jones, T.A. and Kjeldgaard, M. (1993) *O* – The Manual: Manual for *O* version 5.9.

- Kabsch, W.J. and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.
- Kabsch, W.J. (1988a) Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21**, 67-71.
- Kabsch, W.J. (1988b) Evaluation of single-crystal X-ray diffraction from a position-sensitive detector. *J. Appl. Cryst.* **21**, 916-924.
- Kabsch, W.J. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800.
- Khatra, B.S., Herries, D.G. and Brew, K. (1974) Some kinetic properties of human milk galactosyltransferase. *Eur. J. Biochem.* **44**, 537-560.
- Kleywegt, G.J. and Jones, T.A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst.* **D50**, 178-185.
- Kleywegt, G.J. and Jones, T.A. (1995a) Where freedom is given, liberties are taken. *Structure*, 535-540.
- Kleywegt, G.J. and Jones, T.A. (1995b) Braille for pugilists. In "Making the most of your model". Proceedings of the CCP4 Study Weekend. 6-7 January, 1995 (eds. Hunter, W.N., Thornton, J.M. and Bailey, S.) pp. 11-24, CLRC, Daresbury, UK.
- Koga, K. and Berliner, L.J. (1985) Structural elucidation of a hydrophobic box in bovine  $\alpha$ -lactalbumin by NMR: Nuclear Overhauser effects. *Biochemistry* **24**, 7257-7262.
- Kraulis, P.J. (1991) *MOLSCRIPT*: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
- Krezdorn, C.H., Watzele, G., Kleene, R.B., Ivanov, S.X. and Berger, E.G. (1993) Purification and characterization of recombinant human  $\beta$ 1-4 galactosyltransferase expressed in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **212**, 113-120.
- Kronman, M.J., Sinha, S.K. and Brew, K. (1981) Characteristics of the binding of  $\text{Ca}^{2+}$  and other divalent metal ions to bovine alpha-lactalbumin. *J. Biol. Chem.* **256**, 8582-8587.
- Kronman, M.J. and Bratcher, S.C. (1983) An experimental artifact in the use of chelating ion buffers. Binding of chelators to bovine alpha-lactalbumin. *J. Biol. Chem.* **258**, 5707-5709.
- Kronman, M.J. (1989) Metal-ion binding and the molecular conformational properties of  $\alpha$ -lactalbumin. *CRC Crit. Rev. Biochem. Mol. Biol.* **24**, 565-667.
- Kuhn, N.J., Stankiewicz, M. and Ward, S. (1992) Cation activation and stabilisation of golgi  $\beta$ 1,4-galactosyltransferase (lactose synthetase). *Biochem. Soc. Trans.* **20**, 714-716.
- Kumagai, I., Takeda, S., Hibino, T. and Miura, K. (1990) Expression of goat  $\alpha$ -lactalbumin in *Escherichia coli* and its refolding to biologically active protein. *Protein Engineering* **3**, 449-452.

- Kumagai, I., Takeda, S. and Miura, K. (1992) Functional conversion of the homologous proteins  $\alpha$ -lactalbumin and lysozyme by exon exchange. *Proc. Natl. Acad. Sci. USA* **89**, 5887-5891.
- Kumagai, I., Maenaka, K., Sunada, F., Takeda, S. and Miura, K. (1993) Effects of subsite alterations on substrate binding mode in the active site of hen egg-white lysozyme. *FEBS* **212**, 151-156.
- Kumar, S., Cheng, X., Pflugrath, J.W. and Roberts, R.J. (1992) Purification, crystallisation and preliminary X-ray diffraction analysis of a *M. HhaI*-AdoMet Complex. *Biochemistry* **31**, 8648-8653.
- Kuriyan, J., Brünger, A.T., Karplus, M. and Hendrickson, W.A. (1987) X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. *Acta Cryst.* **A45**, 396-409.
- Kuroki, R., Taniyama, Y., Seko, C., Nakamura, H., Kikuchi, M. and Ikehara, M. (1989) Design and creation of a  $\text{Ca}^{2+}$  binding site in human lysozyme to enhance stability. *Proc. Natl. Acad. Sci. USA*. **86**, 6903-6907.
- Kuwajima, K., Harushima, Y. and Sugai, S. (1986) Influence of calcium-binding on the structure and stability of bovine  $\alpha$ -lactalbumin studied by circular dichroism and nuclear magnetic resonance spectra. *Int. J. Pept. Protein Res.* **27**, 18-27.
- Kuwajima, K. (1989) The molten globule as a clue for understanding the folding and cooperativity of globular protein structures. *Prot. Struct. Func. Genet.* **6**, 87-103.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *PROCHECK*: a program to check stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
- Lattman, E.E. (1985) Use of rotation and translation functions. *Meth. Enzymol.* **115**, 55-77.
- Lambright, D.G., Lee, T.K. and Wong, S.S. (1985) Association-dissociation modulation of enzymatic activity: Case of lactose synthase. *Biochemistry* **24**, 910-914.
- Lee, T.K., Wong, L.C. and Wong, S.S. (1983) Photoaffinity labelling of lactose synthase with a UDP-galactose analogue. *J. Biol. Chem.* **258**, 13166-13171.
- Lescar, J., Riottot, M-N., Souchon, H., Chitarra, V., Bentley, G.A., Navaza, J., Alzari, P.M. and Poljak, R.J. (1993) Crystallisation, preliminary X-ray diffraction study, and crystal packing of a complex between anti-hen lysozyme antibody F9.13.7 and guinea-fowl lysozyme. *Prot. Struct. Func. Genet.* **15**, 209-212.
- Leslie, A.G.W. (1991) Heavy atom derivative screening. In "Isomorphous Replacement and Anomalous Scattering" Proceedings of the Daresbury Study Weekend, 25-26 January, 1991. (eds. Wolf, W., Evan, P.R., Leslie, A.G.W.) pp. 9-22, SERC Daresbury Laboratory, Warrington, England.
- Ley, J.M. and Jenness, R. (1970) Lactose synthetase activity of  $\alpha$ -lactalbumins from several species. *Arch Biochem. Biophys.* **138**, 464-469.

- Linzell, J.L. and Peaker, M. (1971) Mechanism of milk secretion. *Physiol. Rev.* **51**, 564-597.
- Luzatti, V. (1952) Traitement statistique des erreurs dans la determination des structures cristallines *Acta Cryst.* **5**, 802-810.
- Machin, P.A. (1985) Editor. "Molecular Replacement" Proceedings of the Daresbury Study Weekend, 15-16 February, 1985. SERC Daresbury Laboratory, Warrington, England.
- Malcolm, B.A., Rosenberg, S., Corey, M.J., Allen, J.S., de Baetselier, A. and Kirsch, J.F. (1989) Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg-white lysozyme. *Proc. Natl. Acad. Sci. USA.* **86**, 133-137.
- Matthews (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
- McKenzie, H.A. and White, F.H. Jr. (1987) Studies on a trace cell lytic activity associated with  $\alpha$ -lactalbumin. *Biochem. Int.* **14**, 347-356.
- McKenzie, H.A. and White, F.H. Jr. (1991) Lysozyme and  $\alpha$ -lactalbumin: Structure, function and interrelationships. *Adv. Prot. Chem.* **41**, 173-315.
- McPherson, A. (1982) in "The Preparation and Analysis of Protein crystals." John Wiley and Sons, New York.
- McPherson, A. (1990) Current approaches to macromolecular crystallisation. *Eur. J. Biochem.* **189**, 1-23.
- McPherson, A., Koszelak, S., Axelrod, H., Day, J., Williams, R., Robinson, L., McGrath, M. Cascio, D. (1986) An experiment regarding crystallisation of soluble proteins in the presence of  $\beta$ -octyl glucoside. *J. Biol. Chem.* **261**, 1969-1975.
- Merritt, E.A. and Murphy, M.E.P. (1994) *RASTER3D* version 2.0: A program for photorealistic molecular graphics. *Acta Cryst.* **D50**, 869-873.
- Morrison, J.F. and Ebner, K.E. (1971a) Studies on galactosyltransferase. 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.
- Morrison, J.F. and Ebner, K.E. (1971c) Studies on galactosyltransferase. Kinetic effects of  $\alpha$ -lactalbumin with N-acetylglucosamine and glucose as galactosyl group acceptors. *J. Biol. Chem.* **246**, 3992-3998.
- Murakami, K. and Berliner, L.J. (1983) A distinct zinc-binding site in the alpha-lactalbumins regulates calcium binding. Is there a physiological role for this control? *Biochemistry* **22**, 3370-3374.
- Musci, G. and Berliner, L.J. (1985) Physiological roles of zinc and calcium binding to alpha-lactalbumin in lactose biosynthesis. *Biochemistry*, **24**, 6945-6948.
- Nagai, K., Evans, P.R., Li, J. and Oubridge, C. (1991) Phase determination using mercury derivatives of engineered cysteine mutants. In "Isomorphous Replacement

- and Anomalous Scattering*" Proceedings of the Daresbury Study Weekend, 25-26 January, 1991. (eds. Wolf, W., Evan, P.R., Leslie, A.G.W.) pp. 141-149, SERC Daresbury Laboratory, Warrington, England.
- Navaza, J. (1987) On the fast rotation function. *Acta. Cryst.* **A43**, 645-653.
- Navaza, J. (1990) Accurate computation of the rotation matrices. *Acta. Cryst.* **A46**, 619-620.
- Navaza, J. (1992) *AMoRe*: A new package for molecular replacement. In "*Molecular Replacement*" Proceedings of the CCP4 Study Weekend, 31 January-1 February, 1992, pp. 87-90. SERC Daresbury Laboratory, Warrington, England.
- Navaza, J. (1994) *AMoRe*: an automated package for molecular replacement. *Acta. Cryst.* **A50**, 157-163.
- Nicholls, A., Sharp, K. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
- Nitta, K., Tsuge, H., Shimazaki, K. and Sugai, S. (1988) Calcium binding lysozymes. *Biol. Chem. Hoppe-Seyler* **369**, 1109-1115.
- Nitta, K. and Sugai, S. (1989) The evolution of lysozyme and alpha-lactalbumin. *Eur. J. Biochem.* **182**, 111-118.
- O'Keefe, E.T., Hill, R.L. and Bell, J.E. (1980a) Active site of bovine galactosyltransferase: Kinetic and fluorescence studies. *Biochemistry* **19**, 4954-4962.
- O'Keefe, E.T., Mordick, T. and Bell, J.E. (1980b) Bovine galactosyltransferase: Interaction with  $\alpha$ -lactalbumin and the role of  $\alpha$ -lactalbumin in lactose synthase. *Biochemistry* **19**, 4962-4966.
- Otwinowski, Z. (1993) Oscillation data reduction program. In "*Data Collection and Processing*" Proceedings of the CCP4 Study Weekend, 29-30 January, 1993, pp. 56-62. SERC Daresbury Laboratory, Warrington, England.
- Patterson A.L.(1934) *Phys. Rev.* **46**, 372.
- Paulson, J.C. and Colley, K.J. (1989) Glycosyltransferases. Structure, localisation, and control of cell type-specific glycosylation. *J. Biol. Chem.* **264**, 17615-17618.
- Peng, Z., Wu, L.C. and Kim, P.S. (1995) Local structure preferences in the  $\alpha$ -lactalbumin molten globule. *Biochemistry* **34**, 3248-3252.
- Perkins, S.J., Johnson, L.N., Machin, P.A. and Phillips, D.C. (1978) Crystal structures of egg-white lysozyme of hen in acetate-free medium and of lysozyme complexes with N-acetylglucosamine and  $\beta$ -methyl N-acetylgulcosaminide. *Biochem. J.* **173**, 607-616.
- Perkins, S.J., Johnson, L.N., Phillips, D.C. and Dwek, R.A. (1981) The binding of monosaccharide inhibitors to hen egg-white lysozyme by proton magnetic resonance at 270MHz and analysis by ring-current calculations. *Biochem. J.* **193**, 553-572.
- Permyakov, E.A., Yarmolenko, V.V., Kalinichenko, L.P., Morozova, L.A. and Burstein, E.A. (1981a) 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.
- Permyakov, E.A., Kalinichenko, L.P., Morozova, L.A., Yarmolenko, V.V. and Burstein, E.A. (1981b)  $\alpha$ -lactalbumin binds magnesium ions: Study by means of intrinsic fluorescence technique. *Biochem. Biophys. Res. Commun.* **102**, 1-7.
- Petsko, G.A. (1975) Protein crystallography at sub-zero temperatures: Cryo-protective mother liquors for protein crystals. *J. Mol. Biol.* **96**, 381-392.
- Petsko, G.A. (1985) Preparation of Isomorphous Heavy-atom Derivatives. *Met. Enzymol.* **114**, 147-156.
- Phillips, D.C., Acharya, K.R., Handoll, H.H.G. and Stuart, D.I. (1987) From lysozyme structure to enzyme engineering. *Biochem. Soc. Trans.* **15**, 737-744.
- Pierce, M., Turley, E.A. and Roth, S. (1980) Cell-surface glycosyltransferase activities. *Int. Rev. Cytol.* **65**, 1-47.
- Pincus, M.R. and Scheraga, H.A. (1981) Prediction of the three-dimensional structure of complexes of lysozyme with cell wall substrates. *Biochemistry* **20**, 3960-3965.
- Post, C.B., Brooks, B.R., Karplus, M., Dobson, C.M., Artymuik, P.J., Cheetham, J.C. and Phillips, D.C. (1986) Molecular dynamics simulations of native and substrate-bound lysozyme. *J. Mol. Biol.* **190**, 455-479.
- Powell, M.J.D. (1977) In "Mathematical Programming" **12**, 241-254.
- Powell, J.T. and Brew, K. (1974) Isolation and characterisation of two forms of bovine galactosyl transferase. *Eur. J. Biochem.* **48**, 217-228.
- Powell, J.T. and Brew, K. (1976a) Metal ion activation of galactosyltransferase. *J. Biol. Chem.* **251**, 3645-3652.
- Powell, J.T. and Brew, K. (1976b) A comparison of the interactions of galactosyltransferase with a glycoprotein substrate (ovalbumin) and with  $\alpha$ -lactalbumin. *J. Biol. Chem.* **251**, 3653-3663.
- Prager, E.M. and Wilson, A. C. (1988) Ancient origin of alpha-lactalbumin from lysozyme: analysis of DNA and amino acid sequences. *J. Mol. Evol.* **27**, 326-335.
- Prasad, R.V., Butkowski, R.J., Hamilton, J.W. and Ebner, K.E. (1982) Amino acid sequence of rat alpha-lactalbumin. A unique alpha-lactalbumin. *Biochemistry* **21**, 1479-1482.
- Prieels, J.P., Dolmans, N., Leonis, J. and Brew, K. (1975) Nitration of tyrosyl residues in human  $\alpha$ -lactalbumin. Effect on lactose synthase specifier activity. *Eur. J. Biochem.* **60**, 533-539.
- Prieels, 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. *Biochemistry* **18**, 1771-1776.
- Qasba, P.K. and Safaya, S.K. (1984) Similarity of the nucleotide sequences of rat alpha-lactalbumin and chicken lysozyme genes. *Nature* **308**, 377-380.



- Qian, M., Haser, R., Buisson, G., Duée, E and Payan, F. (1994) The active center of a mammalian  $\alpha$ -amylase. Structure of the complex of a pancreatic  $\alpha$ -amylase with a carbohydrate inhibitor refined at 2.1Å resolution. *Biochemistry* **33**, 6284-6289.
- Quarforth, G.J. and Jenness, R. (1975) Isolation, composition and functional properties of  $\alpha$ -lactalbumins from several species. *Biochem. Biophys. Acta.* **379**, 476-487.
- Quioco, F.A. (1986) Carbohydrate-binding proteins: Tertiary structures and protein sugar interactions. *Annu. Rev. Biochem.* **55**, 287-315.
- Rao, K.R. and Brew, K. (1989) Calcium regulates folding and disulphide-bond formation in  $\alpha$ -lactalbumin. *Biochem. Biophys. Res. Commun.* **163**, 1390-1396.
- Rao, S.N., Jin, J-H. and Hartsuck, J.A. (1980) Rotation-function space groups. *Acta Cryst.* **A36**, 878-886.
- Radford, S.E. and Dobson, C.M. (1995) Insights into protein folding using physical techniques: studies of lysozyme and  $\alpha$ -lactalbumin. *Phil. Trans. R. Soc. Lond. B.* **348**, 17-25.
- Read R.J. (1986) Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149.
- Read R.J. (1990) Structure-factor probabilities for related structures. *Acta Cryst.* **A46**, 900-912.
- Ren, J., 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, New York.
- Richardson, R.H and Brew, K. (1980) Lactose synthase. *J. Biol. Chem.* **255**, 3377-3385.
- Rius, J. and Miravittles, C. (1986) A full-symmetry translation function - The influence of model misorientation. *Acta Cryst.* **A42**, 402-404.
- Rosenberg, A.H., Lade, B.N., Chui, D-S., Lin, S-W., Duma, J.J. and Studier, F.W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**, 125-135.
- Rossmann, M.G. and Blow, D.M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta. Cryst.* **15**, 24-31.
- Rossmann, M.G. (1972) *The Molecular Replacement Method*. Gordon and Breach, New York, USA.
- Saludjian, P., Prange, T., Navaza, J., Menez, R., Guilloteau, J.P., Ries-Kautt, M. and Ducruix, A. (1992) Structure determination of a dimeric form of erabutoxin-b crystallised from a thiocyanate solution. *Acta Cryst.* **B48**, 520-531.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) "*Molecular Cloning: A Lab Manual.*", pp. 17.38, Cold Spring Harbor Laboratory Press.
- Sambrook, J.F. (1990) The involvement of  $\text{Ca}^{2+}$  in transport of secretory proteins from the endoplasmic reticulum. *Cell* **61**, 197-199.

- Sarkar, G. and Sommer, S.S (1990) The "megaprimer" method of site-directed mutagenesis. *BioTechniques* **8**, 404-407.
- Saxena, V.P. and Wetlaufer, D.B. (1970) Formation of three-dimensional structures in proteins. I. Rapid non-enzymic reactivation of reduced lysozyme. *Biochemistry* **9**, 5015-5022.
- Sawano, H., Koumoto, Y., Ohta, K., Sasaki, Y., Segawa, S. and Tachibana, H. (1992) Efficient *in vitro* folding of the three-disulphide derivatives of hen lysozyme in the presence of glycerol. *FEBS Lett.* **303**, 11-14.
- Schechter, Y., Patchornik, A. and Burstein, Y. (1974) Selective sulphenylation of tryptophan residues in  $\alpha$ -lactalbumin of bovine milk. *J. Biol. Chem.* **249**, 413-419.
- Schindler, M., Sharon, N. and Prieels, 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.
- Shaw, D.C., Messer, M., Scrivener, A.M., Nicholas, K.R. and Griffiths, M. (1993) Isolation, partial characterisation, and amino acid sequence of  $\alpha$ -lactalbumin from platypus (*Ornithorhynchus anatinus*) milk. *Biochem. Biophys. Acta.* **1161**, 177-186.
- Shaper, J.H. and Shaper, N.L. (1992) Enzymes associated with glycosylation. *Curr. Opin. Struct. Biol.* **2**, 701-709.
- Shewale, J.G., Sinha, S.K. and Brew, K. (1984) Evolution of  $\alpha$ -lactalbumins. *J. Biol. Chem.* **259**, 4974-4956.
- Shur, B.D. (1984) The receptor function of galactosyltransferase during cellular interactions. *Mol. Cell. Biochem.* **61**, 143-158.
- Sigler P.B. and Blow D.M. (1965) A means of promoting heavy-atom binding in protein crystals. *J. Mol. Biol.* **14**, 640.
- Sinha, S.K. and Brew, K. (1981) A label selection procedure for determining the location of protein-protein interaction sites by cross-linking with bisimidoesters. *J. Biol. Chem.* **256**, 4193-4204.
- 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 three-dimensional structure of baboon  $\alpha$ -lactalbumin at low resolution. *Biochem. J.* **242**, 353-360.
- 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.
- Song, H., Inaka, K., Maenaka, K. and Matsushima, M. (1994) Structural changes of active site cleft and different saccharide binding modes in human lysozyme co-crystallised with hexa-*N*-acetyl-chitohexaose at pH4.0. *J. Mol. Biol.* **244**, 522-540.
- Stacey, A., Schnieke, A., Kerr, M., Scott, A., McKee ,C., Cottingham, I., Binas, B., Wilde, C. and Colman, A. (1995) Lactation is disrupted by alpha-lactalbumin

- deficiency and can be restored by human alpha-lactalbumin gene replacement in mice *Proc. Natl. Acad. Sci. USA* **92**, 2835-2839.
- Stinnakre, M.G., Vilotte, J.L., Soulier, S. and Mercier, J.C. (1994) Creation and phenotypic analysis of alpha-lactalbumin deficient mice. *Proc. Natl. Acad. Sci. USA* **91**, 6544-6548.
- Strous, G.J. (1986) Golgi and secreted galactosyltransferase. *CRC Crit. Rev. Biochem.* **21**, 119-151.
- Strynadka, N.C.J. and James, N.G. (1989) Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem.* **58**, 951-998.
- Strynadka, N.C.J. and James, N.G. (1991) Lysozyme revisited: Crystallographic evidence for distortion of an N-acetylmuramic acid residue bound in site D. *J. Mol. Biol.* **220**, 401-424.
- Stuart, D.I., Levine, M., Muirhead, H. and Stammers, D.K. (1979) Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6Å. *J. Mol. Biol.* **134**, 109-142.
- 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.
- Studier, F.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.* **185**, 60-89.
- Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.
- Stura, E.A. and Wilson, I.A. (1992) Seeding techniques In "*Crystallization of Proteins and Nucleic acids: A Practical Approach.*" (eds. Ducruix, A. & Giegé, R.) pp. 99-126, IRL Press, Oxford University Press, U.K.
- Takase, K. and Ebner, K.E. (1981) Interactions of substrates and  $\alpha$ -lactalbumin with galactosyltransferase as measured by difference spectroscopy. *J. Biol. Chem.* **256**, 7269-7276.
- Takase, K. and Ebner, K.E. (1984) Interaction of galactosyltransferase with  $\alpha$ -lactalbumin and substrates. *Curr. Top. Cell. Reg.* **24**, 51-62.
- Tickle, I.J. (1985) Review of space group general translation functions that make use of known structure information and can be expanded as fourier series. In "*Molecular Replacement*" Proceedings of the CCP4 Study Weekend, 15-16 February, 1985, (eds. Machin, P.A.) pp. 22-26. SERC Daresbury Laboratory, Warrington, England.
- Tickle, I.J. (1992) Fast fourier translation functions. In "*Molecular Replacement*" Proceedings of the CCP4 Study Weekend, 31 January-1 February, 1992, (eds. Dodson, E.J., Gover, S., Wolf, W.) pp. 20-32. SERC Daresbury Laboratory, Warrington, England.
- Tucker, A.D., Baty, D., Parker, M.W., Pattus, F., Lazdunski, C. and Tsernoglou, D. (1989) Crystallographic phases through genetic engineering: experiences with colicin A. *Prot. Eng.* **2**, 399-405.

- Turkington, R.W., Brew, K., Vanaman, T.C. and Hill, R.L. (1968) The hormonal control of lactose synthetase in the developing mouse mammary gland. *J. Biol. Chem.* **243**, 2282-3387.
- Viaene, A., Volckaert, G., Joniau, M., Baetselier, A. Cauwelaert, F. (1991) Efficient expression of bovine  $\alpha$ -lactalbumin in *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **202**, 471-477.
- 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.
- Wang, Y., Wong, S.S., Fukuda, M.N., Zu, H.Y., Liu, Z.D., Tang, Q.S. and Appert, H.E. (1994) Identification of functional cysteine residues in human galactosyltransferase. *Biochem. Biophys. Res. Commun.* **204**, 701-709.
- Warne, P.K., Momany, F.A., Rumball, S.V., Tuttle, R.W. and Scheraga, H.A. (1974) Computation of structures of homologous proteins. Alpha-lactalbumin from lysozyme. *Biochemistry* **13**, 768-781.
- Watkins, W.M. and Hassid, W.Z. (1962) The synthesis of lactose by particulate enzyme preparations from guinea-pig and bovine mammary glands. *J. Biol. Chem.* **237**, 1432-1440.
- Weber, P.C. (1991) Physical principles of protein crystallisation. *Adv. Prot. Chem.* **41**, 1-36.
- Weaver, L.H., Grütter, M.G. and Matthews, B.W. (1995) The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "gossetype" lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* **245**, 54-68.
- Wilson, K.S. (1994) Refinement at atomic resolution. In "From First Map to Final Model." Proceeding of the CCP4 Study Weekend, 6-7 January, 1994. (Eds. Bailey, S., Hubbard, R. and Waller, D.) pp.141-147, SERC Daresbury Laboratory, Warrington, UK.
- Wu, L.C., Peng, Z-Y. and Kim, P.S. (1995) Bipartite structure of the  $\alpha$ -lactalbumin molten globule. *Nat. Struct. Biol.* **2**, 281-285.
- Yadav, S.P. and Brew, K. (1990) Identification of a region of UDP-galactose N-acetylglucosamine  $\beta$ 1, 4-galactosyltransferase involved in UDP-galactose binding by differential labeling. *J. Biol. Chem.* **265**, 14163-14169.
- Yadav, S.P. and Brew, K. (1991) Structure and function in galactosyltransferase. *J. Biol. Chem.* **266**, 698-703.
- Yu, N-T. (1974) Comparison of protein structure in crystals, in lyophilised state, and in solution by laser Raman scattering. III.  $\alpha$ -lactalbumin. *J. Am. Chem. Soc.* **96**, 4664-4668.
- Zou, J.Y. and Mowbray, S.L. (1994) An evaluation of the use of databases in protein-structure refinement. *Acta Cryst.* **D50**, 237-249.

## References

- Zu, H., Fukuda, M.N., Wong, S.S., Wang, Y., Liu, Z., Tang, Q. and Appert, H.E. (1995) Use of site-directed mutagenesis to identify the galactosyltransferase binding sites for UDP-galactose. *Biochem. Biophys. Res. Commun.* **206**, 362-369.

# Reprints

## Study by Mutagenesis of the Roles of Two Aromatic Clusters of $\alpha$ -Lactalbumin in Aspects of Its Action in the Lactose Synthase System\*

(Received for publication, October 14, 1993, and in revised form, November 1, 1993)

Jay A. Grobler<sup>‡</sup>, Mei Wang<sup>§</sup>, Ashley C. W. Pike<sup>¶</sup>, and Keith Brew<sup>||</sup>

From the Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, Florida 33101

A new system for the bacterial expression of a variant of bovine  $\alpha$ -lactalbumin has been developed. Eighteen mutant proteins containing single site substitutions in a cluster of predominantly aromatic residues adjacent to the cleft (aromatic cluster I) and in the hydrophobic box were expressed. The proteins were extracted from inclusion bodies and treated to generate native folding and disulfide bonds to provide appropriately folded protein samples for nine of the mutants. These were characterized with respect to kinetic parameters reflecting aspects of  $\alpha$ -lactalbumin activity in modulating the specificity of bovine galactosyltransferase. In aromatic cluster I, changes at tryptophan 118 or glutamine 117 were found to specifically reduce affinity for galactosyltransferase, whereas substitutions for phenylalanine 31 or histidine 32 have major effects on the ability to promote glucose binding (13–200-fold) and lesser effects on galactosyltransferase affinity (1.5–70-fold). Substitutions in the hydrophobic box were found to affect folding rather than activity. Thus, the binding of  $\alpha$ -lactalbumin to galactosyltransferase and its ability to promote glucose binding can be separately perturbed and are associated with distinct but adjacent structures. Aromatic cluster I is directly involved in activity whereas the hydrophobic box appears to have a structural rather than functional role.

The milk whey protein,  $\alpha$ -lactalbumin (LA)<sup>1</sup> functions as the regulatory component of the lactose synthase enzyme system (EC 2.4.1.22) which catalyzes and regulates the biosynthesis of lactose in the lactating mammary gland (for reviews, see Refs.

1–3). The catalytic component of lactose synthase is the enzyme UDP-galactose  $\beta$ -*N*-acetylglucosaminide  $\beta$ -4-galactosyltransferase (GT1; EC 2.4.1.38), an intrinsic component of the trans-golgi membranes which, in most cells, has the function of catalyzing the transfer of galactose from UDP-galactose to non-reducing terminal  $\beta$ -linked *N*-acetylglucosaminyl residues in the carbohydrate chains of glycoproteins and glycolipids. Although GT can catalyze lactose synthesis (transfer of galactose to free glucose) *in vitro*, it is ineffective as a catalyst for this reaction under physiological conditions because of its low affinity for glucose, reflected in a  $K_m$  of about 2 M. LA, a specific product of the lactating mammary gland, binds reversibly with GT to form 1:1 complexes with a concomitant 1000-fold decrease in the  $K_m$  for glucose, thereby allowing efficient production of lactose under physiological conditions.

The relationship of structure and function in LA is of particular interest for a number of reasons: (a) it is similar in sequence and three-dimensional structure to the type c lysozymes (4), providing a case of extreme functional divergence in a pair of homologous proteins, (b) it represents an example of enzyme regulation at the level of specificity through a heterologous protein-protein interaction, (c) it contains a tightly bound  $\text{Ca}^{2+}$  ion (5) that has a major influence on its molecular stability and folding properties (6–8), and (d) under mildly denaturing conditions it can assume a stable conformational state with the properties of a “molten globule,” an early intermediate in protein folding (2, 3, 7).

Here, we describe a facile bacterial expression system for a variant of bovine LA with which fully active recombinant LA can be generated in high yields. This system succeeds an earlier system in which bovine LA was expressed as a fusion protein (C-LA) from which a native, active protein was generated by limited proteolysis after extraction and treatment to generate native folding and disulfide bond formation (9). With the new system, site-directed mutagenesis has been used to investigate the roles of structural elements of the bovine LA molecule in its activity. One region, defined by x-ray crystallographic and NMR studies, is designated “aromatic cluster I” (4, 10) and contains the side chains of the invariant amino acids Phe-31, His-32, Gln-117, and Trp-118, together with Tyr-36. This structure is adjacent to the cleft and has been implicated by chemical studies to be involved in the interaction with GT (11–13). An alternative candidate for the GT binding site is a structure, designated the “hydrophobic box” or “aromatic cluster II” which includes Trp-26, Phe-53, Trp-60, Tyr-103, and Trp-104 (14). This structure is found both in LA and lysozyme, but is distinct in LA because of the presence of Tyr-103 (alanine in most lysozymes), which blockades the LA cleft, making it a less open structure than in lysozyme. To probe the roles of these substructures, we have constructed and expressed LA mutants with substitutions at selected sites in both clusters I and II. Mutant proteins that allowed the generation of significant

\* This work was supported in part by National Institutes of Health Research Grant GM21363. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>‡</sup> Recipient of a National Research Service Award.

<sup>§</sup> Current address: Dept. of Medicine, Duke University Medical Center, Durham, NC 27710.

<sup>¶</sup> Current address: Dept. of Biochemistry, Bath University, Bath, UK.

<sup>||</sup> To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, R-629, University of Miami School of Medicine, P.O. Box 016129, Miami, FL 33101. Tel.: 305-547-6297; Fax: 305-547-3955.

<sup>1</sup> The abbreviations used are: LA, bovine  $\alpha$ -lactalbumin; mLA, recombinant bovine  $\alpha$ -lactalbumin with  $\text{NH}_2$ -terminal methionine and valine at position 90; GT, UDP-galactose *N*-acetylglucosaminide  $\beta$ -1,4 galactosyltransferase; ChB, *N,N'*-diacetylchitobiose; HPLC, high performance liquid chromatography; PCR, polymerase chain reaction. Mutants of mLA are designated *XnY*-mLA, where *n* indicates the sequence position in bovine LA, *X*, the amino acid originally present at the site and *Y*, the amino acid introduced at that site (single letter code). It should be noted that the true sequence position in mLA is *n* + 1, because of the presence of the additional residue at the amino terminus.

quantities of native protein after folding, *in vitro*, were characterized with respect to the values of parameters that reflect different aspects of LA action. The relevance of the results to structure-function relationships in LA and the mechanism of action of LA are discussed as well as the facility of the expression system in comparison to others that have been previously described (9, 15, 16).

## EXPERIMENTAL PROCEDURES

### Materials

The sources of reagents for lactose synthase assays, protein, and DNA sequencing, and protein purification are the same as in previous publications (9, 17, 18). T7 promoter and T7 terminator primers and pET3a vector were purchased from Novagen. HP-PEI anion exchange HPLC columns (7.8  $\times$  100 mm) were obtained from Interaction Chromatography Inc., San Jose, CA. Taq polymerase, DNA ligase, and Magic™ PCR Preps DNA purification systems were purchased from Promega Corp., Madison WI. DNA ligase, *Bam*HI, and *Nde*I were from New England Biolabs. *N,N'*-Diacetylchitobiose was supplied by Sigma. Other reagents were of analytical grade.

### Bacterial Expression and Site-directed Mutagenesis

Bovine LA was expressed in *Escherichia coli* strain BL21(DE3) by using a vector (pMLA) generated by cloning the coding sequence for LA into the pET3a vector at a site adjacent to the codon for the initiator methionine. The coding region for mature bovine LA in which the codon for methionine 90 was modified by M13 site-directed mutagenesis (19) to valine was amplified by PCR using the primers designated NF-N and NF-C (see Table I). 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. Further mutations were introduced by PCR using the "megaprimer" method (20) with this construct as the template. The amplification to generate the megaprimer was performed in each case with the synthetic T7 promoter primer or the T7 terminator primer together with an appropriate mutagenic primer. Table I lists the mutagenic primers that were used for the different substitutions. The megaprimer was purified by using the Magic™ PCR Preps Purification Kit and used in a second amplification with the same template and the cognate T7 primer. After purification by agarose gel electrophoresis, the final amplification product was digested and cloned into pET3a as described for mLA. The expression vectors for mLA and variants were characterized by restriction mapping and DNA sequencing (21).

### Generation and Purification of Native Recombinant LAs

**Extraction and Initial Purification**—Cultures of *E. coli* strain BL21(DE3) transformed with pMLA were grown and induced as described previously (8). Harvested cells were suspended in lysis buffer and lysed by treatment with lysozyme and deoxycholate. The extract was subsequently treated with DNAase (22). Inclusion bodies containing mLA were isolated by centrifugation, washed by resuspension in lysis buffer containing 0.5% Triton X-100, and collected by centrifugation. The protein from the inclusion bodies was solubilized by incubation for 1–2 h at 37 °C with 8 M urea containing 20 mM Tris base and 20 mM dithiothreitol (about 10 ml/liter of culture). The pH of the extract was generally about 9.0. After centrifugation at 15,000 rpm for 10 min, the extract (from 4–6 liters of culture) was applied by pumping to a column (2  $\times$  10 cm) of MacroPrep 50Q (Bio-Rad) equilibrated with 20 mM Tris-HCl, pH 8.5, containing 4 M urea at room temperature. After washing with equilibration buffer until the absorbance at 280 nm was reduced nearly to base line, the column was eluted with a linear gradient (total 225 ml) from 0 to 0.5 M NaCl in 20 mM Tris-HCl, pH 8.5, containing 4 M urea. A flow rate of 2.5 ml/min was used throughout. SDS-gel electrophoresis of aliquots of different fractions indicated that a major peak, eluting at about 30 min, contained mLA.

**Folding and Disulfide Bond Formation**—Fractions containing mLA or a variant of mLA were pooled and the concentration of protein (essentially all mLA, based on SDS-gel electrophoresis) was estimated from the absorbance at 280 nm. A value of 2.0 for  $E_{280}^{1\%}$  was assumed for mLA, and corresponding values for mutants were calculated from their contents of tryptophan, tyrosine, and cystine (23). The protein was

diluted to a concentration of <1 mg/ml with 2 M urea containing 0.02 M Tris, pH 8.5, and mercaptoethanol, 2-hydroxyethyl disulfide, and CaCl<sub>2</sub> were added to give final concentrations of 5, 0.5, and 1 mM, respectively. The solution was dialyzed successively against 2 M urea, 1 M urea, and water, all containing 0.02 M Tris-HCl, pH 8.5, 5 mM mercaptoethanol, 0.5 mM 2-hydroxyethyl disulfide, and 1 mM CaCl<sub>2</sub> at 4 °C. In some cases, glycerol (10%) was included in all solutions. Finally, the protein was either precipitated by adjusting the pH to 6.5 and adding ammonium sulfate to 80% saturation, or dialyzed against 30 mM ammonium bicarbonate and lyophilized.

**Final Purification**—Recombinant LA and mutants were purified by gel filtration at room temperature (approximately 22 °C) with columns (5  $\times$  25 cm) of Sephadex G-75 equilibrated with 30 mM ammonium bicarbonate containing 10  $\mu$ M CaCl<sub>2</sub>. Typically, the mLA from a 4–6-liter preparation was dissolved in 10 ml of ammonium bicarbonate and applied to the column. Elution was performed with 30 mM ammonium bicarbonate containing 10  $\mu$ M CaCl<sub>2</sub> at a flow rate of 2.5 ml/min. An additional step of anion exchange HPLC was carried out in some cases using a Hewlett Packard 1090 liquid chromatograph fitted with a HP-PEI anion exchange column. Buffers used for this separation were: buffer A, 20 mM Tris-HCl, pH 7.8, containing 200  $\mu$ M CaCl<sub>2</sub>, and buffer B, 20 mM Tris-HCl containing 200  $\mu$ M CaCl<sub>2</sub> and 1.0 M NaCl. The column was equilibrated with 85% buffer A, 15% buffer B, and elution was performed by a linear gradient from 15 to 60% buffer B over 60 min. A flow rate of 1.0 ml/min was used throughout. Typically, native forms of bovine LA mLA and mutants eluted at 28–30 min, the presence of misfolded forms was indicated by peak(s) eluting at higher concentrations of NaCl.

### GT Assays

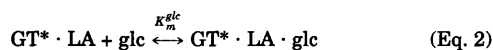
GT assays were carried out by a radiochemical procedure using UDP-[<sup>3</sup>H]galactose, as in previous studies (17, 18). All assays were performed in 10 mM Tris-HCl buffer, pH 7.5, 10 mM MnCl<sub>2</sub>, 0.1% BSA, and 0.3 mM UDP-galactose at 37 °C. Recombinant LAs were compared with bovine LA as activators of lactose synthesis by conducting lactose synthase assays with increasing concentrations of LA at a fixed concentration of GT and 10 mM glucose and as inhibitors of galactose transfer to *N,N'*-diacetylchitobiose (ChB) where the acceptor concentration was 0.5 mM. The enzyme (GT) concentration was the same for each set of assays, and comparisons between assays conducted at different times with GT preparations of different activity were effected by relating activities for lactose synthesis to the apparent  $V_m$  determined by varying the concentration of bovine LA. For inhibition assays with ChB, the activity obtained in the absence of LA at 0.5 mM ChB was arbitrarily designated as 1.0. Kinetic data were analyzed by fitting to appropriate rate equations using the ENZFITTER computer program (Biosoft).

### Interpretation of Steady State Kinetic Measurements of LA Activity

LA modulates the affinity of GT for acceptor substrates. It acts as a competitive inhibitor with respect to extended substrates such as *N,N'*-diacetylchitobiose and enhances the binding of glucose and other monosaccharides (17, 18, 24). With weaker binding substrates, such as glucose, inhibition effects resulting from effects on  $V_m$  are observed at high concentrations of LA and/or acceptor when dead-end inhibitory complexes are formed (25). We have assumed here that the mutant LAs act in a similar manner to bovine and other wild-type LAs.

Although there have been alternative proposals that the mechanism of lactose synthase is best represented by (a) a partially ordered mechanism in which UDP-galactose binds to a GT·Mn<sup>2+</sup> complex prior to a random equilibrium binding of LA and glucose (17, 18) or (b) a completely random equilibrium mechanism (24), there is agreement that the effect of LA involves a random highly synergistic binding with glucose to GT complexes or, in the case of larger acceptor substrates, exemplified by ChB, mutually exclusive binding of LA and acceptor with GT. Under the conditions used here for enzyme assays, GT is essentially saturated with Mn<sup>2+</sup> at both metal binding sites (Mn<sup>2+</sup> concentration of 10 mM versus  $K_d$  values of 2  $\mu$ M and 2 mM) and with UDP-galactose (concentration of 330  $\mu$ M versus  $K_d$  and  $K_m$  values of 25 and 60  $\mu$ M, respectively (17, 26)) so that, irrespective of which mechanism provides the best model for enzyme system, essentially all of the GT will be distributed in complexes containing both Mn<sup>2+</sup> and UDP-galactose. For lactose synthase assays, low glucose concentrations were also used (10 mM or less as compared with a  $K_m$  of 2 M) so that enzyme complexes containing glucose but not LA are insignificant. The section of the mechanism investigated can be consequently represented by the following steps:





where  $GT^*$  represents a  $GT \cdot (Mn^{2+})_2 \cdot UDP$ -galactose complex.

If rapid equilibrium binding of LA and glucose is assumed, as indicated by previous kinetic studies (17, 18, 24) the rate equation (in the absence of products) becomes:

$$v = \frac{V_m [LA][glc]}{K_i^{LA} K_m^{glc} + K_m^{glc} [LA] + [LA][glc]} \quad (\text{Eq. 4})$$

where  $V_m$  is the maximum velocity ( $k_{cat}[GT]$ ),  $[LA]$  and  $[Glc]$  are the concentrations of LA and glucose,  $K_i^{LA}$  is the dissociation constant for the LA from a  $GT \cdot Mn^{2+} \cdot UDP$ -galactose-LA complex, and  $K_m^{glc}$  is the  $K_m$  for glucose at saturating levels of LA. The latter is equal to the equilibrium dissociation constant of glucose from the  $GT \cdot Mn^{2+} \cdot UDP$ -galactose-LA-glucose complex in a rapid equilibrium mechanism.  $K_i^{LA}$  is the dissociation constant of LA from the same complex that produces competitive inhibition with respect to ChB, and can therefore be elucidated separately as an inhibition constant ( $K_i$ ) for that reaction. The action of LA (and its mutants) in activating the catalysis of lactose synthesis by GT at a fixed concentration of glucose reflects an increase in the proportion of enzyme complexes containing both glucose and LA which can give rise to products (*i.e.* also containing  $Mn^{2+}$  and UDP-galactose), and is consequently dependent on both the affinity of LA for GT and its effects on glucose binding. From the above equation, it can be seen that a double reciprocal plot of lactose synthase activity *versus*  $[LA]$  will have an intercept/slope ratio equal to:

$$\frac{([glc] + 1)}{K_i^{LA}} \quad (\text{Eq. 5})$$

from which,  $K_m^{glc}$  can be determined, using the value for  $K_i^{LA}$  obtained from the inhibition of galactose transfer to ChB. Alternatively, the apparent  $V_m$  for lactose synthesis ( $= V_m(\text{true}) / (1 + K_m/[glucose])$ ) can be used to determine  $K_m$  by comparison with the parameters obtained for a corresponding plot for bovine milk LA under the same conditions.  $K_m^{glc}$  is the  $K_m$  for glucose at saturating levels of LA and for a rapid equilibrium mechanism is equal to the equilibrium dissociation constant of glucose from the  $GT \cdot LA$ -glucose complex. By using LA as a competitive inhibitor with respect to ChB, its affinity for the  $GT \cdot gMn^{2+} \cdot gUDP$ -galactose complex can be measured separately from effects resulting from glucose binding, and used to calculate the  $K_m$  for glucose at saturating  $[LA]$ . For the ChB reaction,  $K_i$  values can be calculated from the apparent  $K_i$  values obtained from the Dixon plots (intercept/slope) using the relationship given below:

$$K_i(\text{true}) = \frac{K_i(\text{apparent})}{1 + \frac{[S]}{K_m}} \quad (\text{Eq. 6})$$

where  $K_m$  is the  $K_m$  for ChB (separately determined to be 0.88 mM under these conditions) and  $[S]$ , the concentration of ChB (0.5 mM).

#### Spectroscopy

Near and far UV CD spectra of bovine LA and mutant LAs were determined with a JASCO J-710/720 spectropolarimeter. Twenty spectra were scanned for each sample at a speed of 100 nm/min which were subsequently averaged and smoothed. Near UV CD spectra (250–320 nm) were determined using a cell with a path length of 1 cm, and far UV spectra (200–250 nm) using a cell with a path length of 0.1 cm. Proteins were dissolved in 0.02 M tris-HCl, pH 7.4, containing 0.1 mM  $CaCl_2$ , 0.01 M HCl (pH 2) or 6 M guanidine hydrochloride at concentrations between 130 and 500  $\mu$ g/ml.

#### Other Methods

The  $NH_2$ -terminal sequence of mLA was determined using an Applied Biosystems model 470A Protein/Peptide Sequencer fitted with a model 120A analyzer and a model 900A data analysis system for phenylthiohydantoin-derivative analysis and sequence assignment. Syn-

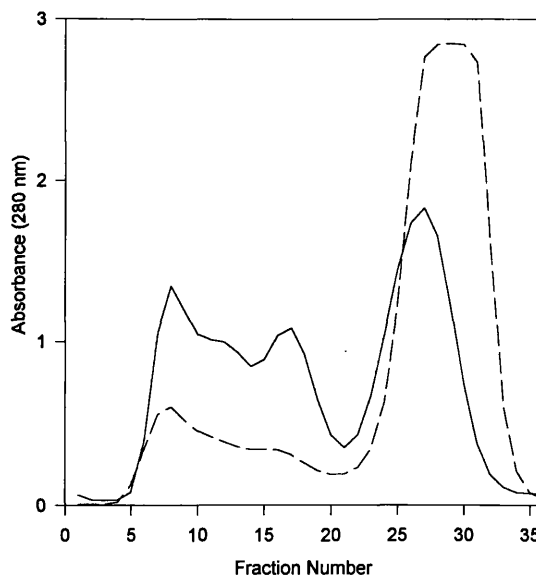


FIG. 1. Purification of selected recombinant LAs by gel filtration with a column of Sephadex G-75. After treatment to allow native folding and disulfide bond formation, the protein was precipitated with 80% ammonium sulfate. The column size and separation conditions are described in the text. 80 ml of eluant was collected before the first fraction. Separation of W118Y-mLA, - - - -; separation of W104Y-mLA, —.

thetic oligonucleotide synthesis was carried out by Dr. Rudolf Werner, Department of Biochemistry and Molecular Biology, University of Miami. Molecular models of mLA and some variants (lacking the  $NH_2$ -terminal methionine) was constructed using the Hyperchem software package (Autodesk Inc., Sausalito, CA) by substituting its sequence into the baboon LA structure (4) and optimizing the geometry by AMBER force field energy minimization to an RMS gradient of  $<0.05$  kcal/( $\text{\AA}$  mol).

## RESULTS

**Preparation and Properties of Recombinant Bovine LA (mLA)**—The system described here was designed to directly express the mature bovine LA sequence in *E. coli*. The substitution of valine for the single methionine at position 90 in the mature LA sequence produced a molecule which is resistant to CNBr cleavage. mLA was expressed, treated and purified as described in "Experimental Procedures." Fig. 1 shows typical elution profiles for the separation of two mLA variants (W118Y-mLA and W104Y-mLA) by gel filtration with Sephadex G-75 after folding. The profile obtained with W118Y-mLA (Fig. 1a) is typical of that obtained with recombinant LAs that folded efficiently, including mLA. The product from this step showed a single band with the same mobility as bovine LA on SDS-gel electrophoresis even at high loadings. The product also ran as a single component on anion exchange HPLC analysis. Final yields of mLA reproducibly exceeded 25 mg/liter of bacterial culture. When mLA was subjected to automated amino-terminal sequencing, a single sequence was obtained with methionine as the amino-terminal residue, followed by the expected amino-terminal sequence of bovine LA.

Because mLA was constructed to be devoid of internal methionine residues, the  $NH_2$ -terminal methionine can be specifically removed by CNBr cleavage without cleaving any internal peptide bonds. However, as described below, mLA is very similar in activity and physical properties to bovine milk LA. Therefore, it was concluded that it is appropriate to use variants of mLA to investigate structure-function relationships in LA without removing the amino-terminal methionyl residues.

**Generation and Purification of Mutants of mLA**—The mutations introduced into mLA are listed in Table I together with

TABLE I  
Primers used for construction of pMLA and for mutagenesis

Primer	Sequence	Orientation
NF-N	CCATGCCACCCATATGGAACAGTTAAC	Coding
NF-C	GCAAAGACAGCGGATCCTCACAACCTCCAC	Complementary
T7 promoter	TAATACGACTCACTATAGGG	Coding
T7 terminator	GCTAGTTATTGCTCAGCGG	Complementary
F31E	CTGTACCACGGAACATACCAGTGG	Coding
F31L	CTGTACCACGTTACATACCAGTGG	Coding
F31S	CTGTACCACGTTTCATACCAGTGG	Coding
F31Y	CTGTACCACGTTATCATACCAGTGG	Coding
H32A	ACCACGTTTGCAACCAGTGG	Coding
H32E	ACCACGTTTGAAACCAGTGG	Coding
H32N	ACCACGTTTAAATACCAGTGG	Coding
H32Y	ACCACGTTTATACCAGTGG	Coding
Y103A	GGCCAACCAGGCGTTAATTC	Complementary
Y103P	GGCCAACCAGGGTTAATTC	Complementary
W104A	GCTTTATGGGCAACGCGTAGTTAATTC	Complementary
W104L	GCTTTATGGGCAACAAGTAGTTAATTC	Complementary
W104Y	GCTTTATGGGCAAGTAGTAGTTAATTC	Complementary
Q117A	GAGAAGCTGGATGCTTGGCTCTG	Coding
Q117D	GAGAAGCTGGATGATGGCTCTG	Coding
W118H	GGATCAGCATCTCTGTGAG	Coding
W118Y	GGATCAGTATCTCTGTGAG	Coding

the oligonucleotides used in generating them. Sites of mutation included all invariant positions in cluster I (Phe-31, His-32, Gln-117, and Trp-118) together with Tyr-103 and Trp-104 of cluster II. Other residues in cluster II were not altered because of variability in different LA sequences (3, 13) or buried nature in both LA and lysozyme (4, 10). The choice of substitutions was based on either the nature of residues found at the corresponding site in various lysozymes (*e.g.* Glu for Phe-31, Asn and Tyr at position 32, Pro and Ala at position 103, Ser and Ala at position 117, and Tyr at position 118), while other substitutions were designed to be structurally conservative or to examine the effects of different side chain types at a particular site. Table II summarizes the yields obtained at different stages during their production. For mLA and variants that folded efficiently *in vitro*, high yields of monomeric protein were obtained on purification by gel filtration (Fig. 1a) and further purification by HPLC anion exchange chromatography gave a single peak that eluted at a similar NaCl concentration to bovine LA and mLA (see Fig. 2a). Variants of mLA that did not fold efficiently *in vitro* in some cases showed a large proportion of aggregated protein that eluted at the void volume on separation by gel filtration. With other mutants, aggregated protein was present together with a peak of monomeric protein which, on separation by HPLC, did not show a peak eluting at the same time as mLA but did show a major peak eluting at high NaCl concentration. Some mutants were analyzed directly by HPLC after folding and were found to be devoid of native protein (H32E-mLA, H32N-mLA, W104A-mLA, and W104L-mLA). Fig. 1b shows the separation of W104Y-mLA by gel filtration after folding in the presence of 10% glycerol; monomeric protein was present which, on separation by HPLC, showed a major component that eluted at a later time and as a broader peak than mLA (Fig 2b); additional material eluted at high NaCl concentrations. The main component was used in further characterization. In the case of H32A-mLA, the final yield of folded protein increased 5-fold when glycerol (10%) was included during the folding process; however, the inclusion of glycerol did not allow the production of native protein from Y103A-mLA. SDS-gel electrophoresis showed that only low levels of protein contaminants were present after the initial separation by anion exchange chromatography in 4 M urea so that the initial levels

TABLE II  
Yields of mLAs and variants during purification

Protein	Yield from O50	Yield from G-75	
		mg/liter of culture	%
mLA	107	40	100
F31E	77.5	8.8 (0) <sup>b</sup>	0 <sup>c</sup>
F31L	62.5	5.5 (0)	0 <sup>c</sup>
F31S	60	16.5 (1.0)	4 <sup>c</sup>
F31Y	32	5.5	46
H32A	55	10.0 (1.0)	5 <sup>c</sup>
H32A <sup>d</sup>	73	28.5 (5.7)	21 <sup>c</sup>
H32E	65.5	ND <sup>e</sup> (0)	0
H32N	55.8	ND (0)	0
H32Y	47	17.5	100
Y103A	26.7	1.7	0
Y103P	53.3	17.7	29
W104A	53.3	ND (0)	0
W104L	42.5	ND (0)	0
W104Y	58.8	11.3 (2)	(9) <sup>f</sup>
Q117A <sup>d</sup>	53.9	42.4	210
Q117D <sup>d</sup>	62.2	29.3	126
W118H <sup>d</sup>	54	15.6	77
W118Y <sup>d</sup>	53	29.8	150

<sup>a</sup> Yield for mLA after final purification was set at 100%. Yields for all other proteins are relative to that of met-LA.

<sup>b</sup> Values in parentheses are yields after ion exchange HPLC.

<sup>c</sup> Yields calculated for protein after HPLC separation.

<sup>d</sup> Proteins were folded in the presence of 10% glycerol.

<sup>e</sup> ND, protein was separated by HPLC directly after treatment to generate native fold.

<sup>f</sup> The major component isolated for W104Y met-LA was not native based on chromatographic and physical properties.

of expression of the different proteins varied only slightly although there were major differences in the final yields of the native protein. Based on minimal yields of folded protein, the following mutations were incompatible with native folding under the conditions used here: F31E, F31L, H32E, H32N, Y103A, W104L, and W104A. Proteins obtained in sufficient yield for detailed characterization had the following substitutions: F31S, F31Y, H32A, H32Y, Y103P, W104Y, Q117D, Q117A, W118Y, and W118H. W104Y-mLA is not included in the list because the major product isolated by HPLC after folding has

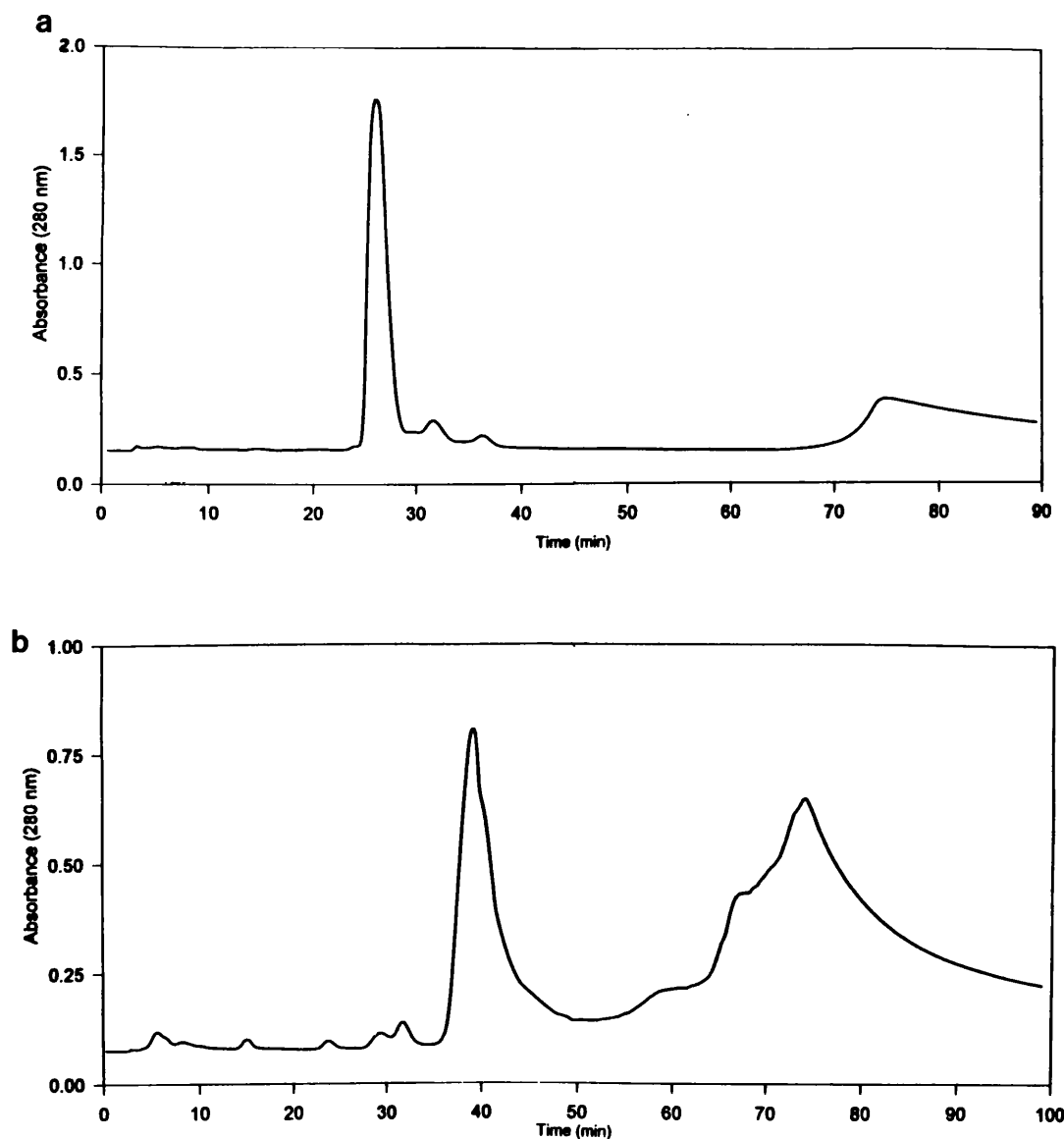


FIG. 2. Separation of representative recombinant LAs by HPLC anion exchange chromatography. *a*, mLA; *b*, W104Y-mLA. The conditions are those described in the text.

properties that suggest it is in a different conformational state from native mLA.

**Structural and Physical Characterization of Mutant LAs**—The purified proteins were all homogeneous on SDS-gel electrophoresis and anion exchange HPLC even at high loadings. The entire sequences of mLA and the mutants were checked by DNA sequencing, confirming the presence of the desired mutation in each case. In two cases, an additional single base change was present resulting in an additional amino acid substitution, presumably being introduced by the PCR mutagenesis procedure. These were in F31E-mLA where methionine was substituted for valine 8, and Q117D-mLA where lysine was substituted for glutamate 113.

The near and far UV CD spectra of mLA, and selected mutant proteins are shown in Fig. 3 in comparison with those of bovine LA at pH 2 (acid or molten globule state), pH 7.4 (native state), and in 6 M guanidine HCl (denatured state). These indicate that most of the mutant proteins which were characterized with respect to functional properties are closely similar in conformation to the native state of bovine LA. W118Y-mLA had a slightly modified spectrum in the near UV range, a feature that can be reasonably attributed to the nature of the sequence

change. There are pronounced differences between the CD spectra of the mutants with changes in the hydrophobic box residues, tyrosine 103 and tryptophan 104. Y103P-mLA, after purification by HPLC had near and far UV spectra that are closely similar to those of bovine LA and mLA, whereas the monomeric form of Y103A-mLA, isolated by gel filtration in low yield, had a molten globule spectrum, similar to that of bovine LA at pH 2 (see Fig. 3, *a* and *c*). The major component isolated from W104Y-mLA by HPLC was quite distinct. It has a pronounced near UV CD spectrum of similar magnitude to mLA, but with major differences (Fig. 3c). The small trough at about 295 nm is missing and the main trough is red-shifted and altered in shape. The far UV CD spectrum also differs in shape from that of mLA.

**Functional Properties of Mutant LAs**—The activities of different mutant LAs and mLA as activators of glucose binding are compared in Fig. 4. mLA and Y103P-mLA are closely similar in activity whereas other mLA variants are less active than the parent protein. Bovine LA is closely similar in activity to mLA (Table III). Visual inspection of the activity profiles suggests that some variants have reduced apparent  $V_m$  values (e.g. F31Y-mLA) while others have increased apparent  $K_m$  values

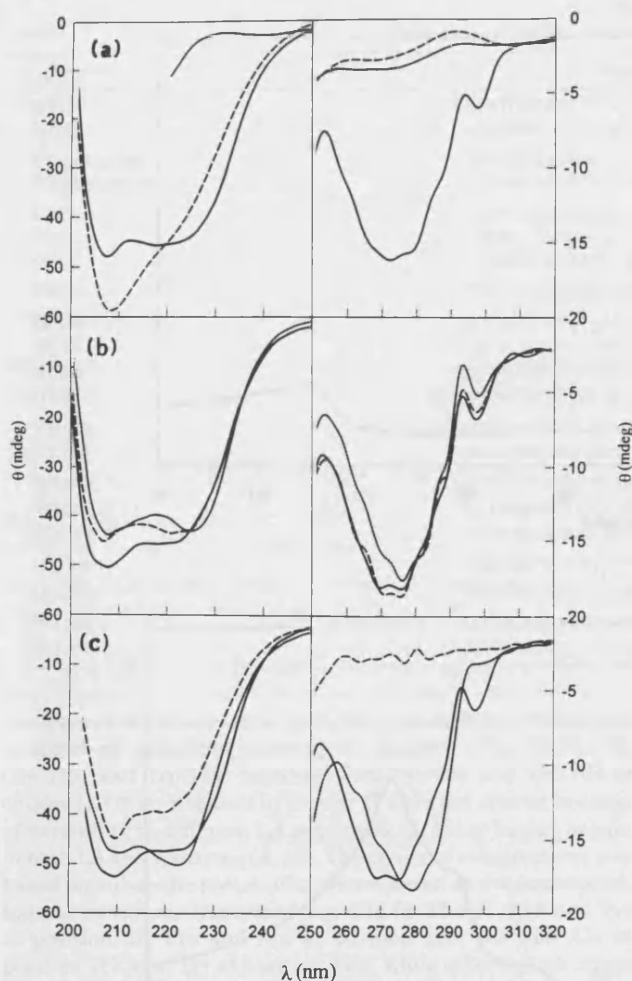


FIG. 3. Near (250–320 nm) and far (200–250 nm) ultraviolet CD spectra of bovine LA and selected mLA variants. *a*, bovine LA at pH 7.0, —; pH 2, - - -; and in 6 M guanidine hydrochloride, ····. *b*, mLA, —; H32A-mLA, - - -; and W118Y-mLA, ····, all at pH 7.0. *c*, mLA, —; Y103A-mLA, - - -; and W104-mLA, ····, all at pH 7.0. All spectra are adjusted to represent protein concentrations of 0.5 mg/ml.

(e.g. W118Y-mLA). Mutants with substitutions for histidine 32 (tyrosine or alanine) showed very little activity over a wide concentration range. W104Y-mLA also had a low activity, but was not isolated in sufficient quantity to perform assays at high concentrations or to perform inhibition assays. The double mutant, Q117D/E113K-mLA displayed an activity profile that is essentially superimposable on that of Q117A-mLA (data not shown).

The different levels of activity observed in the lactose synthase reaction can reflect changes in either or both of two parameters associated with LA function, the affinity for GT ( $K_i^{LA}$ ) and the ability at saturation to promote glucose binding ( $K_m^{glc}$ ). As discussed in "Experimental Procedures," changes in the apparent  $V_m$  in the lactose synthase reaction at a fixed concentration of glucose are expected to reflect changes in the latter parameter. Differences in the affinity for GT were determined using the LAs as inhibitors of the catalysis of galactose transfer to ChB (Fig. 5). Arranged in terms of decreasing effectiveness as inhibitors, the mutant LAs are: Y103P > bovine LA = mLA > F31Y > Q117D > H32A > F31S > W118Y = W118H, the latter two mLA mutants having very low activities in this assay. A comparison of the activities of the various proteins in the two reactions (Figs. 4 and 5) indicates that the relative effectiveness of pairs of mutants, Q117A/F31Y, Q117A/F31S, and

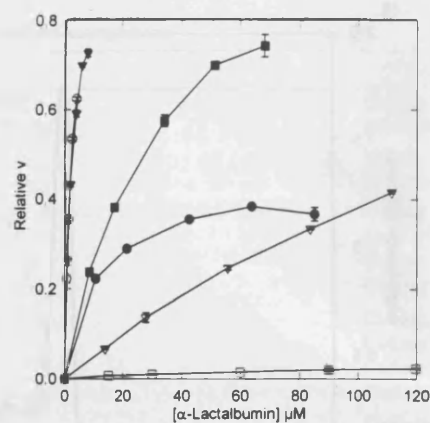


FIG. 4. Effect of increasing concentrations of bovine LA and recombinant mutant LAs on lactose synthase activity of bovine GT. Assays were conducted as described under "Experimental Procedures," using a glucose concentration of 10 mM. The LA variants were: mLA  $\circ$ , Y103P-mLA  $\blacktriangledown$ , Q117A-mLA  $\blacksquare$ , F31Y-mLA  $\bullet$ , W118Y-mLA  $\nabla$ , H32Y-mLA  $\square$ .

TABLE III  
Kinetic parameters determined for mLA and variants with bovine GT<sup>a</sup>

Protein	$K_i^{LA}$ $\mu M$	Change in $K_i^{LA}$	$K_m^{glc}$ mM	Change in $k_m^b$
Bovine LA (Ref. 33)	17	1	0.9	0.9
	26	1	0.8	0.7
mLA	21	1	1.1	1
F31S	875	42	181	165
F31Y	39	2	16	15
H32A	716	34	ND <sup>c</sup>	ND
H32Y	243	12	276	251
Y103P	15	1	2.0	2
Q117A	135	7	3.2	3
Q117D	57	3	2.1	2
W118H	1766	84	ND	ND
W118Y	1488	71	1.7	1.5

<sup>a</sup> Change in  $K_i$  is value for  $K_i^{LA}$  divided by corresponding value for mLA.

<sup>b</sup> Change in  $K_m^{glc}$  divided by corresponding value for mLA.

<sup>c</sup> ND, could not be reliably determined from data.

W118Y/H32A are reversed. The explanation of this is shown in Table III, which gives the values for  $K_i^{LA}$  and  $K_m^{glc}$  for the various proteins. The deficiencies in Q117A-mLA and W118Y-mLA specifically reflect increased values for  $K_i^{LA}$ , whereas F31Y-mLA has a slightly decreased affinity for GT (<2-fold) combined with a much lower ability to promote glucose binding (15-fold increase in  $K_m^{glc}$ ). The latter deficiency is displayed as a reduced lactose synthase activity at high LA concentrations. A similar change is found in other mutants with substitutions for phenylalanine 31 and histidine 32, but combined with major reductions in affinity for GT. Although the low activity levels obtained with F31S-mLA, H32A-mLA, and H32Y-mLA introduce uncertainty regarding the precise values of their kinetic parameters, it appears that the position 32 mutants are even more deficient than the position 31 mutants in their effects on glucose binding. Y103P-mLA has a slightly higher affinity for GT than mLA and bovine LA (Fig. 5 and Table III) in conjunction with a slightly decreased ability to promote glucose binding. However its overall activity in the lactose synthase reaction is closely similar to that of mLA (Fig. 4).

#### DISCUSSION

In this study we describe an expression system for bovine LA that is more expedient than fusion protein systems previously

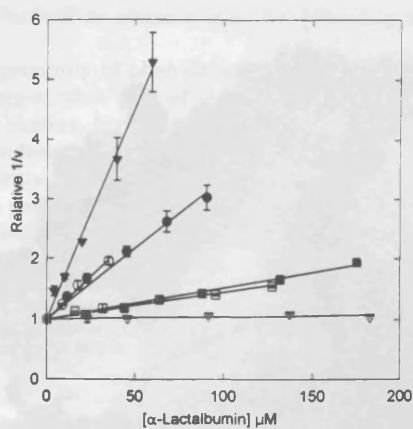


FIG. 5. Inhibition of activity of bovine GT for galactose transfer to  $N,N'$ -diacetylchitobiose by bovine LA and recombinant LAs. The symbols are the same as in Fig. 4.

developed in this laboratory (9, 27). Two sequence changes introduced into the parent recombinant protein (mLA), the presence of an additional methionine at the  $\text{NH}_2$  terminus and the substitution of valine for methionine 90, have insignificant effects on its functional properties. Also, as shown in Fig. 3, mLA and variants used in studies of structure-function relationships have near- and far-UV CD spectra that are closely similar to those of bovine LA, and other studies<sup>2</sup> indicate that the molecular stability of mLA is closely similar to that of bovine milk LA. These properties and the relative ease with which high yields of pure native protein can be generated indicate that the mLA expression system is suitable for investigating the structural basis of activity and stability in LA. Construction of mutants using the PCR "megaprimer" method resulted in the introduction of unintended mutations in two cases, a relatively low error rate when balanced against the facility of the procedure. This system has significant advantages over previously described expression systems for LA (9, 15, 16). Thus, a yeast expression system for a secreted form of LA gave yields of only 2 mg/liter of culture (15), and a bacterial expression system for goat LA as a fusion protein provided a product, after cleavage, with only 12% of the activity of goat LA (16). This system was used to express lysozyme/LA chimeras, but the results are difficult to interpret with regard to structure-function relationships in LA since the physical properties of the recombinant proteins were not examined.

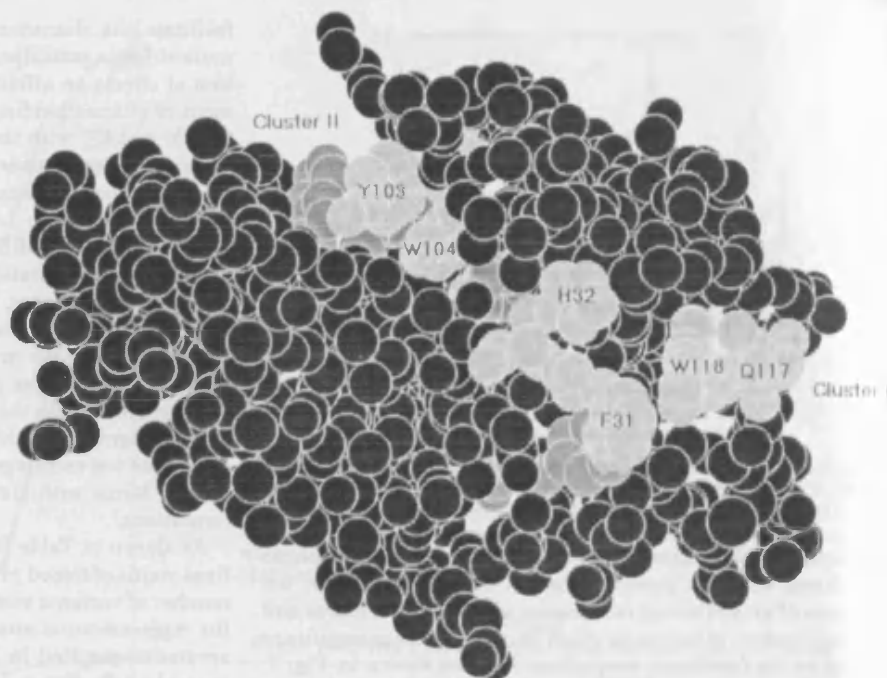
The aim of this study was to use site-directed mutagenesis to probe the roles of two substructures of LA, aromatic clusters I and II, in its action in lactose synthase. Both structures adjoin the cleft in LA and have been suggested as possible binding sites for GT (13, 14). The highly conserved nature of cluster I and previous chemical studies support the view that this structure is important in the activity of LA. Trace-labeling studies indicate that lysines 5 and 114, which flank cluster I, proximate the GT binding site (11, 12) while chemical modification of histidine 32 or tryptophan 118 has been found to disrupt the activity of LA (28–30). Cluster II, or the hydrophobic box, is less conserved but is a site of structural difference between LA and lysozyme, resulting from the blockade of the LA cleft by the side chain of tyrosine 103 (alanine in most lysozymes). GT binding would be expected to be associated with such a region where LA and lysozyme differ in structure since lysozyme does not bind GT. Although nitration of tyrosine 103 does not disrupt the activity of LA (31), the possibility that cluster II plays a functional role in LA could not be excluded by previous data. To

facilitate the characterization of the functional properties of mutant LAs a procedure was developed that allows the distinction of effects on affinity for GT from effects on the enhancement of glucose binding. The validity of the values elucidated for  $K_i^{\text{LA}}$  and  $K_m^{\text{glc}}$  with the various mutants are supported by the close agreement between the values obtained for bovine LA and mLA with those previously obtained from more detailed kinetic studies with bovine LA (see Table III). To ensure that any observed functional differences reflect local structural changes, only protein preparations that appear native based on chromatographic behavior, homogeneity on HPLC anion exchange chromatography and near and far UV CD spectra, were considered suitable for probing structure-function relationships. Chromatography was performed with columns containing an organic matrix with that is derivatized with polyethyleneimine. Chromatography on this support therefore involves adsorption as well as ion exchange and, from our observations, separates protein forms with similar ionic properties but different conformations.

As shown in Table II, despite similar expression yields, the final yields of folded protein varied greatly. Although a limited number of variants were generated, it appears that, in general, the replacement of aromatic residues in the clusters with non-aromatics resulted in low or insignificant folding yields presumably reflecting a destabilization of the native fold. In one case (H32A-mLA) glycerol was found to improve the yield (Table II) as previously reported for marginally stable mutants of other proteins (*e.g.* see Ref. 32). The properties of the mutant proteins confirm and extend previous chemical studies of LA, the distinct effects of substitutions in clusters I and II serving to clarify the respective roles of these structures. In the case of cluster II one of five mutants, namely Y103P-mLA, was isolated in native form in good yield. Monomeric protein obtained from Y103A-mLA did not chromatograph like native LA on HPLC and showed a CD spectrum characteristic of the molten globule state of LA. Two of the three mutants with substitutions for tryptophan 104 (alanine and leucine) also did not fold. However, some monomeric protein was isolated from W104Y-mLA by gel filtration (Fig. 1b) which on further separation by HPLC gave a major peak that eluted later than the native protein. Because of the low yield, this protein was not characterized as fully as other mutants. However, it was found to have a significant level of activity in the lactose synthase reaction and to have a CD spectrum which is distinct from that of mLA but indicates the presence of secondary and tertiary structure (Fig. 3c). Tryptophan 104 is conserved in all known LA and lysozyme sequences. In LA it is buried (4, 10) and is therefore unlikely to play a direct role in interacting with GT or glucose. Its structural importance is supported by a previous study with human LA in which the nitration of tryptophan 104 abolished activity and altered the far UV CD spectrum (31). The changed CD spectrum of W104Y-mLA can be, in part, attributed to the nature of the sequence change. It is more pronounced than the change seen in W118Y-mLA (Fig 3, a and c) which could reflect the different environments of these two tryptophans in the native structure or a conformational change in W104Y-mLA. Energy minimization studies suggest that the substitution of tyrosine in position 104 may result in a rearrangement of the hydrophobic box and a more open cleft,<sup>2</sup> which may serve to explain the increased retention time on HPLC. Studies are in progress aimed at characterizing W104Y-mLA in detail but, based on present evidence, it appears that the reduced activity of this mutant results from a localized conformational change. The ability of the Y103P mutant to produce fully active native protein, in contrast with Y103A-mLA, is interesting since alanine and proline are found at this site in various lysozymes. This may reflect the stabilization of protein native structures

<sup>2</sup> J. A. Grobler, W. K. Linnerooth, and K. Brew, unpublished observations.

FIG. 6. A representation of bovine LA showing the locations of aromatic clusters I and II. The backbone of LA is shown in black and the side chains of residues composing the clusters in gray. Residues that were subjected to mutagenesis in the present study are shown in light gray and are labeled.



that results from the introduction of proline at a compatible site which has been attributed to a resulting decrease in entropy, and concomitant increase in free energy, of the unfolded state (34). Overall, these observations indicate that the hydrophobic box (aromatic cluster II) has a structural rather than functional role in LA (Fig. 6).

Mutant LAs with substitutions for components of cluster I are distinct in properties as a class from those involving components of cluster II. A larger proportion of them (8 out of 12) allow the generation of native protein, and all of these show significant changes in functional properties. Therefore, this region, which is more exposed to solvent than cluster II, is of functional importance in LA. In agreement with previous results, it is shown that histidine 32 and tryptophan 118 play particularly important roles in LA action, all substitutions made at these sites producing major perturbations of activity. Two conserved residues adjacent to these, phenylalanine 31 and glutamine 117, which had not been previously probed by chemical modification, are also implicated in the activity of LA. The kinetic properties of Q117A, W118Y and W118H-mLAs are consistent with direct or indirect roles for residues 117 and 118 in the interaction with GT. Although the activity of W118H-mLA was so low that its effects on glucose binding were not determined, W118Y-mLA appears to be specifically reduced in affinity for GT (Table III). In contrast, the position 31 and 32 mutants show changes in both kinetic parameters. The most informative protein in this group is F31Y-mLA whose low lactose synthase activity is almost entirely attributable to a reduced (15-fold) ability to promote glucose binding. With the position 32 mutants the effect on glucose binding is decreased by up to two orders of magnitude. As far as we are aware, significant effects on this aspect of LA action resulting from a structural modification have not been noted previously. Therefore, the effects of changes in the residues composing cluster I divide them into two groups: residues 117 and 118 which specifically influence the strength of binding with GT, and residues 31 and 32 which influence both GT binding and the enhancement of glucose binding or in the case of F31Y-mLA, affect glucose binding with minimal effects on GT binding.

These results are relevant to the mechanism of action of LA. In one model previously proposed for its mechanism of action

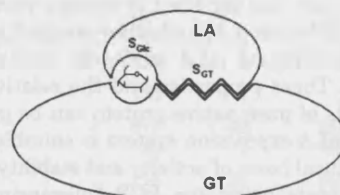


FIG. 7. The "monosaccharide bridge" model for the LA-GT interaction.  $S_{GT}$  denotes the interaction site on LA for GT, and  $S_{Glc}$ , the proposed site through which LA interacts with a monosaccharide in the lactose synthase complex.

(35), LA binds to GT at a site adjacent to the acceptor binding site and provides additional stabilizing interactions for a monosaccharide located at the interface of the two protein components (a "monosaccharide bridge" arrangement as shown schematically in Fig. 7). The same mode of binding sterically prevents the binding of more extended acceptor substrates. The model implies that LA has a binding site for GT ( $S_{GT}$ ) and an adjacent site which can interact favorably with a bound monosaccharide in the lactose synthase complex ( $S_{Glc}$ ). Because the cleft in LA contains several conserved residues whose counterparts in lysozyme act in carbohydrate binding, a region within the shortened cleft in LA may support glucose binding in the lactose synthase complex (12). In this model, the interaction of LA with GT and the enhancement of glucose binding to this complex are associated with distinct but neighboring regions of the LA molecule and would be separately affected by mutagenesis. An alternative mechanism would be for LA to modulate the substrate specificity of GT through an allosteric effect. Although reasonable, this hypothesis implies an unusually extreme level of functional divergence at the molecular level between LA and lysozyme and is also difficult to reconcile with the rapid equilibrium association of LA and GT (17, 24). The properties of the mutant LAs are more consistent with the former model for LA action since single-site substitutions are found to separately change the two aspects of LA activity. If the action of LA on GT was allosteric the two facets of LA action would be intrinsically linked. Therefore, these results tentatively support the "monosaccharide bridge" model although more information regarding the structure of GT and the mo-



lecular basis of its action is required to reach a reliable conclusion.

The proximity of phenylalanine 31 and histidine 32 to the cleft suggest that part of this region may act in supporting glucose binding. However, a contact site with GT that includes only residues 117 and 118 would be clearly insufficient to account for the strength and specificity of the interaction between the two proteins. Studies are currently in progress to attempt to further define the roles of other conserved residues in LA in its action in lactose synthase.

**Acknowledgments**—We thank Vera Ondricek for expert assistance in amino acid sequence analysis, Adriana Vasquez for technical help, and Dr. Walter A. Scott for helpful discussions.

#### REFERENCES

- Hill, R. L., and Brew, K. (1975) *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 411–490
- Kronman, M. J. (1989) *CRC Crit. Rev. Biochem. Mol. Biol.* **24**, 565–667
- Brew, K., and Grobler, J. A. (1992) in *Advanced Dairy Chemistry-1* (Fox, P., ed) pp. 191–229, Elsevier, London
- Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M., and Phillips, D. C. (1989) *J. Mol. Biol.* **208**, 99–127
- Hiroaka, Y., Segawa, T., Kuwajima, K., Sugai, S., and Murai, N. (1980) *Biochem. Biophys. Res. Commun.* **93**, 1098–1104
- Rao, K. R., and Brew, K. (1989) *Biochem. Biophys. Res. Commun.* **163**, 1390–1396
- Kuwajima, K. (1989) *Proteins Struct. Funct. Genet.* **6**, 87–103
- Ewbank, J. J., and Creighton, T. E. (1993) *Biochemistry* **32**, 3677–3693
- Wang, M., Scott, W. A., Rao, K. R., Udey, J., Conner, G. E., and Brew, K. (1989) *J. Biol. Chem.* **264**, 21116–21121
- Alexandrescu, A. T., Broadhurst, R. W., Wormald, C., Chyan, C.-L., Baum, J., and Dobson, C. M. (1992) *Eur. J. Biochem.* **210**, 699–709
- Richardson, R. H., and Brew, K. (1980) *J. Biol. Chem.* **255**, 3377–3385
- Sinha, S. K., and Brew, K. (1981) *J. Biol. Chem.* **256**, 4193–4204
- Shewale, J. G., Sinha, S. K., and Brew, K. (1984) *J. Biol. Chem.* **259**, 4947–4956
- Koga, K., and Berliner, L. J. (1985) *Biochemistry* **24**, 7257–7262
- Viaene, A., Volckaert, G., Joniau, M., De Baestelieer, A., and Van Caewelaert, F. (1991) *Eur. J. Biochem.* **202**, 471–477
- Kumagi, I., Takeda, S., and Miura, K. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5887–5891
- Khatra, B. S., Herries, D. G., and Brew, K. (1974) *Eur. J. Biochem.* **44**, 537–560
- Powell, J. T., and Brew, K. (1976) *J. Biol. Chem.* **251**, 3653–3663
- Kunkel, T. A. (1985) *Proc. Natl. Acad. U. S. A.* **82**, 488–492
- Sarkar, G., and Sommer, S. S. (1990) *BioTechniques* **8**, 404–407
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, pp 17.38–17.39, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Gill, S. C., and von Hippel, E. H. (1989) *Anal. Biochem.* **182**, 319–326
- Bell, J. E., Beyer, T. A., and Hill, R. L. (1976) *J. Biol. Chem.* **251**, 3003–3013
- Morrison, J. F., and Ebner, K. E. (1971) *J. Biol. Chem.* **246**, 3985–3991
- Powell, J. T., and Brew, K. (1976) *J. Biol. Chem.* **251**, 3645–3652
- Brew, K., Grobler, J. A., Wang, M., Scott, W. A., Yadav, S. P., and Malinovsky, V. (1993) in *Advances in Gene Technology: Protein Engineering and Beyond* (Miami Short Reports 3), (Brew, K., Petsko, G. A., Ahmad, F., Bialy, H., Black, S., Fernandez, A., Fenna, R. E., Lee, E. Y. C., and Whelan, W. J., eds) p. 98, IRL Press, Oxford
- Prieels, J.-P., Bell, J. E., Schindler, M., Castellino, F. J., and Hill, R. L. (1979) *Biochemistry* **18**, 1771–1776
- Schechter, Y., Patchornik, A., and Burstein, Y. (1974) *J. Biol. Chem.* **249**, 413–419
- Bell, J. E., Castellino, F. J., Trayer, I. P., and Hill, R. L. (1975) *J. Biol. Chem.* **250**, 7579–7585
- Prieels, J. P., Dolmans, N., Leonis, J., and Brew, K. (1975) *Eur. J. Biochem.* **60**, 533–539
- Sawano, H., Koumoto, Y., Sasaki, Y., Segawa, S., and Tachibana, H. (1992) *FEBS Lett.* **303**, 11–14
- Powell, J. T., and Brew, K. (1974) *Eur. J. Biochem.* **48**, 217–228
- Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6663–6667
- Brew, K., Richardson, R. H., and Sinha, S. K. (1979) in *From Gene to Protein: Information Transfer in Normal and Abnormal Cells* (Miami Winter Symposium 16) (Russell, T. R., Brew, K., Faber, H., and Schultz, J., eds) pp. 443–447, Academic Press, New York