

THE STRUCTURE OF  $\lambda$  LYSOZYME-CHITOHexasaccharide COMPLEX

STRUCTURAL STUDIES OF THE BACTERIOPHAGE LAMBDA LYSOZYME  
COMPLEXED WITH A CHITOHexasaccharide

By

ADELAINÉ KWUN-WAI LEUNG, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

© Copyright by Adelaine Kwun-Wai Leung

Master of Science (1998)  
(Biochemistry)

McMaster University  
Hamilton, Ontario

TITLE: Structural Studies of the Bacteriophage Lambda Lysozyme Complexed With a  
Chitohexasaccharide

AUTHOR: Adelaine Kwun-Wai Leung, B.Sc. (McMaster University)

SUPERVISOR: Professor A.M. Berghuis

NUMBER OF PAGES: *xiii*, 84

## Abstract

Lysozyme from the bacteriophage lambda, laL, complexed with a chitohexasaccharide has been solved to 2.6 Å by molecular replacement using a mutant form of laL as a model. The protein packs as a dimer in the crystal with the backbones of both monomers being nearly identical. The inhibitor molecule resides in the deep cleft in the middle of the bilobal enzyme. Subsites A to D are occupied by one (GlcNAc)<sub>6</sub> molecule, while the remaining sites interact with two rings from the adjacent (GlcNAc)<sub>6</sub> molecule. The binding mode of laL is compared to other lysozymes (HEWL, HuL, GEWL, T4) and Slf70. Interesting differences are noted in the stacking interactions in ring D and the extensive interactions in ring E. It is hypothesized in the thesis that one possible role of the peptide moiety is to interact with Tyr-132, preventing it from forming stacking interactions with ring D, allowing the sugar to penetrate deeper into the active site. Ring E is buried deeply in the enzyme and has a low thermal factor. In addition, the active site is much narrower in laL than in other lysozyme structures. A possible explanation has been suggested that rings E and F stay in the active site longer than those in lysozyme to prevent water molecules instead of the O6 atom of ring D to participate in the nucleophilic attack at the end of the reaction.

## Acknowledgements

I would like to thank my supervisor Dr. Albert Berghuis for his excellent guidance throughout my studies. I especially want to thank him for his continual support at times when progress seemed non-existent. I would never have accomplished as much work without Dr. Berghuis assisting me through each difficult step in my research. I am extremely appreciative of his effort in increasing my self-confidence in the pursuit of science and giving me advice regarding my future aspirations.

To my co-supervisor, Dr. John Honek, I would like to thank him for his endless support in the project. His enthusiasm has definitely encouraged and inspired me. I am extremely grateful for his participation in the synchrotron trip. I would also like to thank Dr. Daniel Yang for his various helpful discussions in science but particularly for his kind sharing of valuable crystallography knowledge with me. I would like to extend my thanks to Dr. Gerry Wright for his critical evaluation of my thesis. I am thankful that he has pointed out my weaknesses so that I can improve in those areas in order to become a more versatile scientist.

I would like to acknowledge Dr. Wai-Ching Hon for showing me how to grow and mount crystals and how to use some computer software when I was still an undergraduate. I would like to thank her for sharing her experiences with me.

I do not think that I have expressed enough gratitude in words for Byron DeLaBarre and Leanne Wybenga's assistance and support during the CHESS trip when my experiments did not succeed. I would like to take this opportunity to thank them once again for their selfless help, which definitely had reduced my frustration level at the time. I would also like to express my gratefulness to Leanne for proofreading this thesis. I would like to extend my thanks to the rest of my labmates, Desiree Fong, Laura Rossi, and Jeff Schwartzenhauer for their friendship, support, and being tolerant of my messy desk!

I would like to thank all my friends who had helped to reduce my stress level in the past two years. I thank Roy Satmaka for taking the risk to rollerblade with me. This thesis would never have been finished without his late night coffee in the lab! I thank Oscar Cheng and Quyen Hoang for keeping me aware of life outside the lab. Further thanks to Michael Wong and Keith Thai for allowing me to sleep over at their house, and for all the food and drinks they had provided all the time! I also like to thank Michael for lending me his laptop. Without it, this thesis would never have been finished as early.

I am very grateful that my parents have given me the freedom to pursue my goal. I am thankful for their unconditional support, understanding, and patience.

My final gratitude goes to my boyfriend, Billy Quan. Thank you for everything!

# Table of Contents

Abstract .....	iii
Acknowledgements .....	iv
Table of Contents .....	vi
List of figures .....	ix
List of tables .....	xi
List of abbreviations .....	xii
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 THE ATYPICAL LYSOZYME – LAMBDA LYSOZYME (LAL) .....	1
1.2 LYSOZYME STRUCTURE AND MECHANISM.....	3
1.2.1 <i>The Phillips Model</i> .....	5
1.3 LYSOZYME-SACCHARIDE COMPLEXES .....	7
1.4 <i>E. COLI</i> LYTIC TRANSGLYCOSYLASE STRUCTURE AND MECHANISM .....	14
1.5 THESIS OBJECTIVE .....	17
<b>CHAPTER 2 GENERAL EXPERIMENTAL METHODS .....</b>	<b>18</b>
2.1 CRYSTALLIZATION .....	18
2.2 CRYOCRYSTALLOGRAPHY .....	19
2.3 DATA COLLECTION .....	19
2.3.1 <i>Wild-type Lambda Lysozyme</i> .....	19
2.3.2 <i>Lambda Lysozyme and Inhibitor Complex</i> .....	19
2.4 DATA PROCESSING.....	19
2.4.1 <i>Autoindexing</i> .....	20
2.4.2 <i>Refinement of Crystal and Detector Parameters</i> .....	20

2.4.3	<i>Integration of the Diffraction Maxima</i> .....	21
2.4.4	<i>Scaling, Merging and Global Refinement</i> .....	21
2.5	STRUCTURE SOLUTION .....	22
2.5.1	<i>Molecular Replacement</i> .....	23
2.6	STRUCTURE REFINEMENT.....	27
2.6.1	<i>Rigid Body Refinement</i> .....	27
2.6.2	<i>Positional Refinement</i> .....	28
2.6.3	<i>Thermal Factor Refinement</i> .....	28
2.6.4	<i>Bulk Solvent Correction</i> .....	29
2.6.5	<i>R factor vs. Free R</i> .....	30
2.6.6	<i>Graphical Manipulations</i> .....	31
2.6.7	<i>Bootstrapping Refinement Scheme</i> .....	31
<b>CHAPTER 3 WILD-TYPE LAMBDA LYSOZYME .....</b>		<b>33</b>
3.1	EXPERIMENTAL PROCEDURES .....	33
3.1.1	<i>Crystallization</i> .....	33
3.2	RESULTS AND DISCUSSION .....	33
<b>CHAPTER 4 LAMBDA LYSOZYME AND CHITOHXAOSE (GLCNAC)<sub>6</sub> COMPLEX, LAL·(GLCNAC)<sub>6</sub>.....</b>		<b>35</b>
4.1	EXPERIMENTAL PROCEDURES .....	35
4.1.1	<i>Crystallization</i> .....	35
4.1.2	<i>Data Collection and Processing</i> .....	36
4.1.3	<i>Molecular Replacement</i> .....	37
4.1.4	<i>Crystallographic Refinement</i> .....	38
4.2	RESULTS AND DISCUSSIONS .....	42
4.2.1	<i>Overall Structure</i> .....	42
4.2.2	<i>Protein-Protein Interactions and Crystal Packing</i> .....	46

4.2.3	<i>Overall Conformational Changes Upon Substrate Binding</i> .....	49
4.2.4	<i>Structure of Chitohexaose Inhibitor</i> .....	53
4.2.5	<i>Changes in Solvent Structure in the Active Site Upon Inhibitor Binding</i> .....	56
4.2.6	<i>Interactions of Sulfate Anion to laL</i> .....	59
4.2.7	<i>Sugar-Binding Mode in laL Compared to Other Lysozyme/Ligand and Slr70/Bulcegin Complexes</i> .....	60
4.2.8	<i>Similarities and Differences Between laL and Other Homologous Proteins</i> .....	71
4.2.9	<i>The Role of the Peptide in laL Catalysis</i> .....	75
4.2.10	<i>The Implication of Ring D</i> .....	76
<b>CHAPTER 5 SUMMARY AND FUTURE WORK</b> .....		<b>78</b>
5.1	SUMMARY .....	78
5.2	FUTURE WORK .....	80
5.3	CONCLUDING REMARK.....	80
<b>REFERENCE</b> .....		<b>81</b>

## List of Figures

FIGURE 1.1 STRUCTURE OF PEPTIDOGLYCAN. ....	2
FIGURE 1.2 REACTION PATHWAYS FOR LAL AND OTHER LYSOZYMES.....	3
FIGURE 1.3 THE PHILLIPS MODEL.....	4
FIGURE 1.4 SURFACE DEPICTION OF HEWL/TETRASACCHARIDE COMPLEX (PDB:1LZC).....	9
FIGURE 1.5 SURFACE DEPICTION OF HUL/HEXASACCHARIDE COMPLEX ((PDB:1LZS).....	10
FIGURE 1.6 SURFACE DEPICTION OF GEWL/(GLCNAC) <sub>3</sub> COMPLEX (PDB:154L).....	12
FIGURE 1.7 SURFACE DEPICTION OF T4 MUTANT COMPLEXED WITH A CELL WALL SUBSTRATE (PDB:148L).....	13
FIGURE 1.8 SCHEMATIC DIAGRAM OF SLT70.....	14
FIGURE 1.9 PROPOSED MECHANISM OF SLT70.....	15
FIGURE 1.10 STRUCTURE OF SLT70 COMPLEXED WITH BULGECIN .....	16
FIGURE 2.1 DIALYSIS BUTTONS .....	18
FIGURE 2.2 STRUCTURE OF AZA-TRYPTOPHANS. ....	23
FIGURE 2.3 EULERIAN ANGLES.....	24
FIGURE 2.4 COMPARISON BETWEEN THE TWO ROTATION METHODS .....	25
FIGURE 2.5 BOOTSTRAPPING REFINEMENT PROCEDURE. ....	32
FIGURE 3.1 LOW POWER LIGHT MICROSCOPE PHOTOGRAPH. ....	33
FIGURE 4.1 CO-CRYSTALS OF LAMBDA LYSOZYME AND HEXAOSE INHIBITOR, (GLCNAC) <sub>6</sub> . ....	35
FIGURE 4.2 DIFFRACTION SPOTS AND INTEGRATION BOXES. ....	36
FIGURE 4.3 RIGID BODY REFINEMENT SCHEME. ....	38
FIGURE 4.4 RAMACHANDRAN PLOT. ....	39
FIGURE 4.5 PLOT OF R-FACTORS IN 20 RESOLUTION BIN IN THE FINAL LAL·(GLCNAC) <sub>6</sub> STRUCTURE AND THE CORRESPONDING PLOT OF LUZZATI ERROR ESTIMATE. ....	41
FIGURE 4.6 THE MONOMERIC CRYSTAL STRUCTURE OF LAL·(GLCNAC) <sub>6</sub> .....	42

FIGURE 4.7 DIFFERENCE DISTANCE MATRIX PLOT (DDMP) OF MONOMER A AND B.....	43
FIGURE 4.8 TOPOLOGY DIAGRAM FOR LAL.....	43
FIGURE 4.9 THE UNIT CELL PACKING OF LAL·(GLCNAC) <sub>6</sub> .....	44
FIGURE 4.10 SIDE VIEW OF THE ACTIVE SITE.....	45
FIGURE 4.11 AVERAGE THERMAL FACTOR IN MAIN CHAIN ATOMS IN EACH RESIDUE.....	46
FIGURE 4.12 PROTEIN-PROTEIN INTERACTIONS BETWEEN MONOMER A AND B.....	47
FIGURE 4.13 PACKING INTERACTIONS BETWEEN MONOMER A AND SYMB.....	48
FIGURE 4.14 DIFFERENCE DISTANCE MATRIX PLOT (DDMP) OF MLAL-O AND MONOMER B OF LAL·(GLCNAC) <sub>6</sub> .....	50
FIGURE 4.15 DIFFERENCE DISTANCE MATRIX PLOT (DDMP) OF MLAL-C AND MONOMER B OF LAL·(GLCNAC) <sub>6</sub> .....	51
FIGURE 4.16 STRUCTURE AND DENSITY OF CHITOHXAOSE INHIBITOR.....	53
FIGURE 4.17 SURFACE DEPICTION OF MONOMER B OF LAL WITH TWO HEXASACCHARIDE BOUND.....	54
FIGURE 4.18 SUPERIMPOSED ACTIVE SITE OF MLAL AND LAL·(GLCNAC) <sub>6</sub> .....	56
FIGURE 4.19 SUPERPOSITION OF MLAL WITH MONOMER B SHOWING WATER DISPLACEMENT UPON SACCHARIDE BINDING.....	57
FIGURE 4.20 (A),(B) STACKING INTERACTIONS IN RINGS B TO D.....	62
FIGURE 4.21 SUPERPOSITION OF MLAL AND LAL·(GLCNAC) <sub>6</sub> SHOWING TRP MOVEMENT UPON SACCHARIDE BINDING.....	63
FIGURE 4.22 CHARACTERISTIC HYDROGEN BONDING NETWORK IN SITE C.....	65
FIGURE 4.23 SUPERPOSITION OF MLAL AND LAL·(GLCNAC) <sub>6</sub> SHOWING TYR-132 MOVEMENT UPON SACCHARIDE BINDING.....	67
FIGURE 4.24 PRODUCTIVE VS. NON-PRODUCTIVE LYSOZYME-CHITOSACCHARIDE COMPLEXES.....	68
FIGURE 4.25 GRAPHICAL REPRESENTATION OF PROPOSED BINDING MODE OF (GLCNAC) <sub>6</sub> IN LAL.....	71
FIGURE 4.26 SUPERPOSITION OF LAL WITH HEWL, HuL, SLT70, T4L AND GEWL.....	72
FIGURE 4.27 LOCATION OF PROPOSED ASP-52 EQUIVALENT IN LAL, ASP-34.....	74
FIGURE 4.28 HYDROGEN BONDING INTERACTIONS IN ASP-34.....	75

## List of Abbreviations

B factor	temperature factor
CHESS	Cornell high energy synchrotron source
DDMP	difference distance matrix plot
E	normalized structure factor
f	structure factor
F <sub>calc</sub>	calculated structure factor
F <sub>obs</sub>	observed structure factor
GEWL	goose egg white lysozyme
(GlcNAc)-NAL	tetrasaccharide lactone
HEWL	hen egg white lysozyme
Hul	human lysozyme
I	intensity
I <sub>corr</sub>	scaled intensity
IPA	isopropanol
LaL	lambda lysozyme
laL-(GlcNAc) <sub>6</sub>	lambda lysozyme and chitohexasaccharide complex
MGM	N-acetylmuramic acid/N-acetylglucosamine/ N-acetylmuramic acid
mlaL	mutant lambda lysozyme
mlaL-c	closed form of the mutant lambda lysozyme
mlaL-o	open form of the mutant lambda lysozyme
MME	monomethyl ether
NAM/MurNAc	N-acetylmuramic acid
NAG/GlcNAc	N-acetylglucosamine
NaOAc	sodium acetate

## List of Tables

TABLE 4.1 DATA COLLECTION STATISTICS FOR LAL·(GLCNAC) <sub>6</sub> COMPLEX.....	37
TABLE 4.2 STATISTICS OF THE FINAL REFINED STRUCTURE OF LAL·(GLCNAC) <sub>6</sub> AT 2.6 Å RESOLUTION.....	40
TABLE 4.3 SUMMARY OF SUBSITE INTERACTIONS.....	55
TABLE 4.4 SUMMARY OF SULFATE ANION INTERACTIONS.....	59

NSLS	National synchrotron light source
PC	patterson correlation
PDB	Brookhaven protein data bank
PEG	polyethylene glycol
PNP	p-nitrophenyl
r.m.s.	root mean squared
Sl <sub>t</sub>	soluble lytic transglycosylase
SO <sub>4</sub> <sup>-</sup>	sulfate anion
T4L	T4 lysozyme
wlaL	wild-type lysozyme

## Chapter 1 Introduction

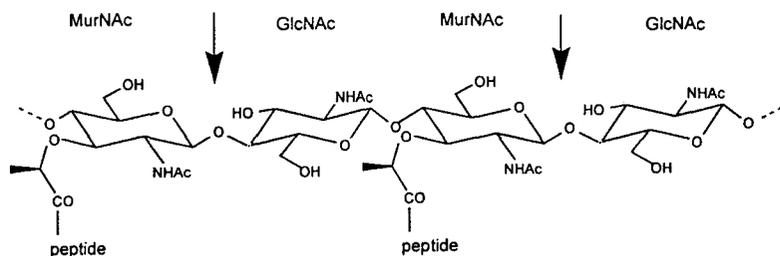
Bacteriophage is a class of viruses that multiply inside bacterial host cells. The particular strain lambda phage reproduces in *Escherichia coli*. The life cycle of  $\lambda$  begins when the tail of the viral particle is adsorbed to the host cell via interactions between the tip of the tail and a component of the outer bacterial cell membrane, followed by the injection of phage DNA into the cytoplasm (Arber, 1983). Two processes can occur after DNA penetration. First, the phage can become a lysogen, in which the phage integrates its DNA into the bacterial chromosome and it is maintained in a dormant state. Secondly, it can enter the lytic pathway where progeny virus replicate inside the host cell. The latter process will eventually lead to lysis of the host cell (Ho & Rosenberg, 1988). The decision of whether the phage will enter the lytic or lysogenic pathway is highly regulated by transcription activators which control the balance of the synthesis of lytic gene repressors and integration enzymes and the expression of the lytic genes (Belfort & Wulff, 1973).

The focus of this thesis is on an enzyme involved in the lytic pathway. One critical process in the lysis of *E. coli* is the breakdown of peptidoglycan of the bacterial cell wall. The lysis operon is made up of three overlapping genes: S, R, Rz. S gene product is a protein that causes porin formation in the cell membrane (Friedman & Gottesman, 1983). Rz encodes for a less-characterized protein, which is believed to act as an endopeptidase (Garrett et al., 1981). Finally, the gene product of R, bacteriophage lambda lysozyme (laL), is the subject of the present thesis.

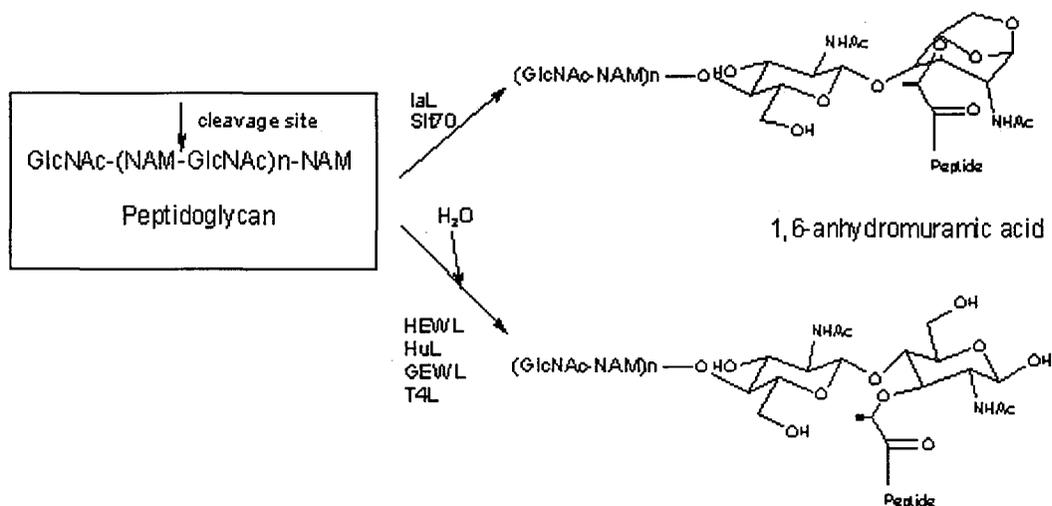
### **1.1 The Atypical Lysozyme - Lambda Lysozyme (laL)**

Lambda lysozyme (laL) is a small basic protein composed of 154 amino acids. Its molecular mass and pI are approximately 18 kDa and 9.7 respectively (Dewel, 1994). Although it is given the name lambda lysozyme, the protein differs in many ways from a typical lysozyme. Nonetheless, researchers have continued to use the name because like other lysozymes, laL

catalyzes the cleavage of the  $\beta$ 1-4 glycosidic bond between N-acetylmuramic acid (NAM) and N-acetylglucosamine (GlcNAc) of peptidoglycan of the bacterial cell wall (Fig. 1.1). However, laL violates the true definition of a lysozyme in that it is **not** a mucopeptide  $\beta$ -1,4-N-acetylmuramyl-hydrolase. A lysozyme requires a water molecule to nucleophilically attack the C1 atom of NAM, producing a reducible end product, thus acquiring the name hydrolase (Fig. 1.2). LaL differs from other lysozymes in the last step of the enzyme reaction where it uses the hydroxyl group (O6) in NAM as the nucleophile instead of water; thus, it cannot be called a hydrolase (Fig. 1.2). The resulting product is a non-reducible 1,6-anhydro-N-acetylmuramic acid. As Bienkowska-Szewczyk, *et al.* (1981) correctly identified for the first time, the lambda R gene product is a murein transglycosylase. To date, only the structure of a mutant laL has been solved (Evrard *et al.*, 1998); its structure will not be introduced here to avoid repetition in the result section of this thesis. Nonetheless, the structure was solved without any ligand bound, thus making it difficult to make a detailed description of the active site.



**Figure 1.1 Structure of peptidoglycan.** A heteropolymer of alternating ( $\beta$ 1-4) linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (NAM) units. The arrows indicate the  $\beta$ 1-4 glycosidic linkage where laL cleaves.

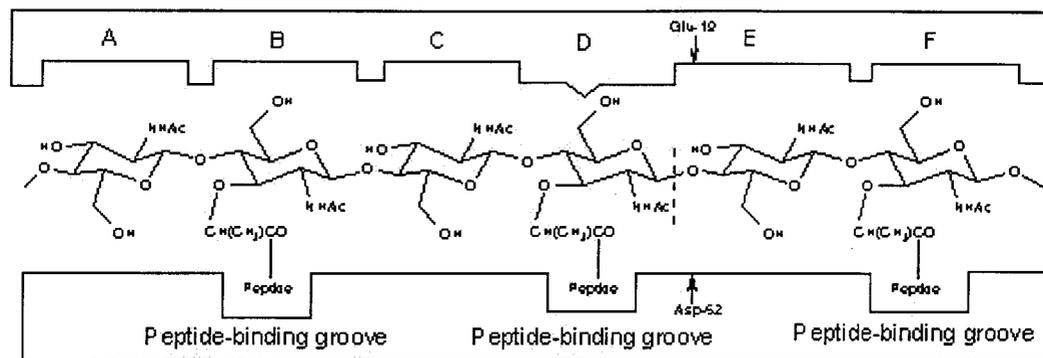


**Figure 1.2 Reaction pathways for IaL and other lysozymes.** In contrast to other lysozymes, IaL produces a non-reducible 1,6-anhydromuramic acid (Taylor, *et al.*, 1975).

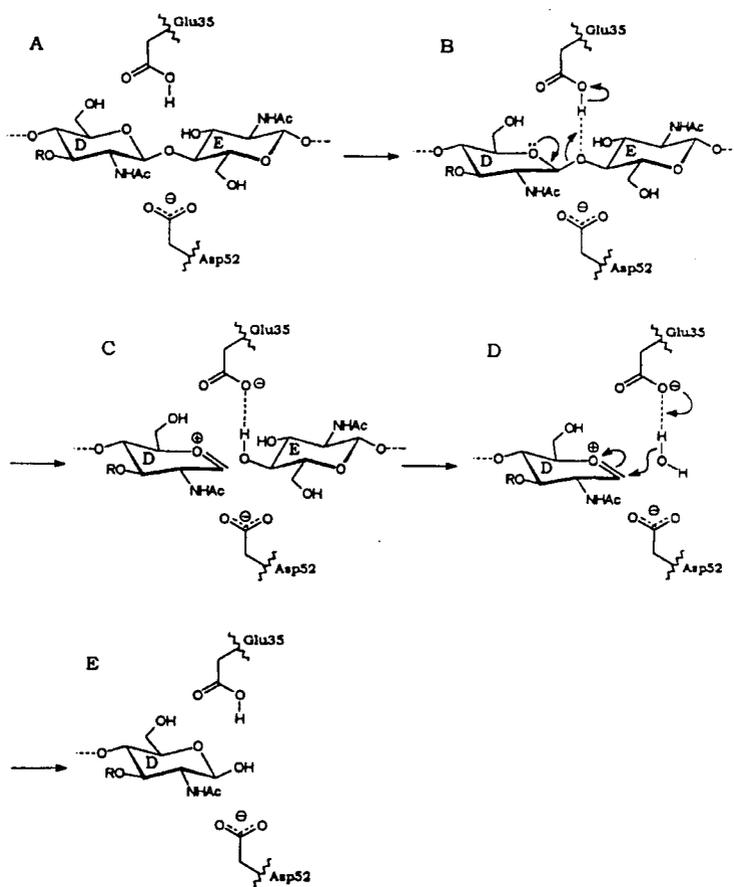
## 1.2 Lysozyme Structure and Mechanism

Sharing the same substrate and first step of the enzyme reaction, it is expected that IaL will have a similar mechanism as other common lysozymes. Therefore, it is worthwhile to introduce the current available models for the enzymatic mechanism of lysozyme for later comparisons.

According to their structural, enzymatic, and physical properties, lysozymes can be categorized into four main families: c (chicken) type, g (goose) type, p (phage) type, and ch (chalaropsis) type. Clear sequence homologies can be observed in lysozymes within the same family but not between different families. As one of the first proteins solved by x-ray crystallography in 1965 (Phillips, *et al.*), the three-dimensional structure of lysozyme has been extensively studied ever since; thus, abundant atomic information is available in the understanding of the enzyme mechanism. The general overall structure of a lysozyme can be described as bilobal, with the active site shaped as a deep crevice located in the middle between the two lobes, namely the upper and lower domains. The protein contains  $\beta$ -sheets in the lower domain, while it is highly helical in the upper domain. The structure of the hen egg white lysozyme (HEWL) complexed to a chitin trisaccharide fills up half the length of the cleft; the region where the sugar



a.



b.

**Figure 1.3 The Phillips model.** (a). Schematic diagram of the proposed hexasaccharide-binding mode in HEWL. The dashed-line shows the cleavage site. (b). Proposed catalytic mechanism for HEWL. (Phillips, 1996)

is bound are designated subsites A to C with A being the site closest to the solvent (Fig. 1.3a) (Phillips, 1966). Three additional sugar rings binding to subsites D to F are subsequently fitted into the active site by molecular modeling (Phillips, 1966). This HEWL/(GlcNAc)<sub>6</sub> complex structure becomes the basis of the current most accepted model for lysozyme action, the Phillips mechanism. Although different mechanisms for lysozyme action exist, only the Phillips model will be introduced here since it is supported by extensive biochemical evidence (Strynadka & James, 1991).

### 1.2.1 The Phillips Model

If the peptidoglycan interacts with the enzyme the same way as chitin, then NAM, which contains a D-lactoyl ether at C3, can only bind to subsites B, D, or F. According to the HEWL/(GlcNAc)<sub>6</sub> model, the O3 atoms in the sugar rings binding to subsites A, C, and E, are pointing into the cleft (Phillips, 1966). As a result, it is sterically impossible for NAM to bind to these sites. The structural data from a T4 mutant/(GlcNAc-NAM) complex show that the peptide moiety of NAM binds across a groove on the protein surface (Figure 1.7) (Kuroki, *et al.*, 1993). Peptide-binding grooves in subsites B, D, and F are depicted schematically in Figure 1.3a. Phillips proposed that the glycosidic bond cleavage occurs between subsites D and E based on the observations that cleavage can only occur between a NAM and GlcNAc residue, and that HEWL complexed with (GlcNAc)<sub>3</sub> binding to subsites A to C is extremely stable. The hypothesis quickly points to the two residues (Glu-35 and Asp-52) flanking subsites D and E to be the catalytic residues in HEWL (Phillips, 1966). In addition to the prediction of the two catalytic residues, the theoretical HEWL/(GlcNAc)<sub>6</sub> model also suggests that in order for the sugar ring in subsite D to interact optimally with the protein, it must adopt a distorted half-chair conformation (Phillips, 1966). The Phillips mechanism elegantly takes into account the two predicted catalytic residues and the suggested distortion in sugar ring D to explain the possible enzyme mechanism, which is summarized in Figure 1.3b (Phillips, 1966).

The cleavage of the glycosidic bond is a general acid/base reaction where the unionized Glu-35 donates a proton to the glycosidic oxygen (O4), leading to the breaking of the C1-O4 bond and the formation of a carbonium ion at subsite D. If the sugar residue D is in a favourable chair conformation, the C6 and O6 atoms will clash into atoms in the enzyme. The close contact, which is depicted schematically in Figure 1.3a, can be avoided if the sugar ring is distorted so that C6 and O6 are in the axial position. This is consistent with the Phillips model since the flattened pyranose ring caused by the double bond in the carbonium ion will place C6 and O6 in the axial position (Fig. 1.3b). The positive distorted carbonium ion is stabilized by better contacts with the enzyme and the ionized negatively charged Asp-52. The disaccharide bound to subsites E and F then diffuses away from the active cleft, followed by a nucleophilic attack of a water molecule on the carbocation, C1, producing a reducible tetrasaccharide.

Various aspects of the Phillips model have been supported by biochemical means and sequence analysis. Both Glu-35 and Asp-52 have been found in analogous position in all lysozymes belonging to the same class as HEWL (type c) (Jolles & Jolles, 1984). Site-specific mutagenesis of these two residues totally eliminates the enzyme activity, indicating their importance in the reaction, as suggested in the Phillips model (Malcolm *et al.*, 1989). Equivalent residues have been located in other types of lysozymes. However, mutation of the equivalent Glu-35 seems to have a more deleterious effect than the equivalent Asp-52, suggesting a more critical role of Glu-35 in the reaction (Matsumura & Kirsch, 1996). In fact, the corresponding Asp-52 residue is absent in some lysozyme structures (e.g. GEWL (Weaver *et al.*, 1995) & Slt (Thunnissen *et al.*, 1994)).

Structural scientists have been studying the validity of the Phillips mechanism by solving various lysozyme structures complexed with saccharides or other similar ligands. At least one lysozyme structure is available in each family. With the exception of type ch lysozymes, it is apparent from these structures that lysozymes exhibit conserved structural motifs regardless of the low sequence identity in lysozymes from different families. In addition to HEWL, the structures of goose egg white lysozyme (GEWL) (Weaver, *et al.*, 1995), bacteriophage T4 lysozyme (T4L)

(Kuroki, *et al.*, 1993), and human lysozyme (HuL) (Song, *et al.*, 1994) have been solved with ligands bound to their active sites. Structural alignment of these structures not only shows the high degree of homology of these proteins, it also demonstrates that they have the same sugar-binding mode suggested by Phillips. In addition, Glu-35 and its equivalents can all be well superimposed in the same position. Section 1.3 will give a summary of some presentative lysozyme/ligand structures available to date and their implications on the Phillips model.

### **1.3 Lysozyme-Saccharide Complexes**

Various lysozyme-saccharide complexes have been solved in an attempt to further understand the enzyme mechanism. In all cases, a sugar binds to the enzyme in a similar fashion proposed by Phillips in which six subsites (A to F) in the protein can accommodate an individual ring. Despite efforts to obtain a structure solution with all sites occupied, only subsites A to D, binding to a saccharide ring, have been observed crystallographically. However, Song *et al.* did report a human lysozyme (HuL) structure complexed with a hexasaccharide (1994). Unfortunately, the saccharide has been cleaved into di- and tetrasaccharide during the crystallization process. The disaccharide resides at a location close to subsites E and F (Song, *et al.*, 1994). To date, no information is available for the atomic environment of the cleaved D-E glycosidic bond.

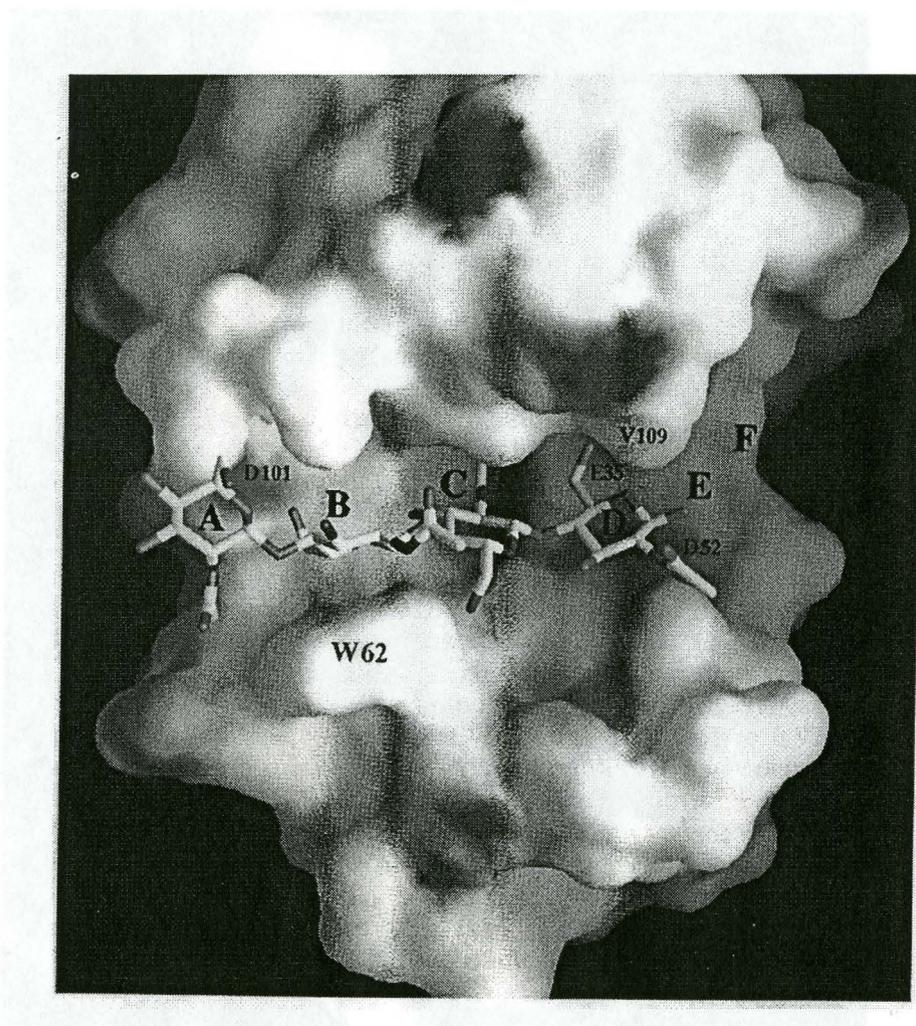
The requirement of a distorted ring D in the Phillips mechanism has been the most controversial aspect of the model. Distorted D rings were observed in all lysozyme structures complexed with a natural substrate, e.g. the HEWL structure complexed with (NAM-NAG-NAM) (Strynadka and James, 1991), and in the T4 mutant structure covalently linked to a peptidylidiasaccharide (Kurok *et al.*, 1993). However, inconsistencies were found in structures complexed with homopolymers of GlcNAc in which both distorted and undistorted ring D's were observed. This section will give a brief account of the binding mode in several lysozyme-saccharide complexes that have been used in comparative analysis with laL as part of this project.

### *HEWL/ligand complexes*

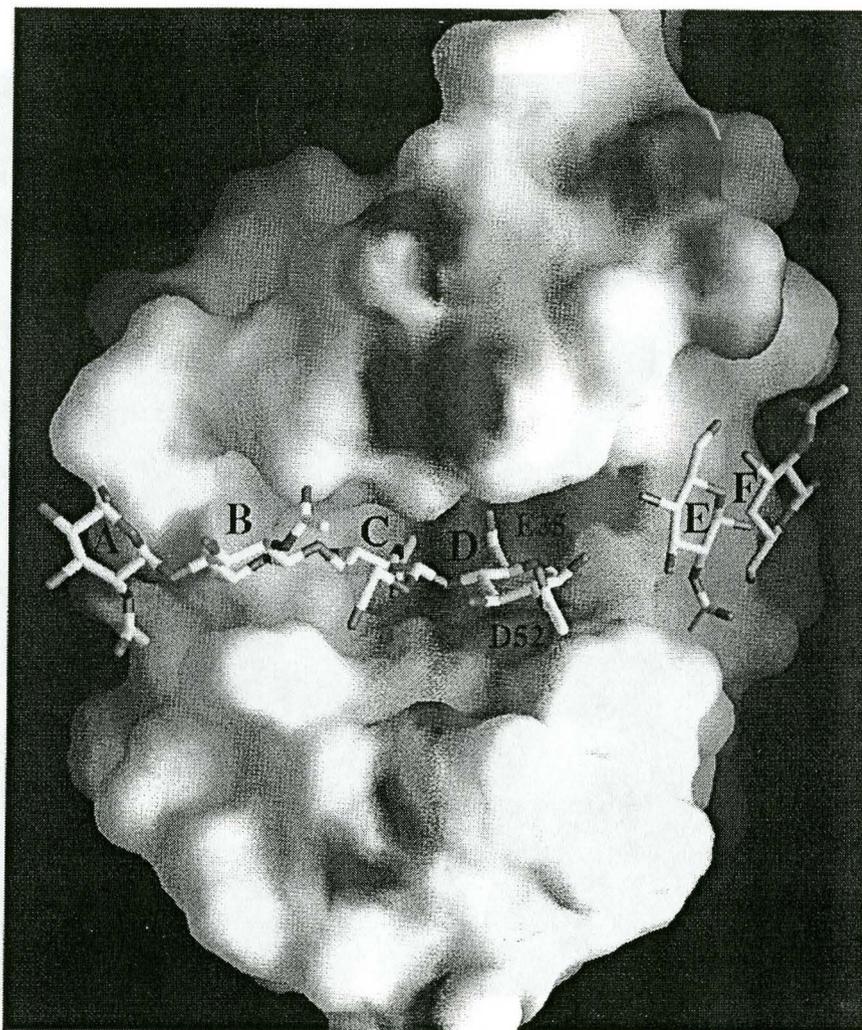
Wild type HEWL complexed with (GlcNAc)<sub>3</sub>, (GlcNAc)<sub>4</sub>, (NAM-GlcNAc-NAM), (GlcNAc)<sub>3</sub>-NAL transition state analog, as well as mutant HEWL complexed with (GlcNAc)<sub>4</sub> have been determined by various groups (Cheetham *et al.*, 1992; Maenaka *et al.*, 1995; Strynadka and James, 1991, Hadfield *et al.*, 1994). Ever since the proposal of a distorted D ring in the transition state was suggested, efforts have been put forth to produce a transition state analog to mimic the hypothesized half-chair D residue. Ford *et al.* (1974) determined the structure of the HEWL/(GlcNAc)<sub>3</sub>-NAL complex where NAL is the  $\gamma$ -lactone of GlcNAc. The conformation of the D residue was described as either a distorted boat or half chair (Ford *et al.*, 1974). Unfortunately, the structure was solved at a relatively low resolution (2.5 Å), making the structural analysis less reliable than the more recent HEWL structures.

The structure of NAM-GlcNAc-NAM, (MGM), complexed to HEWL was solved at 1.5 Å (Strynadka and James, 1991). The substrate binds to sites B, C, and D. Both rings B and C adopt full chair conformations, while ring D is distorted. The geometry is described as a sofa conformation with C2, C1, O5, and C4 being close to coplanar. The NAM residue in site D packs tightly in the protein cleft. The peptide linked to O3 interacts strongly with the protein, causing the C6-O6 group to clash with the 2-acetamido group in residue C. To avoid steric hindrance, the C6-O6 group is required to move to its axial position, which distorts the sugar ring into a sofa conformation. The presence of an extensive hydrogen bonding network to stabilize the distorted D residue gives further evidence to the geometry of this ring.

Various HEWL structures co-crystallized with tri- and tetrasaccharides have been solved. The coordinates from the PDB code representing a HEWL structure complexed to a tetrasaccharide, 1LZC, was used as a representative HEWL/GlcNAc structure for comparison in this project (Figure 1.4). In general, trisaccharides bind to sites A to C while tetrasaccharides bind to sites A to D. In all the complex structures, rings A to C show full conformations. While the majority of these structures are close to identical, the geometry and position of the sugar ring D shows notable variations from different structures. Moreover, the temperature factors for the D



**Figure 1.4** Surface depiction of HEWL/tetrasaccharide complex (PDB:1LZC). The tetrasaccharide binds to subsites A to D. Ring D in this structure adopts a full chair conformation. (Figure copied from Strynadka & James, 1996)



**Figure 1.5** Surface depiction of Hul/Hexasaccharide complex (PDB: 1LZS). The hexasaccharide is cleaved into (GlcNAc)<sub>4</sub> and (GlcNAc)<sub>2</sub>, with the former binds to subsites A to D and the latter binds to locations close to subsites E and F. Ring D in this structure is not distorted. Figure copied from Strynadka & James, 1996.

ring are often higher than the other sugar residues (Strynadka and James, 1996). Of all the HEWL/tetrasaccharide complex structures, distortion in the D ring is only observed in the mutant Asp52Ser (Hadfield *et al.*, 1994). Interestingly, the NAM residue in site D observed in the HEWL/MGM structure is located much deeper into the cleft compared to all the HEWL structures complexed to homopolymers of GlcNAc (Strynadka and James, 1996). It has been hypothesized that the lactoyl group in O3 is important in bringing the D residue closer into the cleft. The absence of the lactoyl group allows greater flexibility of this sugar ring. Without the constraints brought upon by residues E and F, it is more favourable for the D residue to adopt a full chair conformation at a site further away from the cleft to avoid steric hindrance.

#### *HuL(GlcNAc)<sub>6</sub>*

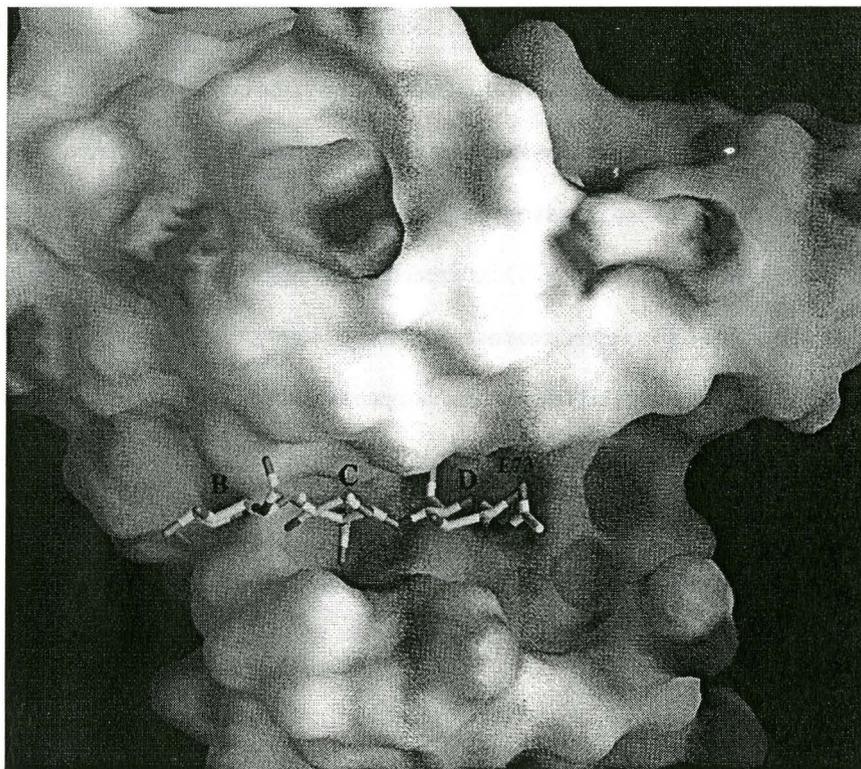
Human lysozyme was co-crystallized with (GlcNAc)<sub>6</sub> in hopes that the hexasaccharide would bind across the active site (Song *et al.*, 1994). Unfortunately, the crystal structure solved at 1.6 Å shows that the (GlcNAc)<sub>6</sub> molecule has been cleaved into (GlcNAc)<sub>4</sub> and (GlcNAc)<sub>2</sub> (PDB code:1LZS). The tetrasaccharide binds to sites A to D, and the disaccharide binds near sites E and F. Distortion is not observed in residue D; as expected, this ring is located close to the surface of the active site cleft. The O1 oxygen atom of ring D has no interaction with the catalytic OE1 atom in Glu35, further indicating that this residue has not penetrated deep enough to form the transition state (Song, *et al.*, 1994). Similar to the HEWL complexes, the authors also suggested that the lactoyl group of saccharide D might play a role in bringing the pyranose ring closer into the cleft. Finally, the Asp52 equivalence in HuL is Asp53; it is thought to have stabilization effects on the proposed oxocarbenium ion transition state.

#### *GEWL(GlcNAc)<sub>3</sub>*

A trisaccharide bound to goose-type lysozyme was solved at 1.6 Å (Weaver, *et al.*, 1995) (PDB code: 154L). (GlcNAc)<sub>3</sub> occupies subsites B, C, and D in full chair conformations. Residue C interacts most strongly with the protein as indicated in the more well-ordered density and lower

temperature factor. Saccharide D has weak interactions with the protein, consistent with the relatively high temperature factor. The Glu35 equivalent catalytic residue in GEWL is Glu73, but no counterpart of Asp52 can be found, thus questioning the role of this residue in stabilizing the transition state oxocarbenium ion.

Unlike HEWL, GEWL activity requires the peptide moiety in the NAM residue at site D (Arnheim *et al.*, 1973). As mentioned previously, Weaver *et al.*, (1995) also suggested that the binding of the peptide might provide additional energy so that saccharide D can form high energy distorted transition state. Not surprisingly, the D ring in this structure is situated further out of the active site than those that are distorted.



**Figure 1.6 Surface depiction of GEWL/(GlcNAc)<sub>3</sub> complex (PDB:154L).** The sugar is shown binding to sites B, C and D. The D ring is in full chair conformation. (Figure is copied from Strynadka & James, 1996).

*T4/(GlcNAc-NAM)-peptide complex*



**Figure 1.7 Surface depiction of T4 mutant complexed with a cell wall substrate (PDB: 148L).** The GlcNAc residue binds to subsite C. The protein linked NAM residue adopts a sofa conformation at the D site. The peptide moiety extends across a groove on the protein surface. (Copied from Strynadka & James, 1996).

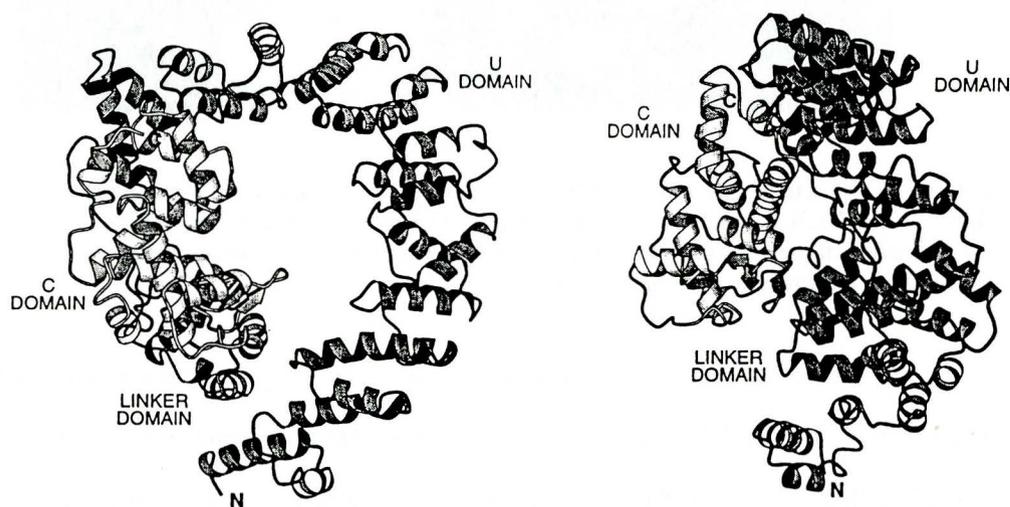
The Thr26Glu T4 mutant with the cell wall substrate covalently bound was solved at 1.9 Å (Kuroki, *et al.*, 1993) (PDB code: 148L). The GlcNAc-NAM sugar component occupies subsites C and D with the O1 atom in the NAM residue covalently attached to the side chain of Glu26.

The D sugar ring shows distortion in the structure. It is in a sofa conformation, which is stabilized by multiple interactions with the enzyme. The peptide component, L-Ala-D-Glu-Diaminopimelic

acid-D-Ala, which is linked to the O3 atom of the NAM residue, is located at a groove on the protein surface. The peptide interacts extensively at the surface groove via direct hydrogen bonds. Like GEWL, T4L activity requires a peptide moiety. The peptide-binding region suggests that the peptide somehow participates in catalysis possibly by providing energy for the distortion in the transition state. Finally, the Glu35 equivalent in T4L is believed to be Glu11, while the Asp52 counterpart is thought to be Asp20.

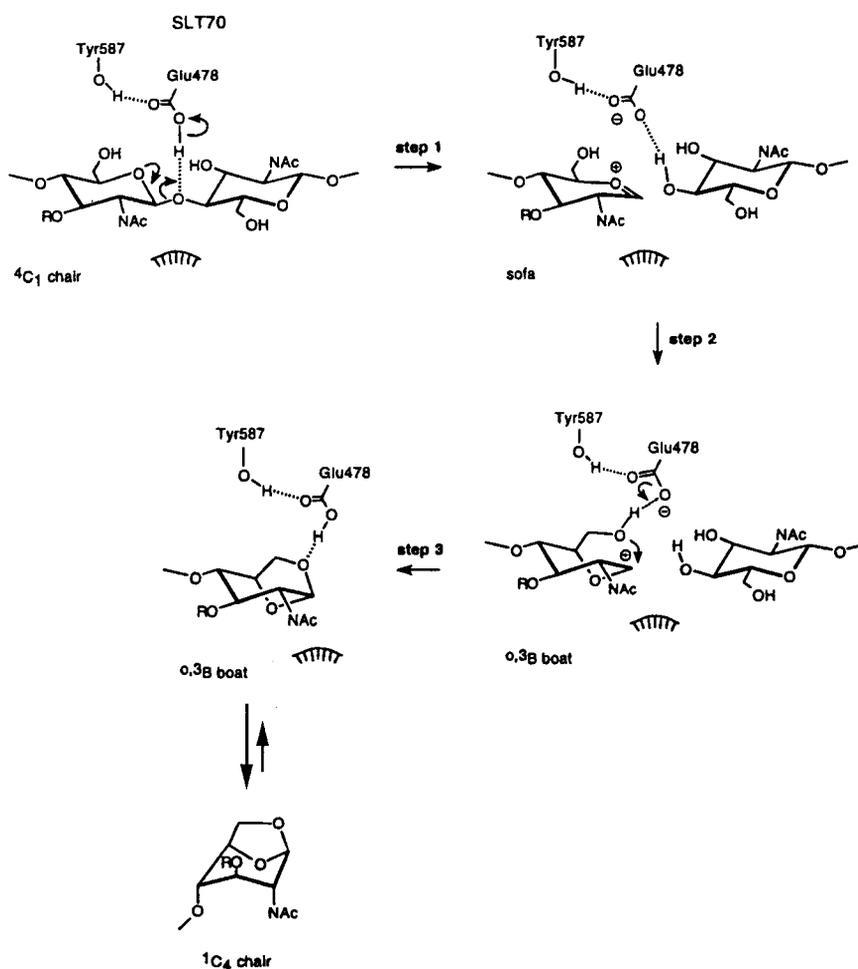
### 1.4 *E. coli* Lytic Transglycosylase Structure and Mechanism

Lambda lysozyme is expected to have a mechanism more analogous to *E. coli* lytic transglycosylase than to common lysozymes. In addition to having the same substrate, *E. coli* transglycosylase also generates the same 1,6-anhydromuramic acid end-product. Participating in the development of *E. coli*, the transglycosylase is involved in the extension of cell wall, where it cleaves the peptidoglycan, so that new cell wall materials can be inserted, allowing the bacteria to grow (Ghuysen, *et al.*, 1973). The crystal structure of a lytic transglycosylase (Slf70) has been solved (Figure 1.8), and its transglycosylase mechanism has been proposed (Figure 1.9) (Thunnissen *et al.*, 1994; Thunnissen *et al.*, 1995 a, b).



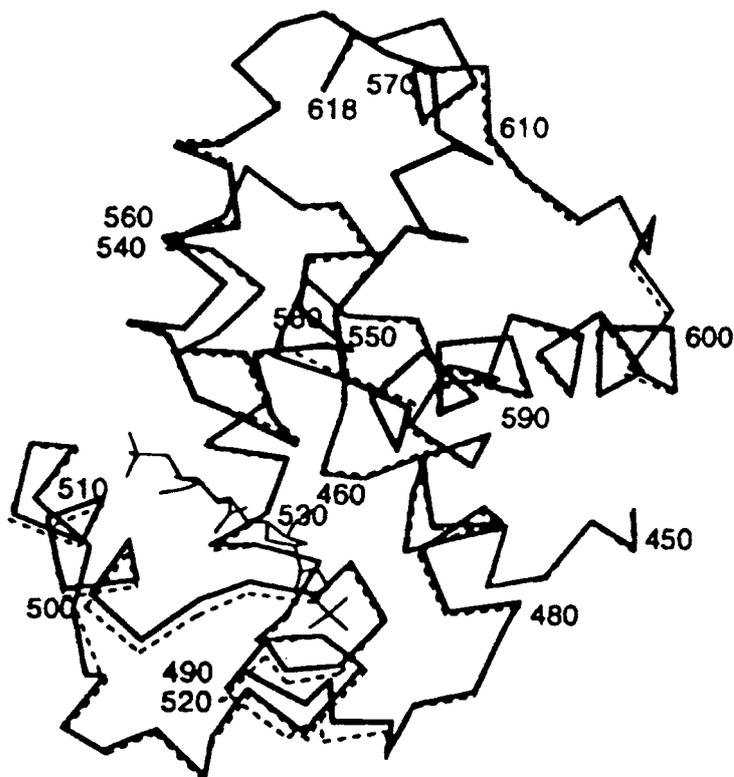
**Figure 1.8 Schematic diagram of Slf70.** The linker domain and U domain form a doughnut ring as shown on the left. The lysozyme-like C domain is located on top of the ring as depicted on the right (Dijkstra & Thunnissen, 1994).

The Slr protein is divided into three domains. The N-terminal domain consists of 22  $\alpha$ -helices packed in a U-shape conformation; thus, this domain is called the U-domain. A small linker domain, containing four  $\alpha$ -helices, join the U-domain and C-terminal domain, whose fold resembles a typical lysozyme. The U-domain and linker domain form a closed ring, featuring a doughnut-shaped structure with the C-domain located on top of the ring (Figure 1.8).



**Figure 1.9** Proposed mechanism for Slr70. Glu-478 is the catalytic residue equivalent to Glu-35 in HEWL. However, no counterpart of Asp-52 can be identified (Thunnissen *et al.*, 1995b). See text for details.

The proposed mechanism for Slf70 is very similar to the Phillips mechanism (Figure 1.9). Glu-478 has the same role as Glu-35 in HEWL in that it acts as a general acid and protonates the glycosidic oxygen. The subsequent bond rearrangement generates the oxocarbenium ion intermediate followed by the nucleophilic attack of the C1 atom by the C6 hydroxyl group of NAM. Slf70 lacks a counterpart to Asp-52, which is proposed to stabilize the positive oxocarbenium ion. Thunnissen *et al.* (1995b) suggested that since the transglycosylation mechanism does not require the diffusion of the disaccharide and a water molecule in and out of the active site, the enzyme has less need to stabilize the oxocarbenium intermediate. Alternatively, it is also proposed that the oxocarbenium ion can be stabilized by a partial negative charge on the



**Figure 1.10 Structure of Slf70 complexed with bulgecin.** Shown here is a superposition of the C $\alpha$  backbone of the lysozyme-like C-domain of Slf70 with and without bulgecin. The bulgecin molecule is shown in thin lines. The uncomplexed Slf70 is depicted in dotted lines (Thunnissen *et al.*, 1995b).

acetamido carbonyl group (Thunnissen *et al.*, 1995a). The structure of Slt70 with a bulgecin molecule bound to the active site was solved to 2.8 Å (PDB code: 1SLY) (Figure 1.10). The bulgecin molecule binds to subsites B, C, and D. The 4-O-sulfonate group occupies site B, GlcNAc at site C, and proline residue at site D. The planar proline part of the bulgecin mimics the oxocarbenium ion in the proposed transition state. Thunnissen *et al.* (1995b) superimposed the disaccharide from the coordinates of the T4L mutant with the bulgecin and discovered that the NAM residue in site D fits well with the proline of bulgecin. Once again, this structure is consistent with the proposed sofa conformation of the D ring in the transition state.

## **1.5 Thesis Objective**

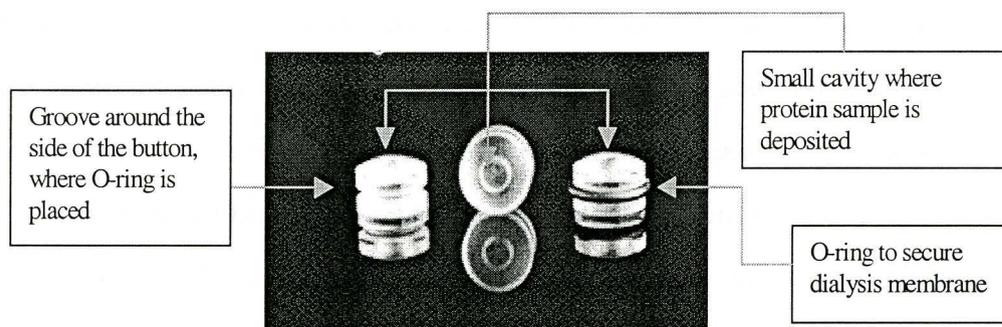
Although the structure of the mutant IaL has been solved by Evrard *et al.* (1998), there is no structural information on the saccharide-binding mode in IaL. The main subject in this thesis is the structure of IaL complexed to a hexasaccharide (IaL·(GlcNAc)<sub>6</sub>). Hitherto, only the HuL structure is available with subsites E and F occupied. The structure of IaL·(GlcNAc)<sub>6</sub> will give further insight into the importance of these two sugar-binding sites. With ample structural information on lysozymes and *E. coli* lytic transglycosylase, the emphasis of the present thesis is on the comparative analysis of the saccharide-binding mode in IaL and various lysozymes and Slt70. In addition, some interesting similarities and differences which might be responsible for the distinctive biochemical properties of IaL will be discussed.

## Chapter 2 General Experimental Methods

### 2.1 Crystallization

The protein used for crystallization was purified using previously described protocol and was prepared by our collaborators, Dr. John Honek and Dr. Henry Duewel (Duewel, *et al.*, 1994).

Wild-type lambda lysozyme crystals (referred to as wlaL) were grown at room temperature using hanging drop vapour diffusion method (Weber, *et al.*, 1997). 2  $\mu$ l of the protein solution (9.2 mg/ml in 6.2 mM phosphate buffer at pH 7.0) was mixed with an aliquot of 2  $\mu$ l of mother liquor (20% PEG 400, 0.1M Tris pH 8.5, 3mM NiCl<sub>2</sub>) and 1  $\mu$ l of 0.1M phenol additive. The mixture was suspended on a siliconized cover slip over 300  $\mu$ l of mother liquor as described above.



**Figure 2.1 Dialysis Buttons (adapted from Hampton Research catalogue, vol.8 n.1, 1998 p.25).** In microdialysis crystallization, the protein sample is separated from the mother liquor by a semi-permeable membrane in which only small molecules can pass through. The membrane is secured by an O-ring around a groove on the side of the button.

Complexes of laL and its inhibitor, (GlcNAc)<sub>6</sub> were prepared using protein solution containing protein/inhibitor ratio of about 1:5 moles. Co-crystals of laL and (GlcNAc)<sub>6</sub> (referred to as laL•(GlcNAc)<sub>6</sub>) were grown using dialysis buttons (Figure 2.1). 5  $\mu$ l of the protein sample was deposited in the 5  $\mu$ l depression on the button, which was then sealed with a dialysis

membrane, and subsequently suspended in a vial containing the mother liquor (0.1 M sodium acetate pH 4.6, 0.1 ammonium sulfate, 20% (w/v) PEG 2000 MME).

## **2.2 Cryocrystallography**

Cryocrystallography technique was used for all data collection. The advantages of this technique include minimization of radiation damage of the crystal, reduction of background scatter and absorption, effective increase in resolution limit, and decrease in thermal parameters (Rodgers, 1997). W1aL and 1aL•(GlcNAc)<sub>6</sub> crystals were first soaked in a drop of cryoprotectant containing the same solutions as the mother liquor except that the PEG concentration was increased to at least 30% for both types of crystals. Then the crystals were scooped up on a loop and flash frozen in liquid nitrogen. The flash freezing step was performed quickly so that water in the protein would not form ice crystals but turn to amorphous ice. All data was collected at -170°C.

## **2.3 Data Collection**

### **2.3.1 Wild-type Lambda Lysozyme**

W1aL diffraction data was collected using the beamline F2 at CHESS and recorded using a ADSC Quantum-4 CCD.

### **2.3.2 Lambda Lysozyme and Inhibitor Complex**

1aL•(GlcNAc)<sub>6</sub> diffraction data set was collected using Rigaku RU200 operating at 50kV and 60 mA equipped with a rotating anode. The beam was passed through a nickel filter to select out CuK $\alpha$  radiation and Super double focusing mirrors to intensify the beam. X-ray diffraction data was recorded using the Rigaku R-axis IIC area detector.

## **2.4 Data Processing**

Both crystal types were initially mounted on the goniometer and exposed for 10 minutes to check their diffraction qualities before actual collections. Using a graphical display program, Xdisp, the crystal qualities were judged by the extent of the resolution and the mosaic spread. All

diffraction data were processed using the HKL package (Xdisp, Denzo, Scalepack) (Otwinowski, Z., *et al.*, 1997). The computer software processes data in five major steps: 1. autoindexing, 2. refinement of the crystal and detector parameters, 3. integration of the diffraction maxima, 4. scaling and merging, 5. global refinement. Each step will be discussed briefly in the following sections.

### 2.4.1 Autoindexing

The first step of autoindexing is to identify a characteristic pattern of diffraction spots from an image and deduce a crystal lattice that matches the pattern. A peak search is first performed by the software. Using Bragg's Law,  $2d\sin\theta=n\lambda$ , each diffraction spot is mapped onto reciprocal space which is related to  $d$ .  $\lambda$  is known prior to the experiment. Since the precise diffraction angles for each reflection are unknown for oscillation data, the centre of the oscillation range is used as the best estimate of  $\theta$  (Otwinowski, Z., *et al.*, 1997). Denzo finds values of each reciprocal index (h, k, or l) independently for all reflections. Finding one index is equivalent to finding a real-space direction of the crystal axis. Following the search, the program finds 3 linearly independent vectors with the smallest determinant (i.e. unit cell volume) which index all or almost all of the peaks identified in the first step. Using these vectors as a basis, the program finds the best cells (dimensions and angles) for all of the 14 Bravais lattices according to the International Tables of Crystallography (Hahn, 1995). Some distortions between these "best cells" and the "ideal cell" from each Bravais lattice will be observed. Generally, the desired Bravais lattice is chosen in a compromise between high symmetry and low distortion.

### 2.4.2 Refinement of Crystal and Detector Parameters

Autoindexing provides a rough estimate of the crystal and detector parameters based on strong reflections. The Bravais lattice chosen in the previous step allows Denzo to predict reflections. The purpose of this step is to refine these predictions using least square methods. More precise crystal and detector orientations are required for the subsequent integration step.

### 2.4.3 Integration of the Diffraction Maxima

The purpose of this step is to calculate the diffraction intensity. In order to obtain the correct value, the detector background must be subtracted from the reflection profile. Denzo fits the profile and measures the intensity of only the portion of a reflection that is located within a spot prediction (Gewirth, 1995). In this step, the user must specify the dimension of the predicted integration spot and the corresponding background regions. One spot shape is used for the whole data set; therefore, it should be chosen so that it covers all the observed reflections without overlapping neighbouring reflections. Denzo computes the diffraction intensity using profile fitting which allows information of the spot shape to be included in the calculations, making the profile fitted intensity more accurate than simply subtracting the background pixels from the total number of pixels in a spot. The major advantage of profile fitting is for measuring high angle weak diffraction spots whose intensities are statistically less reliable. Adding spot shape information allows the software to determine with better accuracy whether or not a weak diffraction is noise. The profile or shape of a spot is predicted using the average profile of the neighboring strong reflections within a defined radius. For example, if the profile of a weak diffraction does not fit well with the average profile determined from neighboring strong reflections, this spot will be rejected even if its absolute intensity is above the threshold. The purpose of this integration step is to assign an intensity value for each hkl index. At the end of this step, Denzo gives out a list of raw measured intensities for each hkl index and their standard deviation.

### 2.4.4 Scaling, Merging and Global Refinement

The intensity measurements from the integration step do not account for variations in redundant measurements between different frames due to experimental set up such as Lorentz factor (the relative speed of a reflection through the Ewald sphere), air absorption and crystal decay during data collection. For each frame in the data set, the software scalepack calculates a correction factor containing a B factor component that accounts for crystal decay and a scale factor component

that accounts for other experimental errors. This correction factor and Lorentz corrections are applied to each raw intensity measurement,  $I$ , to give the scaled intensity,  $I_{corr}$ . Outlier reflections are rejected and systematic errors of the data set are analyzed. Finally, redundant measurements of intensity are averaged and their standard deviations are estimated.  $R_{sym}$  is a measurement of the internal agreement of the data set:

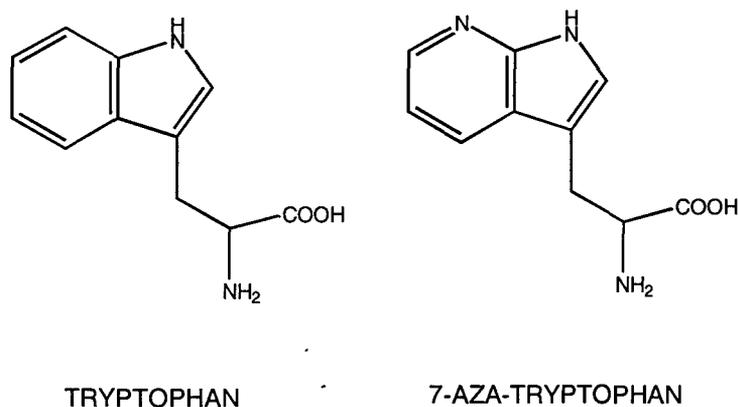
$$R_{sym}(I) = \frac{\sum_h \sum_i |I_h - I_{hi}|}{\sum_h \sum_i I_{hi}} \times 100 \quad (4)$$

where  $I_h$  is the weighted mean measured intensity of the observation  $h$ ,  $I_{hi}$  is the intensity of a given symmetry-related reflection which should be the same as  $I_h$ .

The crystal and detector parameters are refined previously from a single image. Global refinement allows a more precise determination of these parameters using the whole data set. As these parameters are refined, the correction factor for all films are shifted to minimize the difference between average  $I_{corr}$ ,  $\langle I_{corr} \rangle$ , and  $I_{corr}$  for each individual measurement. A new  $\langle I_{corr} \rangle$  is calculated after each round of refinement. The process is repeated until convergence is reached.

## 2.5 Structure Solution

The structure of a mutant form of laL, mlaL, has been solved to 2.3 Å (Evrard *et al.*, 1998). Four tryptophan residues were mutated to 7-aza-tryptophans (Figure 2.2). The mlaL packs in the crystal as a trimer. Monomer A was used as the search model for molecular replacement. The computation was performed using XPLOR version 3.8. The following sections give a general background on the molecular replacement technique with emphasis on the approach used in XPLOR. A more specific description of the experimental procedures will be given in subsequent chapters



**Figure 2.2 Structure of 7-aza-tryptophan.**

### 2.5.1 Molecular Replacement

Molecular replacement can be used to solve the phase problem when there is a homologous structure available. The relationship of the two homologous objects,  $(x_1, x_2)$ , can be described in the following equation:

$$x_1 = R \times x_2 + t \quad (5)$$

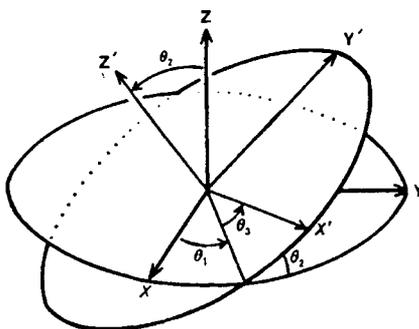
where  $R$  is the rotation matrix and  $t$  is the translation vector. In XPLOR, this six-dimensional search is divided into rotation and translation search. Essentially, the search model is first rotated then translated in the unit cell of the target crystal to obtain the best agreement between the calculated and the observed diffraction data.

#### 2.5.1.1 Rotation Search

##### *Rotation angle conventions*

The Eulerian coordinate system is widely used to describe angles in the rotation function. Figure 2.3 shows the definition of the eulerian angles  $\theta_1, \theta_2, \theta_3$ .  $\theta_1$  is the rotation of the  $x$  axis around the  $z$  axis.  $\theta_2$  is the rotation of the  $z$  axis around the moved  $x$  axis.  $\theta_3$  is the rotation around the moved  $z$  axis producing a new  $y$  axis. A disadvantage of using the Eulerian angle

system is that when  $\theta_2$  is close to  $0^\circ$  or  $180^\circ$ ,  $\theta_1$  and  $\theta_3$  are rotations of nearly parallel axis and are strongly coupled. When the rotation vector  $R$  is plotted on a  $\theta_1/\theta_3$  grid, peaks in regions where  $\theta_2$  is close to  $0^\circ$  or  $180^\circ$  are streaky and distorted. The problem of uneven and inefficient sampling is solved by using pseudo-orthogonal Eulerian angles ( $\theta_+ = \theta_1 + \theta_3$ ,  $\theta_- = \theta_1 - \theta_3$ ,  $\theta_2 = \theta_2$ ) instead (Lattman, 1985).



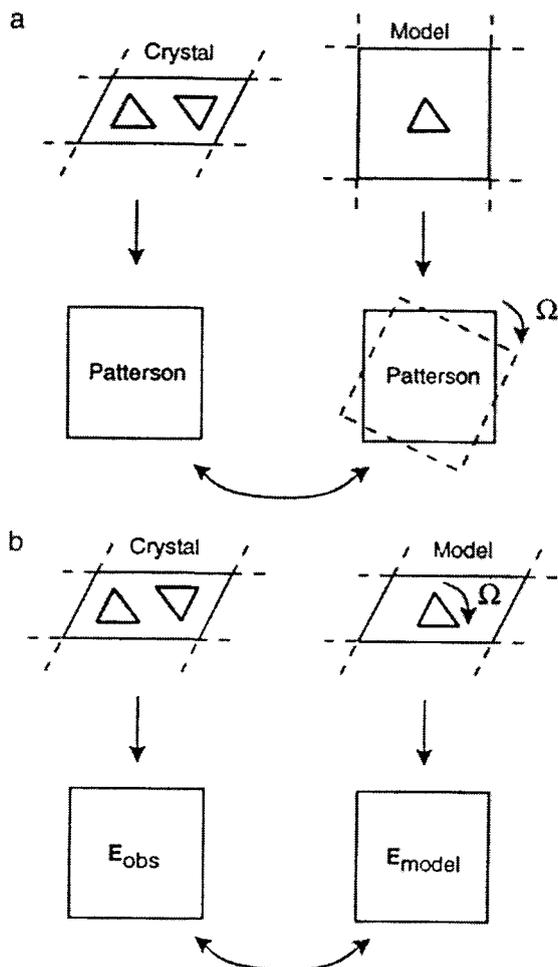
**Figure 2.3 Eulerian angles.** See text for definition. (Copied from Rossman & Blow, 1962).

#### *Real space Patterson search*

A Patterson function is a special Fourier function that contains all the intensity information in a data set. The corresponding Patterson map contains all the interatomic vectors within a molecule. The Patterson maps of the search model,  $P_1$ , and the crystal,  $P_2$ , are calculated by Fast Fourier Transform. The two maps are rotated with respect to each other using the Eulerian angles. The orientations of the search model are sampled using pseudo-orthogonal Eulerian angles. For every sampled rotation, the correlation between the two maps are determined by computing the product function (Eq. 6) between the rotated model peaks and the interpolated peaks of the crystal:

$$RoI(\Omega) = \int_U P_x(\mathbf{r}) P_m(\Omega \mathbf{r}) dV \quad (6)$$

where  $\Omega$  is a  $3 \times 3$  rotation matrix represented by the three Eulerian angles,  $\mathbf{r}$  is the integration variable,  $U$  is the volume of the integration,  $P_x$  is the Patterson function of the crystal and  $P_m$  is the Patterson function of the search model.



**Figure 2.4 Comparison between the two rotation methods. a). Conventional real space Patterson search.** The Patterson maps of the search model and the crystal are rotated with respect to each other to find the best correlation between the two. **b). Direct rotation search.** The search model is rotated. The correlation between the observed and calculated normalized structure factors is computed. (Copied from Brünger, 1997.)

#### *Direct rotation search*

Besides real space Patterson search, the XPLOR package also includes direct rotation search strategy. The comparison between the two methods is illustrated in Figure 2.4. The search model is placed in a cell with  $P_1$  symmetry and has cell dimensions and angles equal to those of the target crystal. The model (the Patterson map in the case of Patterson search) is rotated directly. The correlation between the observed ( $E_{obs}$ ) and calculated ( $E_m$ ) normalized structure factors is

computed for every sampled orientation:

$$Rot_{direct}(\Omega) = PC = \frac{\langle |E_{obs}|^2 |E_m(\Omega)|^2 - \langle |E_{obs}|^2 \rangle \langle |E_m(\Omega)|^2 \rangle \rangle}{\sqrt{\langle |E_{obs}|^4 - \langle |E_{obs}|^2 \rangle^2 \rangle \langle |E_m(\Omega)|^4 - \langle |E_m(\Omega)|^2 \rangle^2 \rangle}} \quad (7)$$

where  $\Omega$  is the rotation matrix and  $\langle \rangle$  notes an average calculated over all observed reflections in a  $P_1$  cell. Although direct rotation search is computationally expensive, it has the greatest signal-to-noise ratio when compared to other rotation search methods (Brünger, 1997).

#### *Patterson correlation refinement (PC-refinement)*

PC-refinement of the highest peaks generated from the rotation search improves the discrimination between correct and incorrect orientations of the search model. The refinement can greatly improve the accuracy of the search model before translation search, which increases the success of molecular replacement. The basis of PC-refinement is to introduce a parameter  $p$ , which describes the expected differences between the search model and the crystal. For example, in a multi-domain structure, one would expect the relative position of each domain to change in a different crystal form. The atoms from each domain can be grouped into a rigid body. Then, the parameters  $p$ 's are refined against negative PC (-PC). In the minimization algorithms, finding the smallest -PC is equivalent to finding the maximum PC.

### **2.5.1.2 Translation Search**

The final step of molecular replacement involves placing the correctly oriented search model relative to crystallographic space group symmetry elements. When noncrystallographic symmetry exists in the crystal, the positioning of the subunit relative to each other is also performed. The search involves placing the model over a three-dimensional grid, generating all the symmetry-related molecules and finally, computing the fitting between calculated and observed data.

## 2.6 Structure Refinement

The initial phase obtained from molecular replacement is approximate and contains errors. Refinement involves adjusting the position and thermal parameter of each atom in order to improve the agreement between the observed data and the calculated data. As each parameter is adjusted, the calculated phase also changes. The phase and the electron density map should improve as refinement proceeds.

The number of observations should significantly exceed the number of parameters to be adjusted. The greater this ratio is, the more reliable the model will be. Depending on the quality of the data, either the number of observations should be maximized or the number of parameters used to describe the model should be minimized. One challenge in refining a protein structure is that the resolution does not usually exceed 1.5 Å, which greatly limits the number of observations available. Fortunately, some structural features of the molecule are known and can be included in the refinement. This additional information includes bond lengths, bond angles, dihedral torsion angles, chiral centers, planarity of aromatic rings, van der Waals, and hydrogen bonding interactions. Refinement is converged when further changes in parameters no longer improve the agreement between observed and calculated structure factors. Refinements of  $\text{laL} \cdot (\text{GlcNAc})_6$  were done using XPLOR version 3.8 (Brünger, 1992). The software uses the conjugate gradient minimization algorithm for both positional and B factor refinements. The general refinement scheme is described in the following sections. Specific strategies will be discussed in subsequent chapters.

### 2.6.1 Rigid Body Refinement

After the initial model is obtained, a rigid body refinement can be run using the same idea as in PC-refinement for the rotation search. This procedure is especially useful in multi-domain structures where domain movements are observed. As in PC-refinement, a group of atoms can be treated as a rigid body, in which they do not move relative to each other. The rotational and translational parameters of each domain are refined.

## 2.6.2 Positional Refinement

The goal of positional refinement is to find the global minimum of the target function:

$$E = E_{chem} + w_{x-ray} E_{x-ray} \quad (8)$$

where  $E_{chem}$  describes the geometry of the model based on information obtained from small molecule crystallography. This empirical energy term accounts for the covalent interactions such as bond lengths, bond angles, dihedral torsion angles, chiral centers, planarity of aromatic rings, and also the nonbonded interactions such as van der Waals, hydrogen bonding and electrostatic forces.  $E_{x-ray}$  term describes the discrepancies between the observed and calculated diffraction data:

$$E_{x-ray} = \sum \left( |F_o| - k|F_c| \right)^2 \quad (9)$$

where  $|F_o|$  is the amplitude of the observed structure factor,  $|F_c|$  is the amplitude of the calculated structure factor and  $k$  is a scale factor chosen to minimize the difference between the two amplitudes.  $w_{x-ray}$  is the ideal weight between  $E_{chem}$  and  $E_{x-ray}$  computed before the start of refinement. A greater weight puts more emphasis on the  $E_{x-ray}$  term during minimization of this target function. If the diffraction data is weak, one should use a smaller weight to put more emphasis on the geometry of the molecule during refinement.

## 2.6.3 Thermal Factor Refinement

The scattering factor,  $f$ , describes the scattering power of a stationary atom. In reality, atoms are never at rest; they are constantly vibrating. The extent of this vibration depends on the temperature, mass of the atom, and how strongly it is interacting with the surrounding atoms

(Stout and Jensen, 1989). Therefore, a proper scattering factor should include a mathematical term, which accounts for this thermal motion as shown below:

$$f = f_o e^{-B(\sin^2 \theta) / \lambda^2} \quad (10)$$

As a result, the calculated structure factor,  $F_{\text{calc}}$ , which is a summation of all the scattering factors in the model, should include this thermal factor as well. During refinement, this term must be adjusted according to the position of each atom and the x-ray term. XPLOR provides several approaches to this problem. First, an overall B factor refinement, which calculates one B factor for the whole model, can be computed isotropically or anisotropically. The latter case requires a refinement of six parameters, which describe the anisotropic B-factor tensor, while the former case only refines one parameter since the value is the same in all directions. Second, grouped B-factor refinement, which calculates two B-factors for each residue (usually one for main chain atoms and one for side chain atoms), can be done. Third, restrained individual isotropic B-factor refinement calculates an appropriate B-factor for each atom. The general rule of thumb is to include as many parameters as possible without over fitting the data; this will be discussed in more detail later.

#### 2.6.4 Bulk Solvent Correction

The protein molecules in a crystal are surrounded by solvent. Density from this bulk solvent should be accounted for in  $F_{\text{calc}}$  to bring it closer to  $F_{\text{obs}}$ . XPLOR computes a bulk solvent mask by setting all the grid points lying within half a van der Waals radii of the model atom and half the solvent radii to zero. After going through all atoms in the model, XPLOR defines the bulk solvent mask as the non-zero regions. Since the solvent density is not known, the solvent structure factors are scaled to the  $F_{\text{calc}}$  terms. The bulk solvent mask has sharp edges in the protein/mask boundary, causing a Fourier rippling effect (Brünger, 1992). The rippling occurrence would produce undesirable effects in the high-resolution terms. The addition of a large temperature factor can solve this problem by smoothening the sharp edges. In essence, the calculation of the

scaled bulk solvent structure factors involve finding the optimal values for the scale factor,  $k_{\text{solvent}}$  and the temperature factor,  $B_{\text{solvent}}$ . XPLOR first performs a crude search on the optimal value of  $k$ , which minimizes the R factor in the lowest resolution shell without adversely affecting the high-resolution R values. This optimal  $k$  value is used for subsequent crude search for the optimal B value. Finally, the process is repeated over a finer set of both  $k$  and B values. The bulk solvent structure factors can be incorporated into  $F_{\text{calc}}$  in subsequent refinement.

### 2.6.5 R factor vs. Free R

The correctness of the model is monitored by the conventional R factor:

$$R = \frac{\sum \| |F_o| - k|F_c| \|}{\sum |F_o|} \quad (11)$$

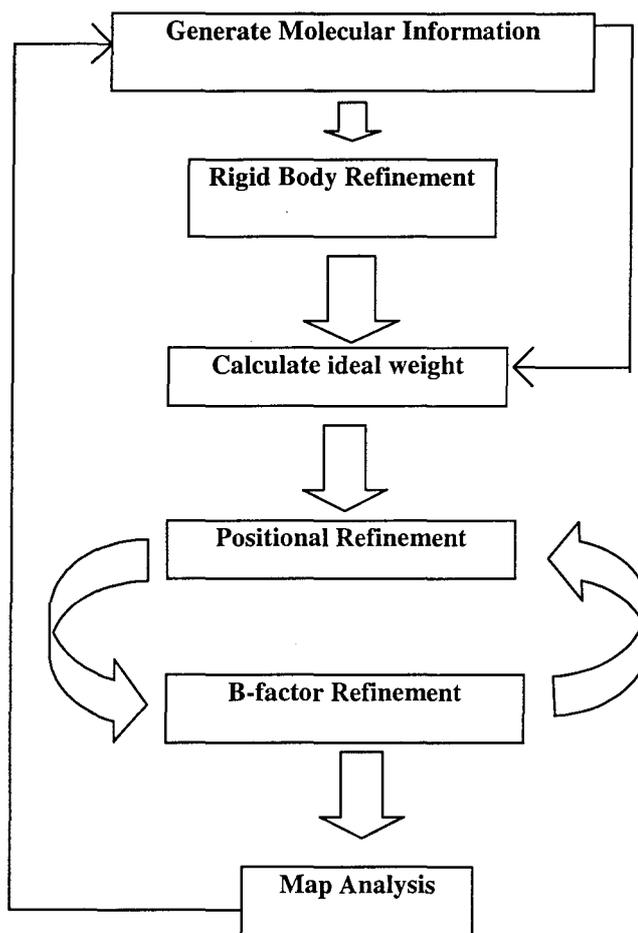
where  $|F_o|$  is the amplitude of the observed structure factor,  $|F_c|$  is the amplitude of the calculated structure factor and  $k$  is a scale factor chosen to minimize the difference between the two amplitudes. In a perfect structure, the R should equal zero. In a typical protein structure, a reasonable R factor should be around 0.20. R is closely related to the crystallographic residual,  $E_{x\text{-ray}}$  mentioned earlier. Essentially, the purpose of refinement is to minimize  $E_{x\text{-ray}}$ , hence increasing the accuracy of the model. R can be made arbitrarily small by refining more model parameters without improving the data at all. Thus, a low R does not necessarily reflect a reliable structure. Brünger (1992, 1993) proposed a method of cross validation using the Free R value. The procedure implemented in XPLOR randomly selects a certain percentage (usually 10%) of the experimental data, which comprise a 'test' set of reflections not used during refinement. The equation for calculating Free R is the same as that of the conventional R factor but only reflections in the 'test' set are included in the computation. Free R is highly correlated to phase accuracy (Brünger, 1993); therefore, it is a reliable way to monitor whether over fitting has occurred. A general rule of thumb is to include as many model parameters as possible in refinement without increasing the Free R value.

## 2.6.6 Graphical Manipulations

The last step in the bootstrapping refinement scheme is the calculation of electron density maps and manual manipulations in graphical software. The electron density maps were calculated by Fast Fourier Transform of the structure factors obtained from the model using 2Fo-Fc and Fo-Fc coefficients. The program O version 6.2 was used to inspect how well the model fit in the density (Jones, *et al.*, 1991). In flexible regions where no apparent density is observed, an omit map, which is an electron density map calculated without information of the flexible regions, is used. This type of map is also used when ambiguous density is observed because an omit map is the real space equivalent of the Free R value (as described in section 2.6.5). A simulated annealing omit map will give an unbiased density of the problematic regions.

## 2.6.7 Bootstrapping Refinement Scheme

A general refinement scheme is shown in Figure 2.5. The first step in the bootstrapping refinement scheme is to generate a molecular structure file which gives XPLOR information on how the molecule is connected. The initial solution from molecular replacement has to go through several cycles of rigid body refinement until no further improvement of R-free is observed. This step is only necessary in the first round of refinement. It is excluded in subsequent refinement cycles. Then the weight described in section 2.6.2 is calculated, followed by several cycles of positional refinement and B-factor refinement. Similar to rigid body refinement, the cycle alternating positional refinement and B-factor refinement continues until no further drop in R-free is observed. Finally, an electron density map is computed and graphical manipulation proceeds. After regions of the model are better adjusted to the electron density, a new molecular structure file must be generated and the whole refinement scheme is repeated without rigid body refinement.



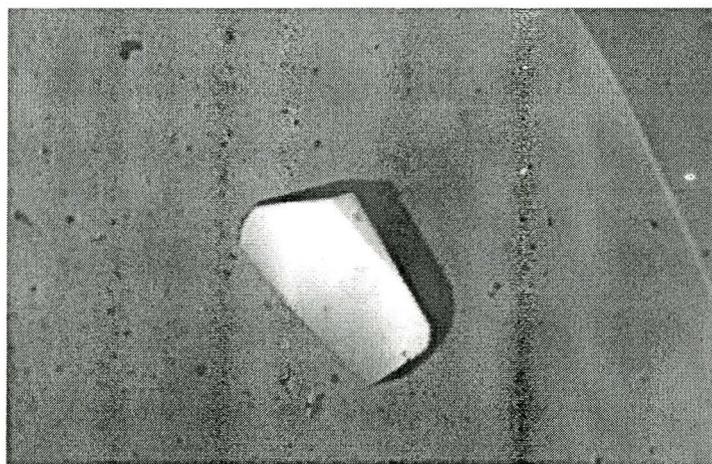
**Figure2.5** Bootstrapping refinement procedure.

## Chapter 3 Wild-type Lambda Lysozyme

### 3.1 Experimental Procedures

#### 3.1.1 Crystallization

WlaL crystals were grown using conventional hanging drop method. Small plate-like crystals appeared in a day from one of the 100 conditions in Hampton Research crystal screens. The condition contains 20% PEG 2000 MME, 0.01 M NiCl<sub>2</sub> hexahydrate and 0.1 M Tris at pH 8.5. Attempt to fine tune this condition followed. Bigger plate-like crystals with dimensions 0.5 x 0.4 x 0.1 mm were produced (Figure 3.1). Unfortunately, these crystals did not diffract better than 3.5 Å.



**Figure 3.1** Low power light microscope photograph. Crystals of wlaL grown from 20% PEG 400, 1 μM NiCl<sub>2</sub>, 0.1 M Tris pH 8.5 at room temperature. The crystal had dimensions approximately 0.6 x 0.4 x 0.2 mm.

### 3.2 Results and Discussion

WlaL crystals diffract very poorly to 3.5 Å using the X-ray machine at McMaster University (see chapter 2). Two data sets diffracted to 3.0 Å were collected at the CHESS beamline F2. Due to the poor quality of these native data sets, attempts to solve the native structure was not further

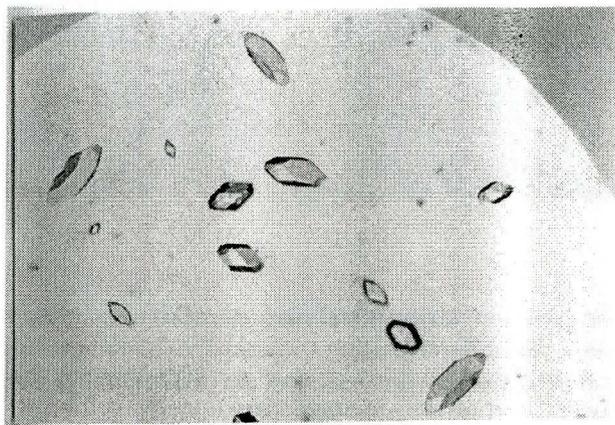
pursued. The comparison between the co-crystal structure and the uncomplexed structure were done using the mutant IaL structure solved by Evrard *et al.* (1998). The high degree of backbone similarity between the two structures and the ability of the mutant protein to function indicate that the wild type IaL and the mutant IaL closely resemble one another (see chapter 4).

## Chapter 4 Lambda Lysozyme and Chitohexaose (GlcNAc)<sub>6</sub>, complex, LaL•(GlcNAc)<sub>6</sub>

### 4.1 Experimental Procedures

#### 4.1.1 Crystallization

Crystal Screen kits I and II purchased from Hampton Research were used to find crystal conditions for laL•(GlcNAc)<sub>6</sub>. Of 100 solutions, laL•(GlcNAc)<sub>6</sub> crystals appeared in a wide variety of precipitants and salts. Through further fine screening, the mixture containing 0.1 M NH<sub>4</sub>SO<sub>4</sub>, 0.1 M NaOAc pH 4.7, and 20% (w/v) PEG 2000 MME was identified as the best condition for laL•(GlcNAc)<sub>6</sub> crystal growth. Needle-like, good quality crystals grown from this condition had dimensions around 0.5 x 0.2 x 0.2 mm (Figure 4.1).

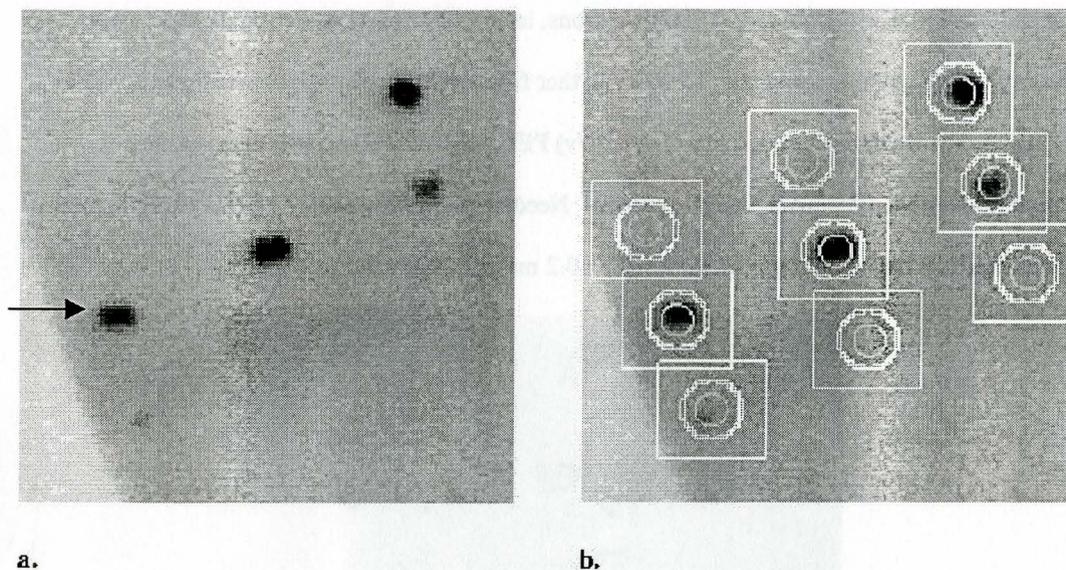


**Figure 4.1** Co-crystals of lambda lysozyme and hexaose inhibitor, (GlcNAc)<sub>6</sub>. The largest crystal in the picture has dimensions: 0.5 x 0.2 x 0.2 mm.

These crystals diffracted to at least 2.8 Å. One crystal was brought to the NSLS beamline X12-C and it diffracted to 2.3 Å. The diffraction patterns for these crystals were highly anisotropic. Diffraction along the c\* axis was usually about 0.5 Å worse. In addition, twinning was observed in all of these data. Twinning occurs when two crystal lattices grow on top of each other, producing duplicated diffraction spots in each hkl index. The twinning problem seemed to be

intrinsic to the crystallization condition. Attempts to solve the twinning problem using techniques such as microseeding, sitting drop vapour diffusion method, and application of the Hampton Research additive screen were used, but no satisfactory results were observed. Finally, one crystal grown using a dialysis button instead of the conventional hanging drop vapour diffusion method was twinned less severely so that further processing was possible. A data set from this crystal was collected at McMaster University (chapter 2).

#### 4.1.2 Data Collection and Processing



**Figure 4.2 Diffraction spots and integration boxes.** a) and b) were taken from the same section of the diffraction pattern. The diffraction spots correspond to resolution at about 3.8 Å. a). The arrow is pointing to a slightly twinned spot (see small dip in the middle). b). This picture shows the integration boundaries covering the widespread spots.

A data set for  $\text{IaL} \cdot (\text{GlcNAc})_6$  was collected as described in chapter 2. The data collection statistics are summarized in Table 4.1. Specifically, the crystal to detector distance was 166.0 mm. Each image was collected with a 1.5° oscillation angle. The crystal belongs to the orthorhombic space group  $p2_12_12_1$  with cell dimensions  $a=57.29$ ,  $b=61.08$ ,  $c=122.53$  Å. The mosaicity was 1.2°, which is quite reasonable considering the slight twinning observed in the data. During data processing, special attention was given to high resolution spots to ensure that the integration spot size was big enough to cover the widespread reflections caused by twinning

(Figure 4.2). Reflections up to 2.6 Å were processed; the overall  $R_{\text{sym}}$  was 11.8%. However in the highest resolution shell, the data was only 55.3% complete. Also, the error in intensity measurement was very high in the two highest resolution shells (2.6 & 2.7 Å). Realistically, the structure should be considered to have a resolution of 2.8 Å.

**Table 4.1** Data collection statistics for laL•(GlcNAc)<sub>6</sub> complex<sup>a</sup>.

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions (Å)	a=57.29, b=61.08, c=122.53
Highest resolution (Å)	2.6
Observed reflections	108652
Unique reflections	14138
$R_{\text{sym}}$ <sup>b</sup> (%)	11.8
Completeness (%)	92.0

<sup>a</sup> Data from this table was taken from the final scalepack output

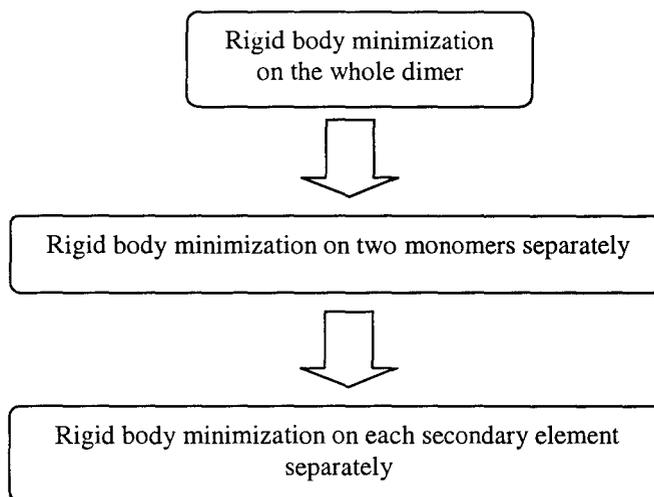
<sup>b</sup> For definition of  $R_{\text{sym}}$ , see section 2.4.4.

#### 4.1.3 Molecular Replacement

As mentioned in chapter 2, both rotation and translation searches were performed using XPLOR version 3.8. Monomer A from the mutated laL (mlaL) was used as the search model. Using a method proposed by Matthews, (Matthews, 1968), it was estimated that there are two monomers in an asymmetric unit. Therefore, the rotation search should find at least two peaks. Real space Patterson search method generated only one prominent peak after PC-refinement. At the same time, a more computationally intense direct rotation search was also done because it had the greatest signal-to-noise ratio compared to the other rotation search methods (Brünger, 1997). PC-refinement of the direct rotation search generated two top peaks approximately 267° from each other. The orientation with the highest peak was used in subsequent translation searches. After the position of one monomer was determined, the partial structure factors of this monomer was kept constant while a second translation search for the second monomer was performed on a lattice that covered the complete unit cell. One prominent peak was generated from the second translation search, thus an unambiguous solution was found.

#### 4.1.4 Crystallographic Refinement

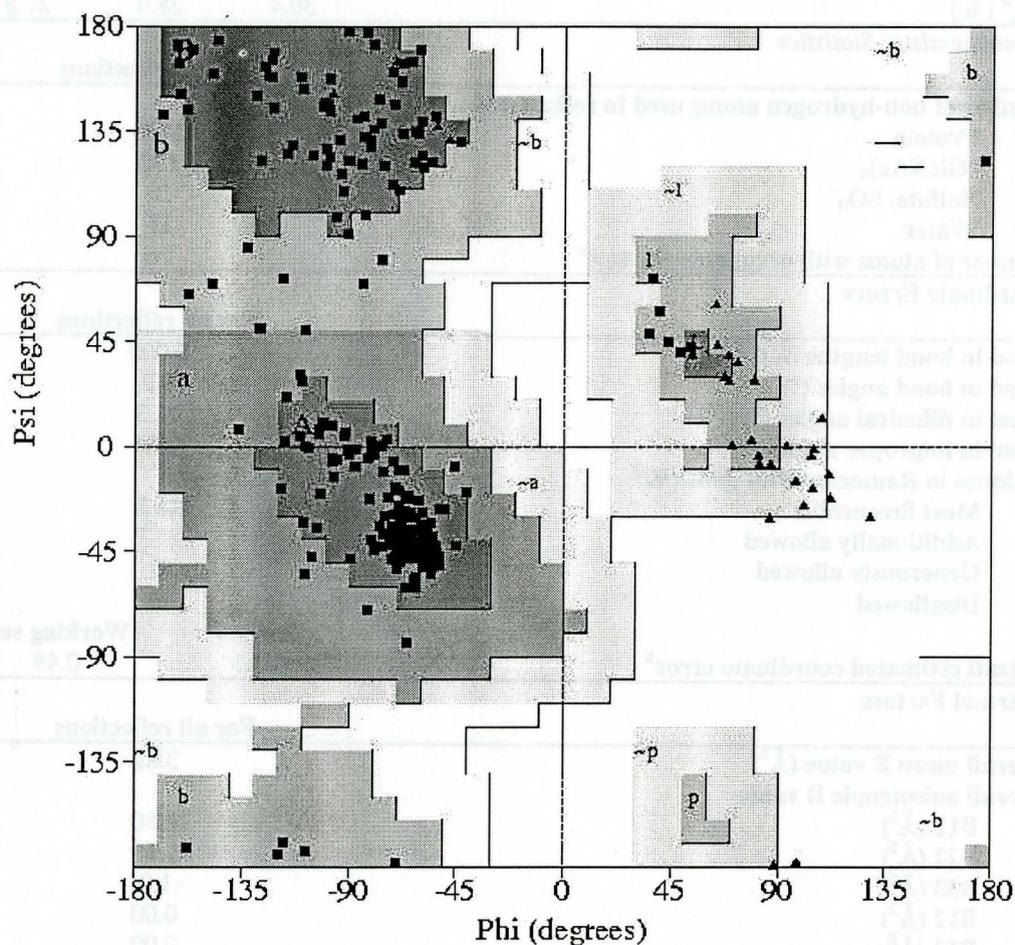
Rigid body refinement using the scheme shown in Figure 4.3 brought the R-free and R factor down to 0.423, 0.415 from 0.449, 0.439 respectively. The majority of the model fit nicely into density without manual intervention, except for two loop regions (residues 55-60 and residues



**Figure 4.3 Rigid body refinement scheme.**

129-138). These two regions corresponded to the same flexible regions where most variations were observed for each monomer in the trimeric mutant structure (Evrard, *et al.* 1998). Rough density for the inhibitor was observed near the active site quite early during refinement. The topology parameters of the inhibitor were based on small molecule crystallographic data. During early refinement, only reflections with  $F/\sigma F$  of at least 2.0 and those in the resolution range of 6–2.8 Å were used. During the initial rounds of positional refinement, one third of the suggested ideal weight was used in order to put more emphasis on the geometry of the structure. One half of the ideal weight was used when most of the structures had been fitted in density and values of  $E_{\text{chem}}$  were reasonably small. After positional refinement, restrained individual isotropic B-factor refinement was computed. A bulk solvent correction was also calculated after each cycle of B-factor refinement (see section 2.6.4). In addition, anisotropic correction was performed since the

diffraction pattern was highly anisotropic as discussed in section 4.1.1. Four sulfate anions and thirty-one water molecules were included in the final structure. Convergence was reached when R-free and R factor reached 0.286 and 0.214, respectively. The refinement and stereochemistry statistics of the final structure of  $1\alpha L \cdot (\text{GlcNAc})_6$  are summarized in Table 4.2.



**Figure 4.4 Ramachandran plot.** Plot of all phi psi angles in the final refined  $1\alpha L \cdot (\text{GlcNAc})_6$  dimeric structure. The shaded areas represent sterically allowed phi, psi combinations. The darker the shade, the more favourable the phi psi angles are. Phi psi angles for Gly are shown as triangles (all other residues are shown as squares). Some Gly phi psi angles fall in the disallowed regions because the side chain of Gly only consists of a hydrogen atom. As a result, it can adopt a wider range of conformations without any steric hindrance (Branden & Tooze, 1991).

**Table 4.2** Statistics of the final refined structure of laL•(GlcNAc)<sub>6</sub> at 2.6 Å resolution.

<i>Refinement Statistics</i>			
	F/σF=0 <sup>a</sup>	F/σF=1 <sup>a</sup>	F/σF=2 <sup>a</sup>
Number of reflections used for R factor calculations	11803	10883	9906
Resolution Range Å	40-2.6	40-2.6	40-2.6
Overall Completeness (%) <sup>b</sup>			
2.6 Å	92.0	84.3	76.7
2.7 Å	96.2	88.2	80.2
2.8 Å	97.2	89.1	81.1
R <sub>conventional</sub> <sup>c</sup> (%)	23.2	21.4	20.7
R <sub>free</sub> <sup>d</sup> (%)	30.4	28.6	27.8
<i>Stereochemistry Statistics</i>			
	For all reflections		
Number of non-hydrogen atoms used in refinement			
Protein		2450	
(GlcNAc) <sub>6</sub>		170	
Sulfate, SO <sub>4</sub> <sup>-</sup>		20	
Water		31	
Number of atoms with occupancy set to 0 <sup>e</sup>		76	
<i>Coordinate Errors</i>			
	For all reflections		
Rmsd in bond lengths (Å)		0.005	
Rmsd in bond angles (°)		1.2	
Rmsd in dihedral angles (°)		21.3	
Rmsd in improper angles (°)		0.96	
Residues in Ramachandran plot (%) <sup>f</sup>			
Most favourable		88.3	
Additionally allowed		11.7	
Generously allowed		0	
Disallowed		0	
		Test set <sup>g</sup>	Working set <sup>g</sup>
Luzzati estimated coordinate error <sup>h</sup>		0.61	0.48
<i>Thermal Factors</i>			
	For all reflections		
Overall mean B value (Å <sup>2</sup> )		32.2	
Overall anisotropic B value			
B11 (Å <sup>2</sup> )		0.50	
B22 (Å <sup>2</sup> )		0.51	
B33 (Å <sup>2</sup> )		-1.01	
B12 (Å <sup>2</sup> )		0.00	
B13 (Å <sup>2</sup> )		0.00	
B23 (Å <sup>2</sup> )		0.00	
Isotropic thermal model		Restrained	
Isotropic thermal factor restraints	RMS	SIGMA	
Main-Chain Bond (Å <sup>2</sup> )	4.31	1.50	
Main-Chain Angle (Å <sup>2</sup> )	6.40	2.00	
Side-Chain Bond (Å <sup>2</sup> )	6.70	2.00	
Side-Chain Angle (Å <sup>2</sup> )	9.23	2.50	

<sup>a</sup>F/σF=2 was used in the initial rounds of refinement. F/σF=1 was used in subsequent refinement cycles. The R factors in this table were taken from the output of a positional refinement on the final model using the corresponding F/σF cut off.

<sup>b</sup> The completeness was taken from the final output of scalepack

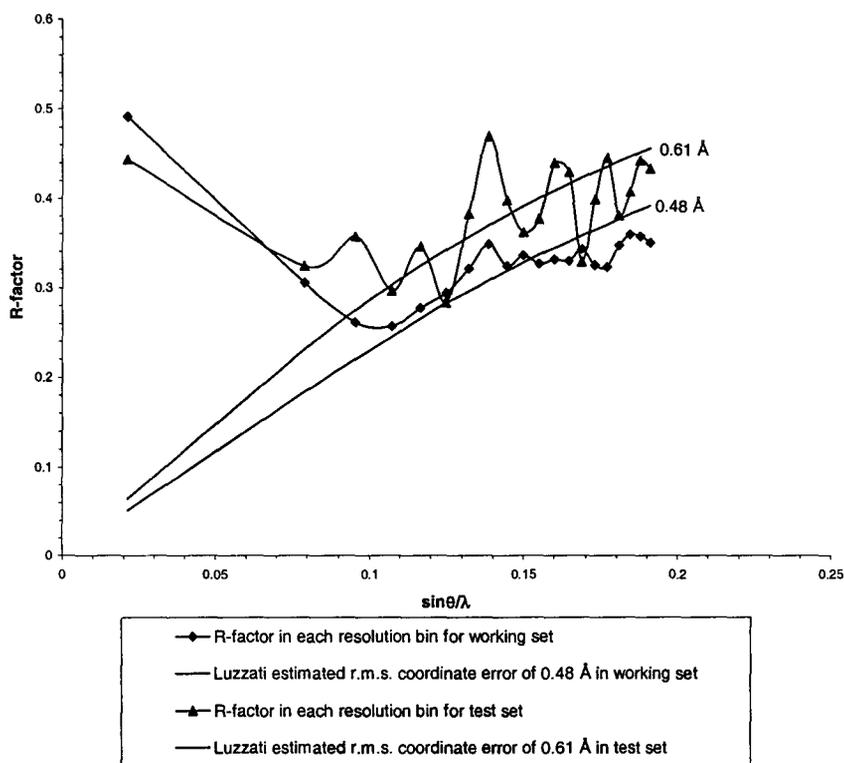
<sup>c, d</sup> For definition of  $R_{\text{conventional}}$  and  $R_{\text{free}}$ , see section (2.6.5)

<sup>e</sup> The occupancies of these atoms were set to zero because the density in these regions were either ambiguous or not present at all. See discussion for more details

<sup>f</sup> Values were calculated using PROCHECK (Laskowski, 1993). See also Figure 4.4 for graphical representation.

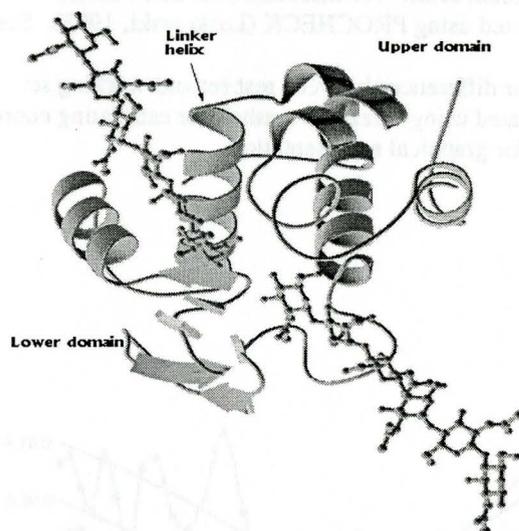
<sup>g</sup> See section 2.6.5 for differences between test set and working set

<sup>h</sup> Values were calculated using Luzzati's method for estimating coordinate errors (Luzzati, 1952). See also Figure 4.5 for graphical representation.



**Figure 4.5** Plot of R factors in 20 resolution bin in the final IaL•(GlcNAc)<sub>6</sub> structure and the corresponding plot of Luzzati error estimate. The plots for both the working set and the test set are drawn according to the definition in the legend. This plot shows a Luzzati estimated r.m.s. coordinate error of 0.48 Å and 0.61 Å for the working set and test set respectively.

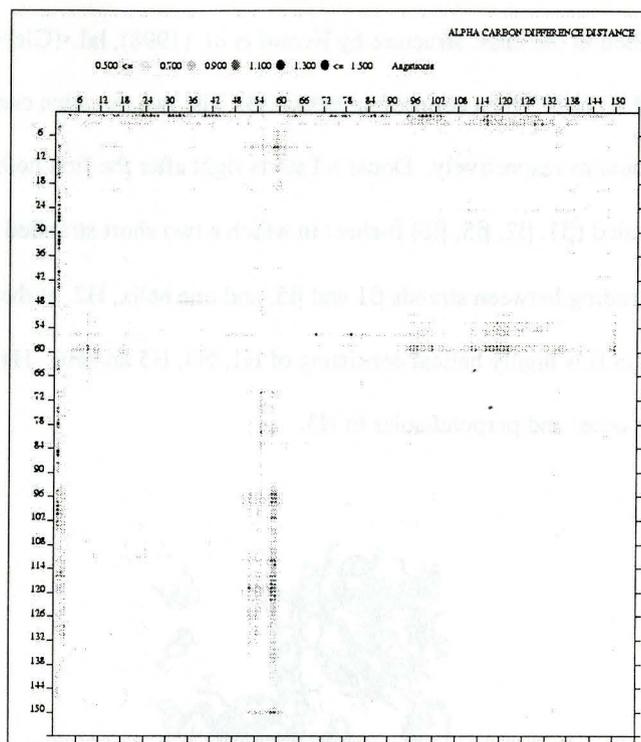
## 4.2 Results and Discussions



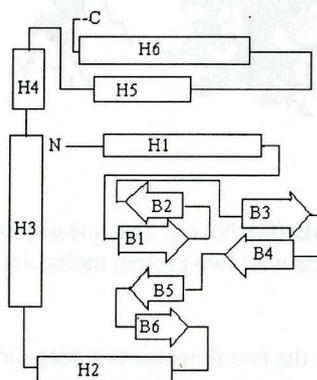
**Figure 4.6** The monomeric crystal structure of  $laL \cdot (GlcNAc)_6$ . The ball-and-stick model represents the  $(GlcNAc)_6$  inhibitor molecule. The upper and lower domains and the linker helix are labeled as shown. (This figure was produced using Molscript, (Kraulis, 1991), and Raster3d, (Bacon, & Anderson, 1988))

### 4.2.1 Overall Structure

The structure of  $laL \cdot (GlcNAc)_6$  will be discussed using conventions as close as possible to those used in describing the mutant  $laL$  ( $mlaL$ ) structure (Evrard, *et al.*, 1998). The protein packs as a dimer (monomers will be referred to as A and B) in the crystal (Figure 4.6). The superposition of the monomer backbones has a r.m.s.d of 0.381 and 0.870 when the side chains are included in the calculation. The difference distance matrix plot (DDMP) in Figure 4.7 shows that the major differences between the backbones of both monomers are located in regions from residues 50 to 60 and a few residues at the N-terminus. The two monomers are related to each other by eulerian angles  $(255.0, 3.7, 16.6)^\circ$  and a translation vector 185.7, -17.1, 35.4. Since the two monomers are almost identical, monomer A will be used for subsequent description unless stated otherwise. Figure 4.8 shows the topology diagram of  $laL \cdot (GlcNAc)_6$ . The secondary structures are labeled with reference to  $mlaL$ .

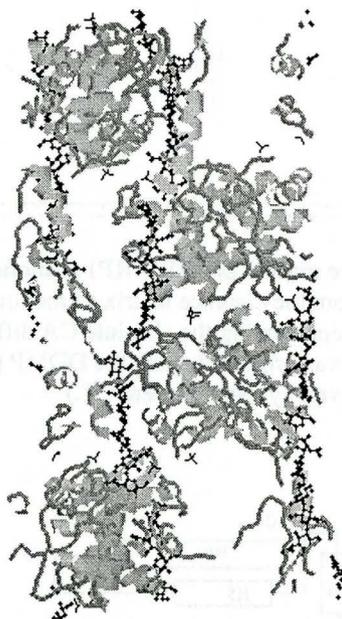


**Figure 4.7** Difference distance matrix plot (DDMP) of monomer A and B. The distance matrix of monomer A is subtracted from the distance matrix of monomer B. The delta values are plotted with different gray scale dots representing the absolute CA differences of 0.5 to 1.5 Å according to the scale at the top. (The plot was produced using the DDMP program from the Center for Structural Biology at Yale University, New Haven, CT.)



**Figure 4.8** Topology diagram for IaL. H1=(6-18); H2=(71-81); H3=(88-101); H4=(105-109); H5=(113-120); H6=(139-147);  $\beta$ 1=(34-37);  $\beta$ 2=(40-42);  $\beta$ 3=(52-55);  $\beta$ 4=(59-62);  $\beta$ 5=(63-65);  $\beta$ 6=(68-70). All the  $\beta$  strands and H2 correspond to the lower domain. H3 is the connecting helix. H1, H4-6 all correspond to the upper domain. Secondary structure elements were identified using the program DSSP (Kabsch & Sander, 1983).

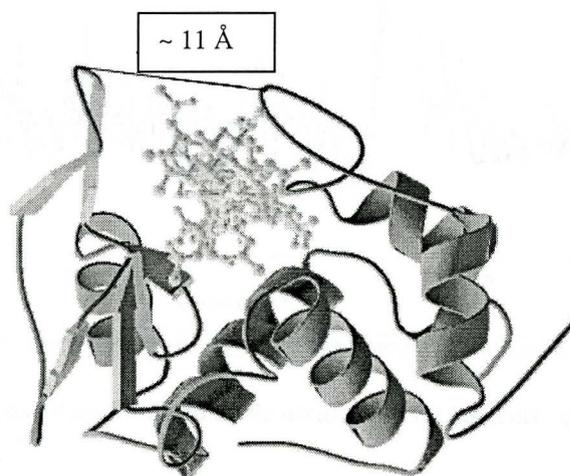
As was described in the *mlaL* structure by Evrard *et al.*, (1998),  $\text{laL}\cdot(\text{GlcNAc})_6$  is divided into two domains (I and II) linked by a connecting helix, H3. The two domains can also be viewed as upper and lower domains respectively. Domain I starts right after the first helix, H1. It consists of one four short stranded ( $\beta 1, \beta 2, \beta 5, \beta 6$ )  $\beta$ -sheet in which a two short stranded ( $\beta 3, \beta 4$ ) hairpin-like  $\beta$ -sheet is protruding between strands  $\beta 1$  and  $\beta 5$ , and one helix, H2, at the C-terminal end of the domain. Domain II is highly helical consisting of H1, H4, H5 and H6. H4 and H5 are antiparallel to each other and perpendicular to H3.



**Figure 4.9** The unit cell packing of  $\text{laL}\cdot(\text{GlcNAc})_6$ . The linearity of the inhibitor-binding mode is depicted here. Each inhibitor interacts with two protein molecules, linking each monomer in a linear fashion.

The active site located in between the two domains is a deep elongated cleft where the inhibitor binds to the protein in a linear fashion (Figure 4.6, 4.9). Each monomer interacts with four pyranose rings from one  $(\text{GlcNAc})_6$  molecule and two pyranose rings from the adjacent  $(\text{GlcNAc})_6$  molecule. The specific interactions will be discussed in following sections. The

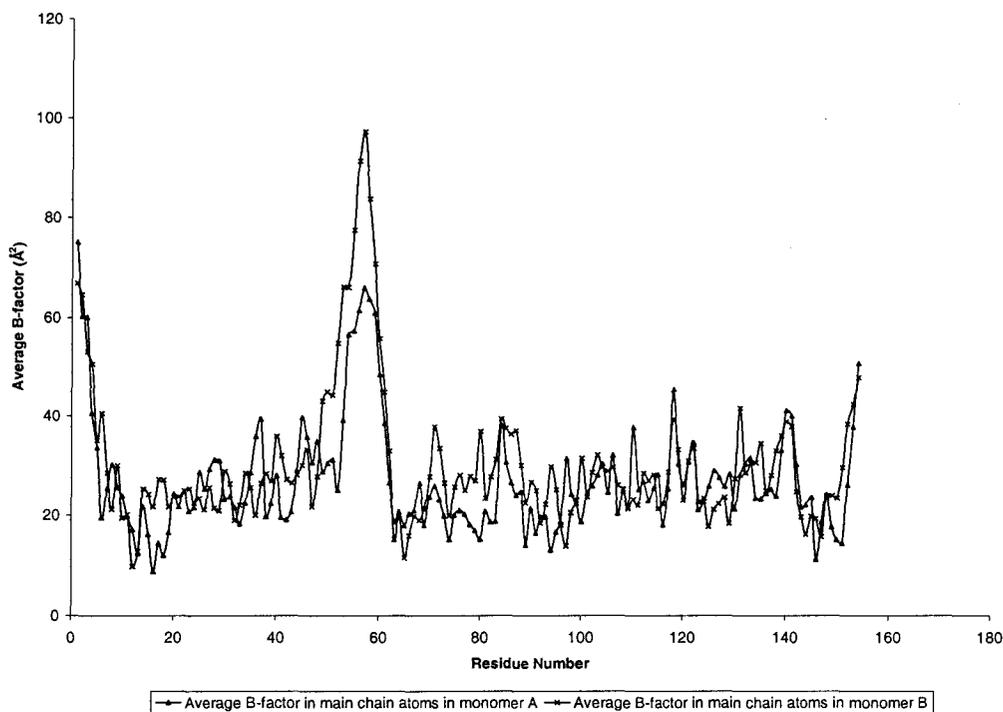
hairpin-like  $\beta$ -sheet wraps over the active site, holding the inhibitor in place. Together with the loop (residues 130 - 140) located on the other side of the active site, they form the lower lip and upper lip, respectively. In the trimeric mlal, both open and closed conformations were observed (Evrard, *et al.*, 1998). The open or closed conformation is defined from the distance between the upper and lower lips of the active site. In the open conformation, the distance between the two lips is approximately 23 Å, while in the closed conformation, this distance is reduced to around 10 Å. In both laL•(GlcNAc)<sub>6</sub> monomers, the distance between the upper and lower lips is approximately 11 Å, indicating a closed conformation (Figure 4.10).



**Figure 4.10** Side view of the active site showing the inhibitor binding to the cleft with the upper and lower lips closed.

Electron density is observed for almost all atoms in the final structure. The occupancies of 76 atoms are set to zero due to ambiguity or absence of electron density in these regions. Most of these atoms belong to the side chains of the loop regions exposed to the solvent. The N-terminal Met-1, Leu-55, Leu-59 and Lys-60 in monomer B are the only residues in which the occupancies for the main chain atoms are also set to zero. The thermal factors distribution for both monomers are similar (Figure 4.11). Consistent with the DDMP (Figure 4.7), the most flexible regions are

the N-terminus and residues 50 to 60, which correspond to the lower lip of the active site mentioned before.

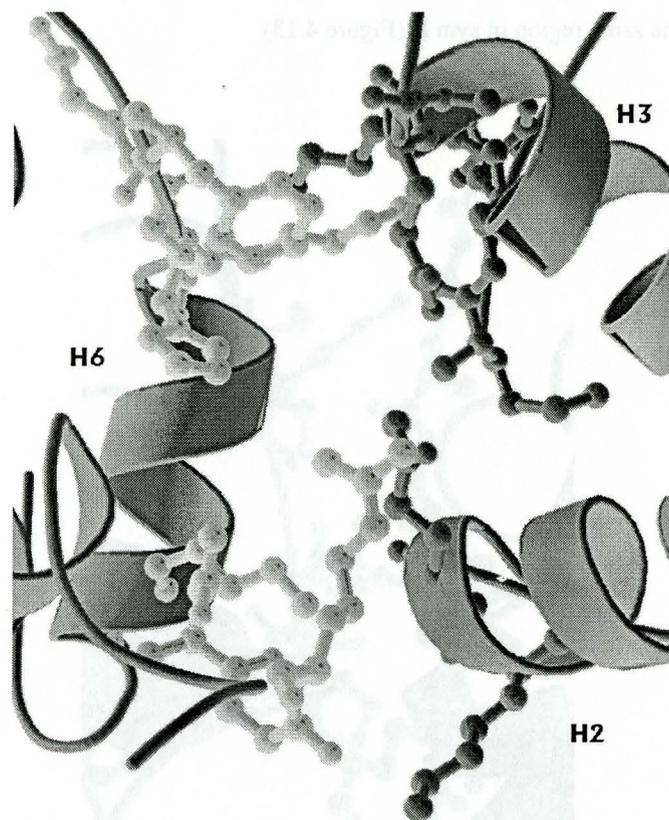


**Figure 4.11 Average thermal factor in main chain atoms in each residue.**

#### 4.2.2 Protein-Protein Interactions and Crystal Packing

The interactions holding the dimer together are a combination of hydrogen bonding and van der Waals forces. The inhibitor molecule and the sulfate anions also play a role in holding the dimer together by interacting with both monomers simultaneously. The major contributing forces seem to be the concurrent interactions with the anions and the inhibitors because only three hydrogen bonding interactions were present according to the algorithm used by INSIGHT II version 2.3.0 (1993). The interacting surface was relatively weak as well – the total solvent-accessible surface area occluded by the dimer formation was only 661 Å<sup>2</sup>. The major protein-protein interactions are as follows: 1) portion of the N-terminal loop (residues 1-3) and C-terminal end of H3 (residues 100-101) of monomer A interacts with the C-terminal end of the loop between

H5 and H6 (residues 135-138) in monomer B (this region is also called the upper lip of the active site) (Figure 4.12); 2) the C-terminal end of H2 (residues 79-80) in monomer A interacts with the middle portion of the loop between H1 and B1 (residues 23-25) and a small section in the middle part of H6 (residue 145) in monomer B (Figure 4.12).



**Figure 4.12 Protein-protein interactions between monomer A and B.** The ball-and-stick model represents the interacting side chains in the protein-protein interface. The darker shaded residues belong to monomer A while the lighter residues belong to monomer B. See text for more details.

Interestingly, the protein-protein interactions between monomer A and a symmetry-related (symmetry operator  $3, -x, y+1/2, -z+1/2$ ) monomer B (sym B) is much greater. Although only three hydrogen bonds were present, the total solvent-accessible surface area occluded by the symmetry interactions was  $1429 \text{ \AA}^2$ , more than twice the value of the dimer. Monomer A and sym B appeared to have the same orientation, but their backbones are opposite in directions. The major protein-protein interactions are as follows: 1) the N-terminal end of the loop between H1 and B1

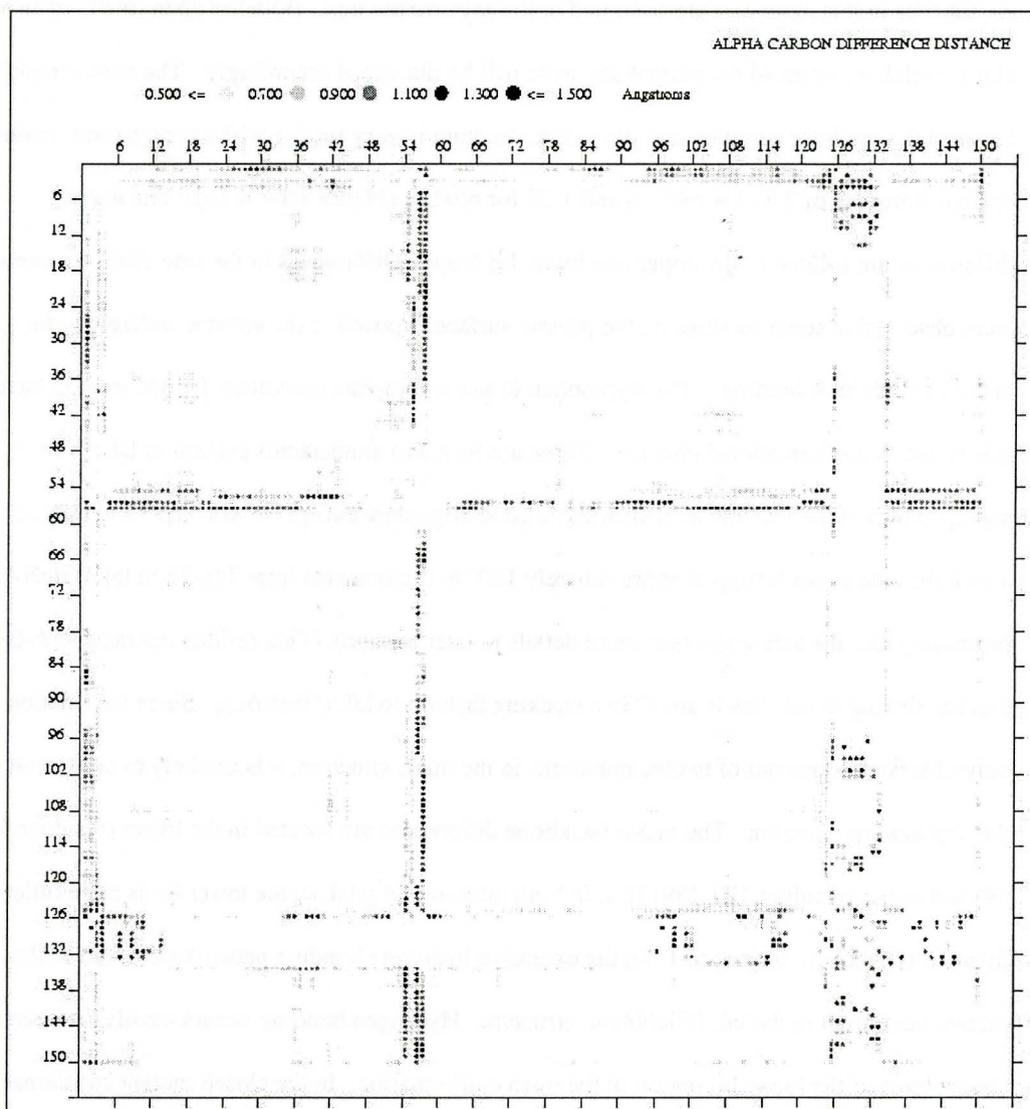
(residues 29-31) of monomer A interacts with the N-terminal end of H3 (residues 87-89) in sym B;  
2) the N-terminal end of H3 (residues 87-89) in monomer A interacts with the N-terminal end of the loop between H1 and B1 (residues 29-31) of sym B (this is the same interactions as in 1) but in opposite direction); 3) the middle portion of the C-terminal loop (residues 148-150) in monomer A interacts with the same region in sym B (Figure 4.13).



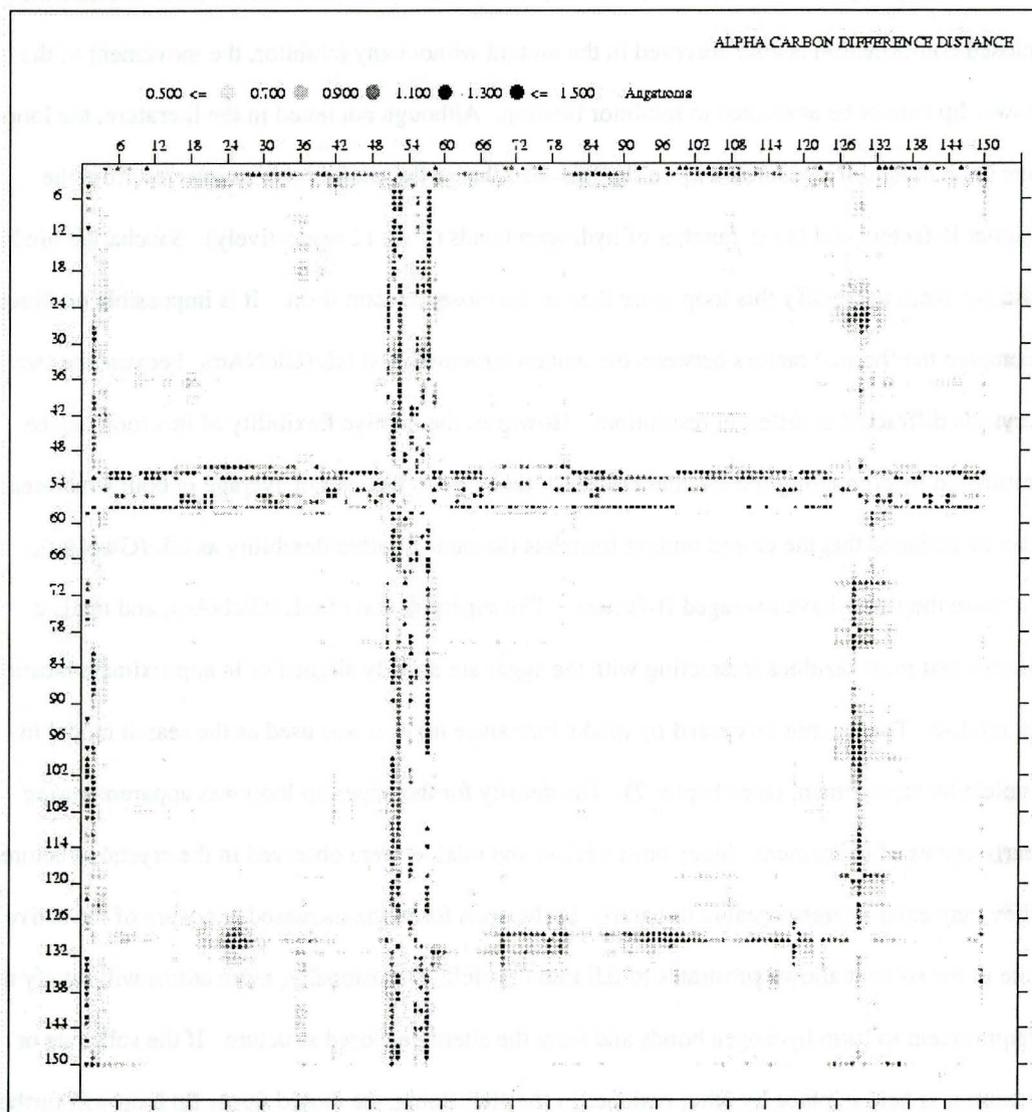
**Figure 4.13 Packing interactions between monomer A and symB.** The ball-and-stick model represents the interacting side chains in the protein-protein interface. The darker shaded residues belong to symB while the lighter ones belong to monomer A. See text for more details.

### 4.2.3 Overall Conformational Changes Upon Substrate Binding

Monomer B was used to compare the structure between  $\text{laL}\cdot(\text{GlcNAc})_6$  and  $\text{mlaL}$  because all six subsites in this monomer are occupied in the asymmetric unit. Both the open ( $\text{mlaL-o}$ ) and the closed ( $\text{mlaL-c}$ ) forms of the mutant structure will be discussed accordingly. The conformation between the inhibitor complex and the native structure is very similar with average r.m.s. main-chain differences of 2.39 for  $\text{mlaL-o}$  and 1.27 for  $\text{mlaL-c}$  (Figure 4.14, 4.15). The major differences are located in the upper and lower lip loops. Differences in the side chain rotamers were observed in some residues on the protein surface exposed to the solvent, unlikely to be caused by inhibitor binding. The tryptophan to aza-tryptophan mutations did not seem to have caused much conformational changes. There are four Trp amino acids present in  $\text{laL}$ . The superposition of the four mutated sites show good alignment except for aza-Trp-73 in  $\text{mlaL-c}$  in which the side chain is flipped approximately  $120^\circ$  anti-clockwise from Trp-73 in  $\text{laL}\cdot(\text{GlcNAc})_6$ , protruding into the active site (see more details in later section). This residue interacts with the saccharide ring in subsites B and C in a stacking fashion in  $\text{laL}\cdot(\text{GlcNAc})_6$ . Since the rotation is only observed in one out of twelve mutations in the  $\text{mlaL}$  structure, it is unlikely to be a result of the Trp-azaTrp mutation. The major backbone differences are located in the lower (residues 51-59) and upper (residues 127-139) lips. In both  $\text{mlaL-o}$  and  $\text{mlaL-c}$ , the lower lip is more ordered than  $\text{laL}\cdot(\text{GlcNAc})_6$  as judged from the extensive hydrogen bonding network observed in the mutant but absent in the  $\text{laL}\cdot(\text{GlcNAc})_6$  structure. Hydrogen bonding occurs mostly between main chain atoms of the lower lip region in the open conformation. In the closed mutant conformation, several ordered water molecules are present in the region. In contrast, the lower lip loop displays extremely weak electron density in the  $\text{laL}\cdot(\text{GlcNAc})_6$  dimer. It is difficult to assess whether the increased flexibility is caused by inhibitor binding to accommodate the large cell wall upon interacting to peptidoglycan or due to worse crystal packing in  $\text{laL}\cdot(\text{GlcNAc})_6$  since the crystal diffracts only to  $2.6 \text{ \AA}$  compared to  $2.3 \text{ \AA}$  in the mutant structure.



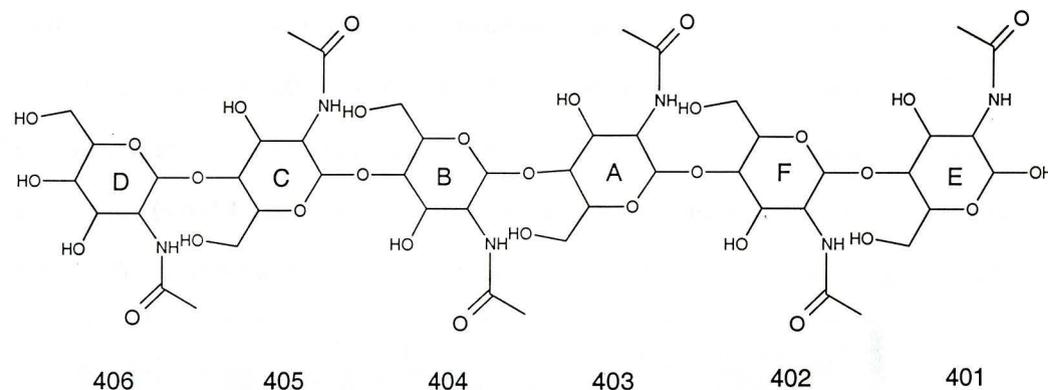
**Figure 4.14** Difference distance matrix plot (DDMP) of mlaL-o and monomer B of laL-(GlcNAc)<sub>6</sub>. See legend in Figure 4.7. The plot shows flexibility of the same region as depicted in Figure 4.7. The differences are much more prominent here as indicated by the darker colour of the dots. (The plot was produced using the DDMP program from the Center for Structural Biology at Yale University, New Haven, CT.)



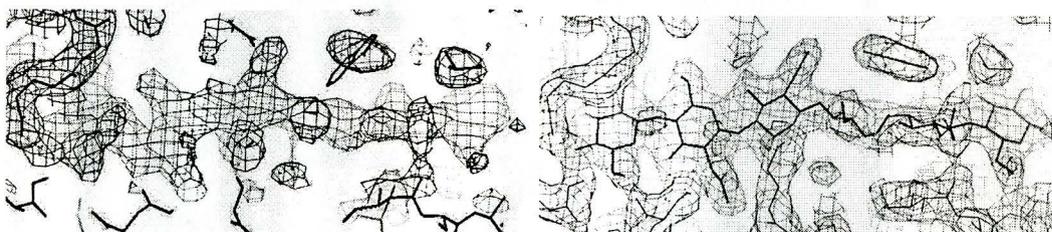
**Figure 4.15** Difference distance matrix plot (DDMP) of mlaL-c and monomer B of laL-(GlcNAc)<sub>6</sub>. See legend in Figure 4.7. This figure also shows flexibility of the same regions depicted in Figure 4.7 and 4.15. Judging from the number of black dots in the lower lip regions, there seems to be more differences between mlaL-c and laL-(GlcNAc)<sub>6</sub> than between mlaL-o and laL-(GlcNAc)<sub>6</sub>. (The plot was produced using the DDMP program from the Center for Structural Biology at Yale University, New Haven, CT.)

The upper lip loop is much more ordered than the lower lip loop as evaluated from the well-defined density and averaged thermal factor. As mentioned before, the distance between the upper and lower lip suggests that the laL·(GlcNAc)<sub>6</sub> structure is in a closed conformation. Since the closed conformation is also observed in the mutant without any inhibitor, the movement of the lower lip cannot be attributed to inhibitor binding. Although not noted in the literature, the loop in the open mutant form is much more flexible than that in the closed form as inferred from the higher B-factors and fewer number of hydrogen bonds (7 vs. 12 respectively). Saccharide binding did not seem to rigidify this loop more than in the closed mutant form. It is impossible to directly compare the thermal factors between the mutant structures and laL·(GlcNAc)<sub>6</sub> because the two crystals diffracted at different resolutions. However, the relative flexibility of this loop can be estimated by comparing the thermal factors of this loop to the overall average in both structures. It can be deduced that the closed mutant form has the same relative flexibility as laL·(GlcNAc)<sub>6</sub> because they both have averaged B-factors. The superposition of laL·(GlcNAc)<sub>6</sub> and mlaL-c shows that most residues interacting with the sugar are already aligned or in approximate location in mlaL-c. This cannot be caused by model bias since mlaL-o was used as the search model in molecular replacement (see chapter 2). The density for the upper lip loop was apparent during early cycles of refinement. Since both mlaL-c and mlaL-o were observed in the crystal structure, they may exist interchangeably in nature. In the open form, the increased exposure of the active site to the solvent allows substrates to fall into the cleft. Occasionally, more atoms will satisfy the requirement to form hydrogen bonds and form the alternate closed structure. If the substrate or inhibitor is held in place by other residues in the cleft lining, the closed upper lip loop will further stabilize the sugar by interacting with it. If no substrate or inhibitor is present, the relatively weak hydrogen bonding network can be disrupted by the solvent, changing it back to the open form.

## 4.2.4 Structure of Chitohexaose Inhibitor

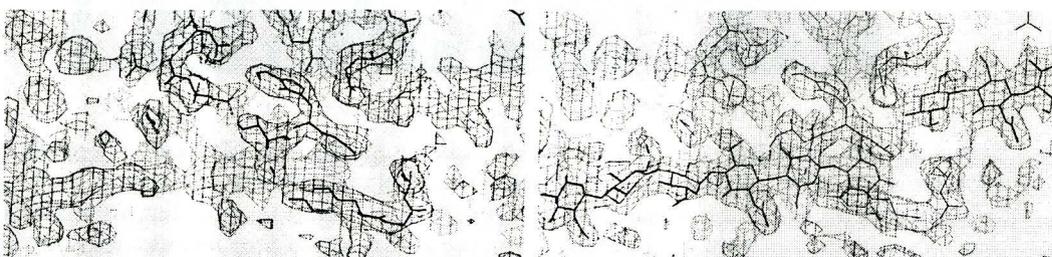


(a).



(b).

(c).

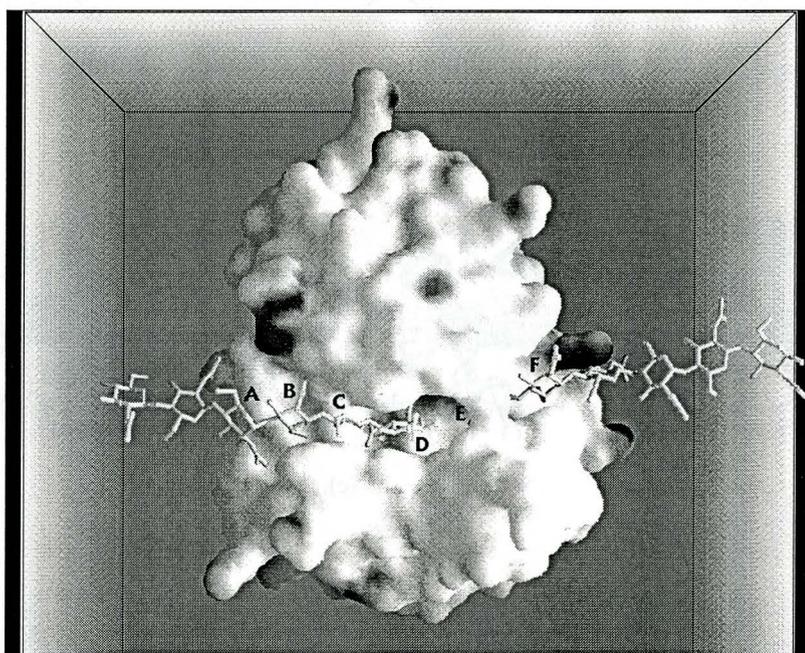


(d).

(e).

**Figure 4.16 Structure and density of chitohexaose inhibitor.** (a). Structure of  $(\text{GlcNAc})_6$  used in co-crystallization. The letters represent the equivalent sugar rings in the Phillips models. See text for more explanation. (b),(c). Density of the inhibitor binding mostly to monomer A (GlcNAc-1). (b) shows apparent density before modeling in the inhibitor and (c) shows the final density at the end of refinement. (d),(e). Density of the inhibitor binding mostly to monomer B (GlcNAc-2). (b) shows apparent density before modeling in the inhibitor and (c) shows the final density at the end of refinement. Three rings from the adjacent GlcNAc1 molecule are also shown here.

Chitohexaose has six GlcNAc moieties linked by a  $\beta(1\rightarrow4)$  glycosidic bond. LaL does not accept chitooligosaccharide as a substrate but the sugar compound can act as an inhibitor to laL activity (Duewel, *et al.*, 1995). Differential scanning calorimetry studies show that the binding of the inhibitor interacts with laL and stabilizes it (Duewel, *et al.*, 1995). Residue numbers 401 to 406 were assigned to each inhibitor, with each number representing a sugar ring (Figure 4.16a). The density of the GlcNAc appeared in early refinement cycles (Figure 4.16b-e). The orientation of each ring is confirmed by calculating a simulated annealing omit map from the final model.



**Figure 4.17** Surface depiction of monomer B with two hexasaccharides bound (GRASP; Nicholls *et al.*, 1991). Four residues from one molecule bind to subsites A to D while two residues from the adjacent molecule bind to subsites E and F. Note that the cleft is much closer than those observed in lysozymes (Chapter 1). Residue E is buried deep into the cleft, making extensive interactions with the enzyme.

The binding of  $(\text{GlcNAc})_6$  is shared between two protein molecules throughout the crystal (Figure 4.17). GlcNAc residues A to D interact with one monomer and residues E and F interact with the adjacent monomer. From other lysozyme/GlcNAc models, the saccharide binding subsites are designated A to F with the glycosidic bond cleavage occurring between subsites D and E (Chipman, *et al.*, 1969). Observing the atomic interactions at the active site, it was immediately

**Table 4.3 Summary of Subsite Interactions.**

Inhibitor Ring	Corresponding Subsite	Polar contacts <sup>a</sup>	Total van der Waals contacts <sup>b</sup>	Average thermal factor ( $\text{\AA}^2$ )
GlcNAc-1:406	D	(O-6) — NH 125	18	28.42
GlcNAc-1:405	C	(N-2) — CO 123	32	17.21
		(O-6) — NH 70		
		(O-3) — Gln 98		
GlcNAc-1:404	B	(O-7) — Asn 122	13	27.32
		(O-6) — Tyr 77		
GlcNAc-1:403	A	(N-2) — Arg 25	15	24.23
		(O-3) — Arg 25		
GlcNAc-1:402	F	None	13	17.06
GlcNAc-1:401	E	(N-2) — CO 19	39	15.41
		(O-7) — NH 38		
		(O-3) — Glu 19		
		(O-3) — Gln 68		
		(O-6) — CO 135		
GlcNAc-2:406	D	(O-6) — NH 125	19	41.89
GlcNAc-2:405	C	(N-2) — CO 123	31	23.21
		(O-7) — NH 70		
		(O-3) — Gln 98		
		(O-5) — Tyr 132		
GlcNAc-2:404	B	(O-7) — Asn 122	20	26.50
		(O-4) — Glu 101		
		(O-6) — Tyr 77		
		(O-6) — Gln 98		
GlcNAc-2:403	A	None	12	40.04
GlcNAc-2:402	F	None	12	21.95
GlcNAc-2:401	E	(N-2) — CO 19	35	21.81
		(O-7) — NH 38		
		(O-3) — Gln 68		
		(O-4) — Glu 19		
		(O-6) — CO 135		

<sup>a</sup>Polar contacts are those at less than 3.5 Å between hydrogen bonding atoms, as calculated from atomic positions in the crystallographic model.

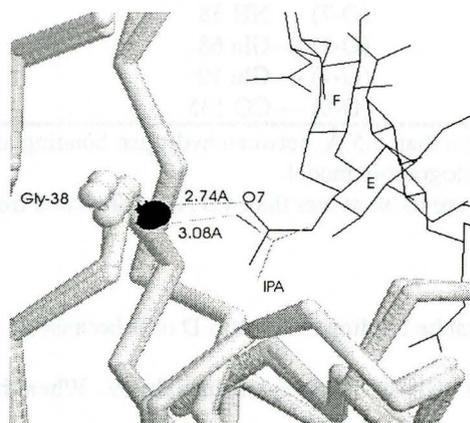
<sup>b</sup>The van der Waals contacts are those at less than 4.0 Å as calculated from atomic positions in the crystallographic model.

realized that GlcNAc-406 must be binding to subsites D or E because its glycosidic oxygen, O4, is in hydrogen bonding distance with the catalytic residue Glu-19. When the helix containing the catalytic residue of several known lysozyme/GlcNAc complexes were superimposed with that of 1aL, the GlcNAc rings were well aligned with those of 1aL. It was apparent from the superposition that GlcNAc-403, 404, 405, 406 occupy sites A, B, C, and D respectively. The 3-OH groups of GlcNAc-404, 406, 402 (B, D, and F) are situated such that it is sterically possible to link a peptide.

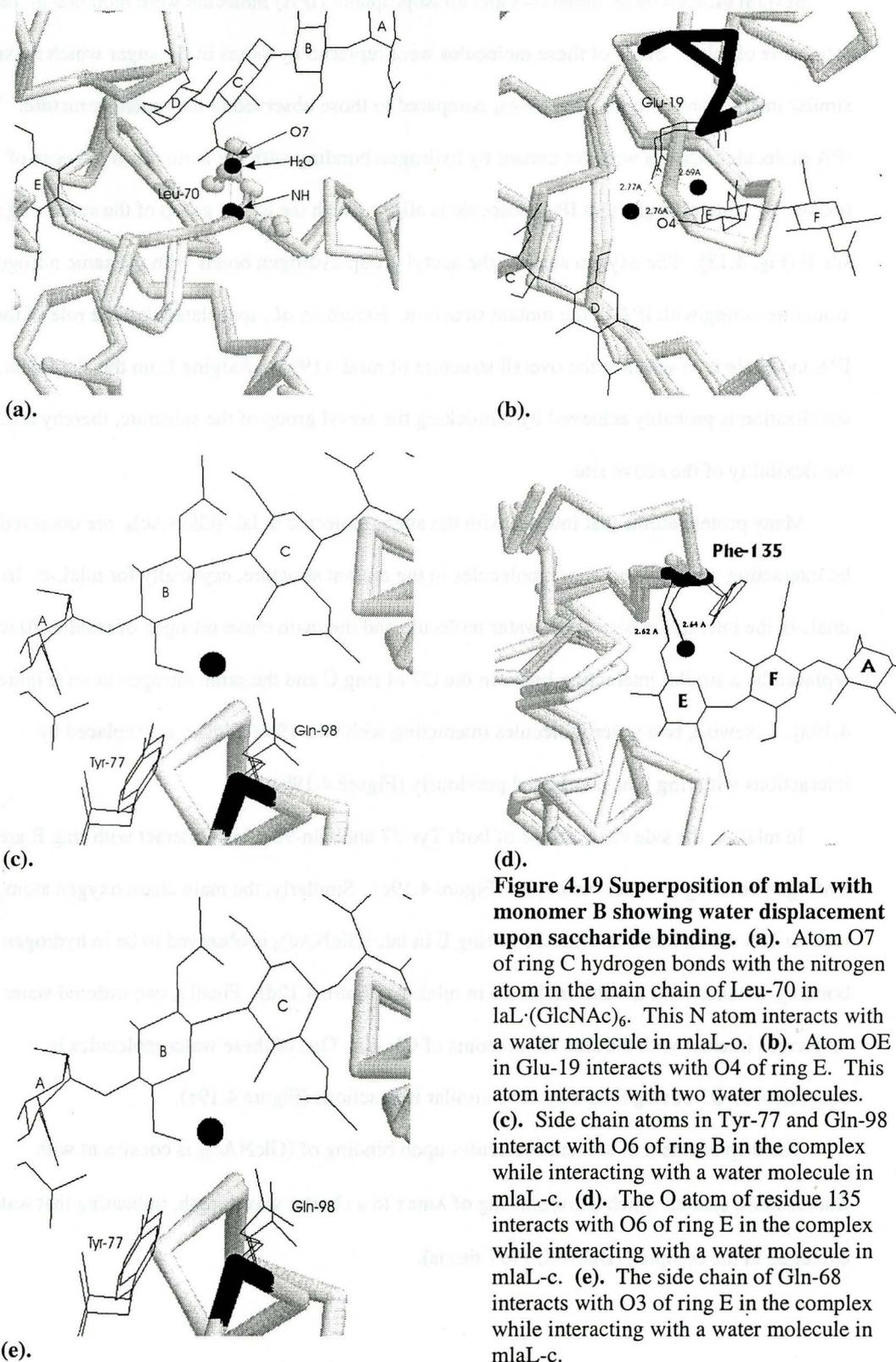
This observation further supports the binding site assignments because N-acetylmuramic acid (NAM) in peptidoglycan can only bind subsites B, D, and F as shown in other lysozyme/GlcNAc models (Phillips, 1966; Blade *et al.*, 1965, 1967a,b).

The average thermal factors for GlcNAc-1 and GlcNAc-2 are  $21.61 \text{ \AA}^2$  and  $29.23 \text{ \AA}^2$  respectively. The individual thermal factor for each ring indicates that rings C, E, and F bind most tightly to the protein. Stability in ring C is consistent with previous observations that this residue is very important to substrate binding interactions (Weaver, *et al.*, 1995; Strynadka & James, 1991). Currently, only HuL, with sites E and F occupied, is available for comparison. The relative thermal factors for E and F compared to other sugar residues in HuL are higher than that in laL, indicating stronger binding of these two rings in laL (Song, *et al.*, 1994). The polar and van der Waals contacts for both inhibitors to the corresponding monomer is similar but not identical (Table 4.3). This indicates that the binding is not highly specific, which is not too surprising because  $(\text{GlcNAc})_6$  is not a substrate for laL.

#### 4.25 Changes in Solvent Structure in the Active Site Upon Inhibitor Binding



**Figure 4.18** Superimposed active site of mlaL and laL· $(\text{GlcNAc})_6$ . As shown in this figure, the oxygen atom from the IPA molecule interacts with the same nitrogen atom (residue 38) the oxygen of the acetyl group in ring E is interacting.



Several ordered water molecules and an isopropanol (IPA) molecule were modeled in the active site of mlaL. Some of these molecules were replaced by atoms in the sugar which make similar interactions with laL·(GlcNAc)<sub>6</sub> compared to those observed in the mutant structure. The IPA molecule interacts with the mutant by hydrogen bonding with the main chain nitrogen of residue 38. Interestingly, this IPA molecule is aligned with the acetyl group of the sugar ring in site E (Fig. 4.18). The oxygen atom in the acetyl group hydrogen bonds with the same nitrogen atom interacting with IPA in the mutant structure. Evrard *et. al.*, speculated that the role of the IPA molecule is to stabilize the overall structure of mlaL (1997). Judging from the alignment, the stabilization is probably achieved by mimicking the acetyl group of the substrate, thereby reducing the flexibility of the active site.

Many protein atoms that interact with the sugar molecule in laL·(GlcNAc)<sub>6</sub> are observed to be interacting with ordered water molecules in the mutant structure, especially for mlaL-c. In mlaL-o, the interaction between a water molecule and the main chain nitrogen of residue 70 is replaced by a similar interaction between the O7 of ring C and the same nitrogen atom (Figure 4.19a). Likewise, two water molecules interacting with Glu-19 in mlaL-o are replaced by interactions with ring E as mentioned previously (Figure 4.19b).

In mlaL-c, the side chain atoms of both Tyr-77 and Gln-98 which interact with ring B are hydrogen bonding to water molecules (Figure 4.19c). Similarly, the main chain oxygen atom in residue 135 which interacts with O6 of ring E in laL·(GlcNAc)<sub>6</sub> is observed to be in hydrogen bonding distance with a water molecule in mlaL-c (Figure 4.19d). Finally, two ordered water molecules interact with the side chain atoms of Gln-68. One of these water molecules is equivalent to O3 of ring E as they have similar interactions (Figure 4.19e).

The displacement of solvent molecules upon binding of (GlcNAc)<sub>6</sub> is consistent with fluorescence studies which show shifting of  $\lambda_{max}$  to a shorter wavelength, indicating that water is excluded in the complex (Dewel, 1997 thesis).

#### 4.2.6 Interactions of Sulfate Anion to IaL

Four sulfate anions were modeled in consideration of the appearance of some puzzling large densities in regions surrounded by positive side chains. Since IaL is a basic protein, it is not surprising to see anions participating in the crystallization process. The addition of either phosphate or sulfate anions were investigated since both anions were present in the crystallizing drop. Due to similar chemical structure, it was difficult to distinguish the two anions solely by the shape of the density. Sulfate anions were chosen on grounds of its higher concentration in the crystal (6 mM phosphate vs. 100 mM sulfate).

**Table 4.4 Summary of sulfate anion interactions**

Sulfate Anion	Polar contacts <sup>a</sup>	Total van der Waals contacts <sup>b</sup>	Average thermal factors ( $\text{\AA}^2$ )
SO <sub>4</sub> <sup>-</sup> 200	(O-1) — Lys 138 (O-1) — NH 3 (O-3) — Lys 138 (O-4) — Arg 8	13	94.49
SO <sub>4</sub> <sup>-</sup> 201	(O-1) — NH 138 (O-2) — Gln 97 (O-3) — His 137 (O-3) — Ser 141 (O-4) — Ser 141	11	39.29
SO <sub>4</sub> <sup>-</sup> 202	(O-1) — Lys 100 (O-2) — Gln 97 (O-2) — His 137 (O-3) — Ser 141 (O-3) — NH 138 (O-3) — His 137 (O-4) — Ser 141	27	37.89
SO <sub>4</sub> <sup>-</sup> 203	(O-1) — Lys 138 (O-2) — Arg 8 (O-3) — Arg 8 (O-4) — NH 3 (O-4) — Arg 8 (O-4) — CO 3	6	81.88

<sup>a</sup>Polar contacts are those at less than 3.5 Å between hydrogen bonding atoms, as calculated from atomic positions in the crystallographic model.

<sup>b</sup>The van der Waals contacts are those at less than 4.0 Å as calculated from atomic positions in the crystallographic model.

Table 4.4 summarizes the interactions between the sulfate anions and the protein. The four negative sulfate anions are surrounded by positive lysine residues. Besides residue 200 that interacts solely with monomer A, the other sulfate molecules interact with two protein molecules. Residues 201 and 203 are simultaneously in contact with monomer A and a symmetry-related molecule; residue 202 interacts with both monomers A and B. The presence of residues 200 and 203 are questionable as they both have extremely high temperature factors. A simulated annealing omit map calculated at the end of refinement showed density at the anion positions. As a result, residues 200 and 203 are kept in the model despite their high temperature factors.

#### 4.2.7 Sugar-Binding Mode in laL Compared to Other Lysozyme/Ligand and Slt70/Bulcegin Complexes

As it has already been established that the overall fold of laL is homologous to c and v type lysozyme (Evrard, *et. al.*, 1998), the comparison in this section will emphasize the sugar-binding modes in laL and other lysozyme/ligand complexes. Moreover, the recent elucidation of the Slt70 structure reveals a resemblance of its C-terminal domain to basic lysozyme folds (Thunnissen, *et. al.*, 1994, 1995a, 1995b). LaL was superimposed with the Slt70/bulcegin in order to identify any similar features in the active site since both enzymes produce the same anhydro end product. Also included in the end of this section is a discussion of some interesting similarities and differences between several residues of laL and other homologous structures.

As explained in the introduction, it is generally accepted that the mechanism of lysozyme action involves a Glu residue acting as a general acid in the reaction. Moreover, this catalytic residue is located at the C-terminal end of a helix in all forms of lysozyme. The negative dipole moment of the C-terminal end of the helix makes it more favourable for the catalytic residue to donate a proton to the glycosidic oxygen. Based on this information, the helix containing the catalytic residue in laL and various lysozymes were superimposed with the catalytic residue fixed at one end.

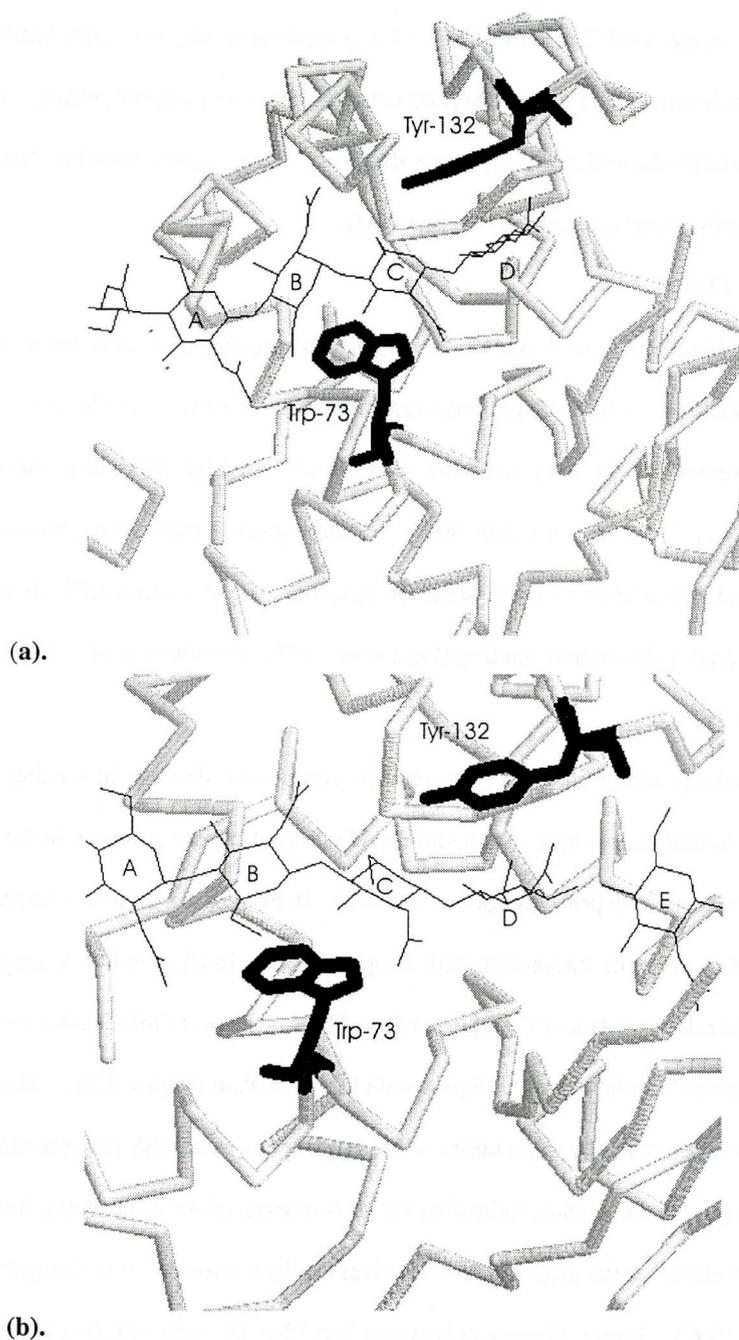
Coordinates of the c-type lysozymes, HEWL/(GlcNAc)<sub>4</sub> (1LZC) and HuL/(GlcNAc)<sub>6</sub> (1LZS), g-type lysozyme, GEWL/(GlcNAc)<sub>3</sub> (154L), p-type lysozyme, mT4L/(NAM-NAG) (148L) and SlT70/bulgecin (1SLY) were overlaid with monomer b of laL·(GlcNAc)<sub>6</sub>. The superimposed structures showed relatively good alignment of the ligands bound to their corresponding protein, indicating similar binding mode.

#### *Subsite A (residue 1137)*

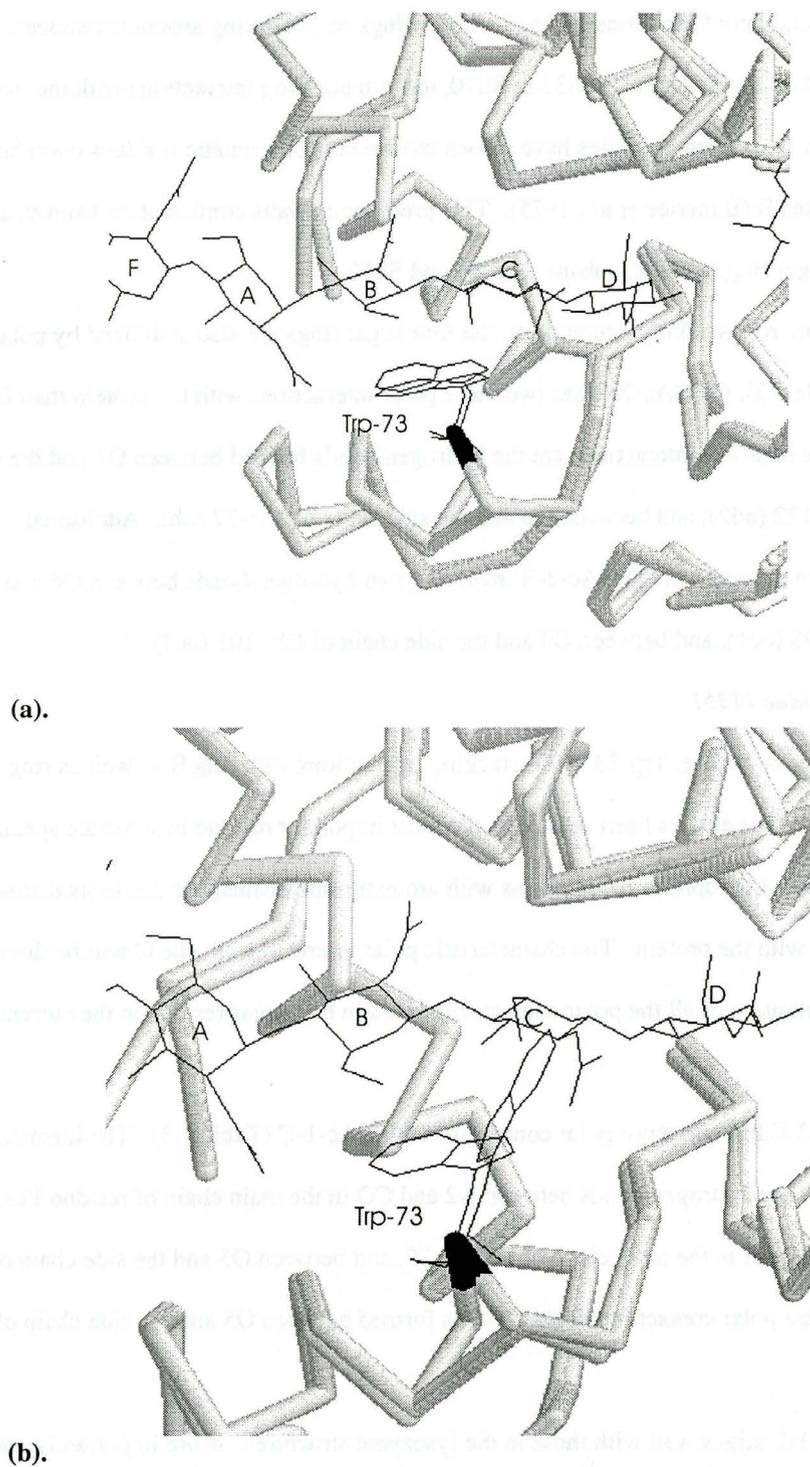
The sugar ring binding to site A is relatively flexible due to its exposure to the solvent at the end of the cleft. GlcNAc1-A has two polar contacts with Arg-25 while GlcNAc-2-A has none at all. The lack of interaction with the protein is consistent with the high B-factor in this sugar, especially for GlcNAc-2-A where it equals 40.04 Å<sup>2</sup>. Subsites A in other lysozyme structures also show the same weak interactions to the protein. In fact, this residue is often difficult to model due to low occupancy and/or disordered binding (Ford *et al.*, 1974; Cheetham *et al.*, 1992).

#### *Subsite B (residue 1136)*

Extensive stacking interactions are observed between saccharide rings in subsites B to D and aromatic residues in both monomers. Such van der Waals contacts are common in the recognition of carbohydrate molecules by proteins (Quioco, 1986). In most cases, an aromatic residue makes partial or face-to-face stacking interaction with a sugar residue. In laL, the indole ring of Trp-73 is aligned with saccharide rings B and C (Figure 4.20). The side chain of this residue from mlaL-c protrudes into subsite C while it is well aligned with laL in mlaL-o (Figure 4.21). The movement of Trp-73 closer to the active site upon inhibitor binding is consistent with fluorescence experiments which show a blueshift, indicating the environment of the dominating fluorophore (Trp) has changed either by the displacement of solvent or by conformational changes in the protein (Dewel, 1997). Similar alignment between Trp-73 in laL with ring B is also present in HuL, HEWL and GEWL where the equivalent residues involved in the stacking interactions are Tyr-63, Trp-62 and Phe-123, respectively. In contrast, no residue capable of forming stacking interactions in the subsite B of T4 or SlT70 can be identified. Since both T4 and SlT structures used for superposition lack a sugar ring binding to subsite B, it is possible that the binding of ring B



**Figure 4.20 (a),(b) Stacking interactions in rings B to D.** Highlighted in these two figures are the residues participating in stacking interactions from saccharides B to D. Trp-73 interacts with rings B and C while Tyr-132 stacks with ring D. Both (a) and (b) depict the same figure in a different orientation to show the perfectly parallel stacking between Tyr-132 and the pyranose ring D.



**Figure 4.21 Superposition of mlaL and laL·(GlcNAc)<sub>6</sub> showing Trp movement upon saccharide binding.** (a). Superposition of mlaL-o and laL·(GlcNAc)<sub>6</sub>. The side chain of Trp-73 in mlaL-o aligns well with that of laL·(GlcNAc)<sub>6</sub>. (b). Superposition of mlaL-c and laL·(GlcNAc)<sub>6</sub>. The side chain of Trp-73 in mlaL-c protrudes into subsite C.

can induce a slight conformational change which brings neighbouring aromatic residues, such as Phe-104 in T4 or Tyr-552 and Tyr-533 in Slf70, to form stacking interactions with the pyranose ring. Previous fluorescence studies have shown movements of aromatic residues upon binding of substrate in sites B (Banerjee *et al.*, 1975). This prediction awaits confirmation from structural data with a sugar ring bound to subsite B in T4 and Slf70.

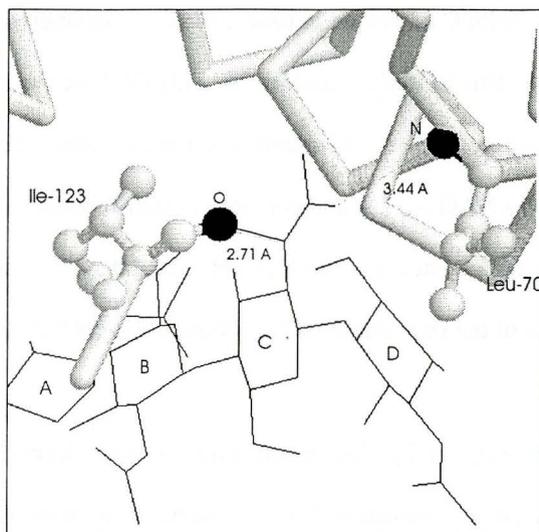
Apart from ring stacking interactions, the four sugar rings are also stabilized by polar contacts (Table 4.3). GlcNAc-2-B has two more polar interactions with the protein than GlcNAc-1-B does. The identical interactions are the hydrogen bonds formed between O7 and the side chain of Asn-122 (nd2), and between O6 and the side chain of Tyr-77 (oh). Additional interactions are observed in GlcNAc-2-B involving two hydrogen bonds between O6 and the side chain of Gln-98 (oe1), and between O4 and the side chain of Glu-101 (oe1).

#### *Subsite C (residue 1135)*

As mentioned before, Trp-73 forms stacking interactions with ring B as well as ring C. However, ring C has always been viewed as the most important residue in substrate specificity, not because of its hydrophobic interactions with aromatic side chains, but due to its distinctive polar contacts with the protein. The characteristic polar interactions in site C will be discussed after a brief summary of all the polar contacts observed in this sugar residue in the current structure.

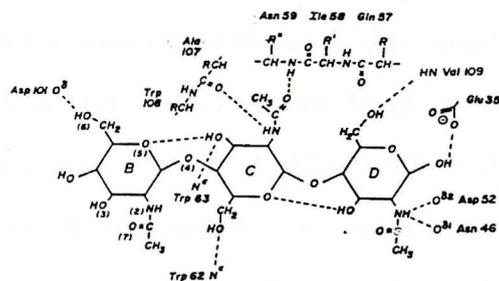
GlcNAc-2-C has one more polar contact than GlcNAc-1-C (Table 4.3). The identical interactions are the hydrogen bonds between N2 and CO in the main chain of residue 123, between O7 and NH in the main chain of residue 70, and between O3 and the side chain of Gln-98 (oe1). The extra polar contact in GlcNAc-2-C is formed between O5 and the side chain of Tyr-132 (oh).

C ring in laL aligns well with those in the lysozyme structures. More importantly, the characteristic hydrogen bonding network in the 2-acetamido group, where it is located between two backbone segments of the protein, and forms a pair of bridging hydrogen bonds observed in HEWL, T4, GEWL, HuL, and Slf70, is also present in laL (Fig. 4.22) (Weaver, *et al.*, 1995). The

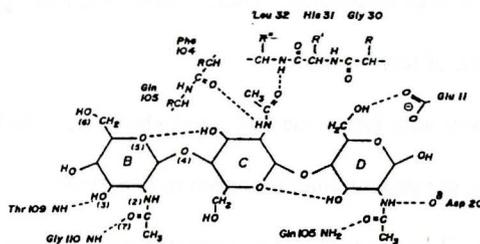


(a).

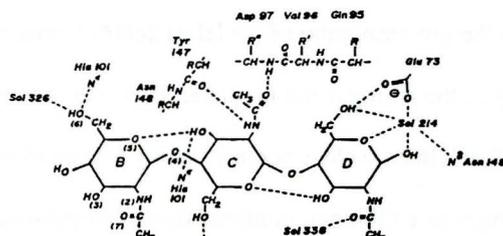
(A) HEWL



(B) T4L



(C) GEWL



(b).

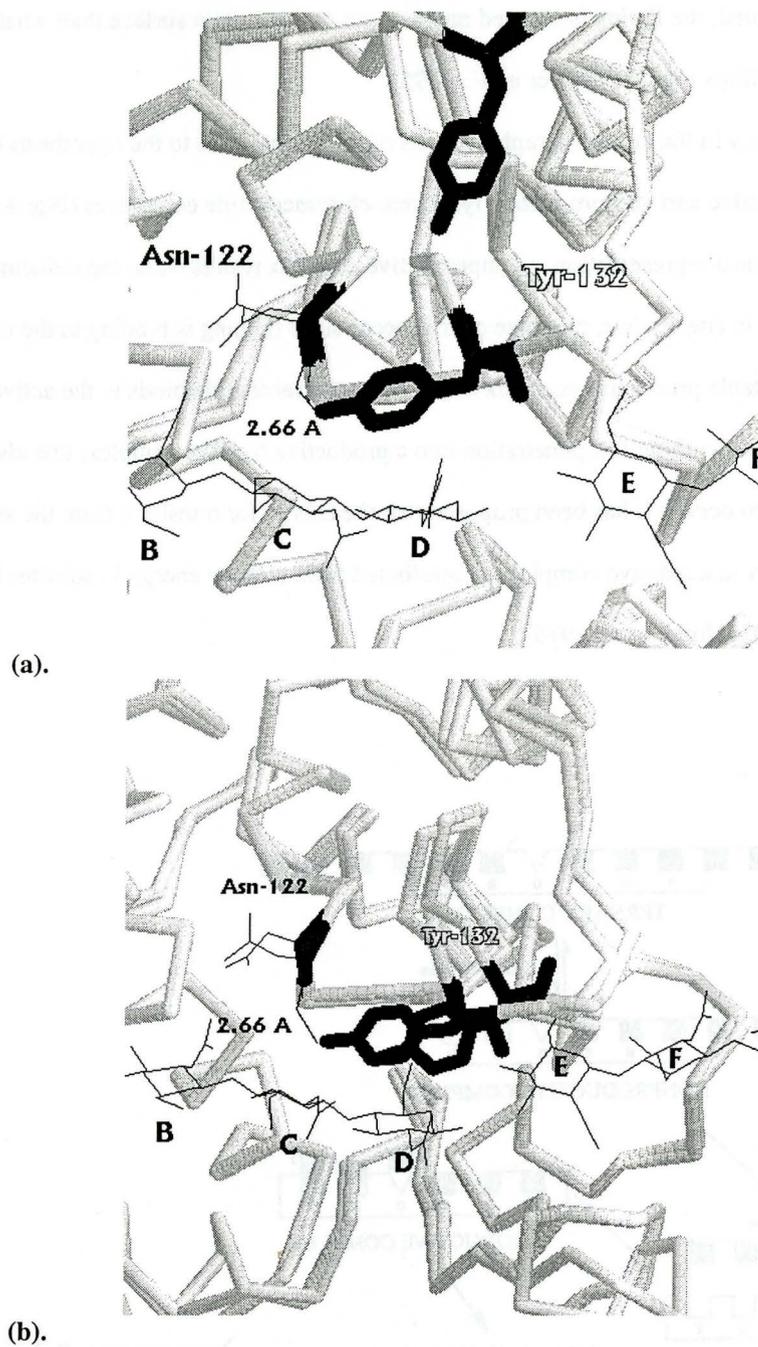
**Figure 4.22 Characteristic hydrogen bonding network in site C.** (a). The N atom in the 2-acetamido group of ring C interacts with main chain O atom of residue 123 while the O atom interacts with the main chain N atom of residue 70. Figure (b) shows the same interactions at the 2-acetamido group. Although not shown here, HuL and Slf70 also have these distinctive interactions (Weaver, *et al.*, 1995).

conserved interaction in subsite C is thought to have importance in determining the cleavage specificity of the enzymes. Due to steric consideration, only GlcNAc can bind to this site, thus, a NAM residue must bind to site D. Consequently, the cleavage site is always the glycosidic bond between a NAM and a NAG residue and not vice versa (Thunnissen, *et. al.*, 1995b). The inability of maltotetraose and cellotetraose (both of which lack the 2-acetamido group) to inhibit laL gives further evidence of the importance of this group for (GlcNAc)<sub>n</sub> binding (Duewel, 1997).

#### *Subsite D (residue 1134)*

Similar to Trp-63, the rings of Tyr-132 in both monomers also form stacking interactions with the D ring. However, the alignment of Tyr-132 and ring D in monomer B is not as parallel as the one in monomer A. This residue is held in place by hydrogen bonding to the carbonyl oxygen of Asn-122 (Figure 4.23). Compared to other residues in the active site, Tyr-132 shows a large movement from the ligand-free mutant, especially for mlaL-o in which the residue moves approximately 8 Å (Figure 4.23). Such an observation is not surprising since this residue is located at the flexible upper lip. In contrast to the other sugar rings, ring D forms the same polar interactions in both monomers. There is only one hydrogen bond observed between O6 and NH of the main chain of residue 125.

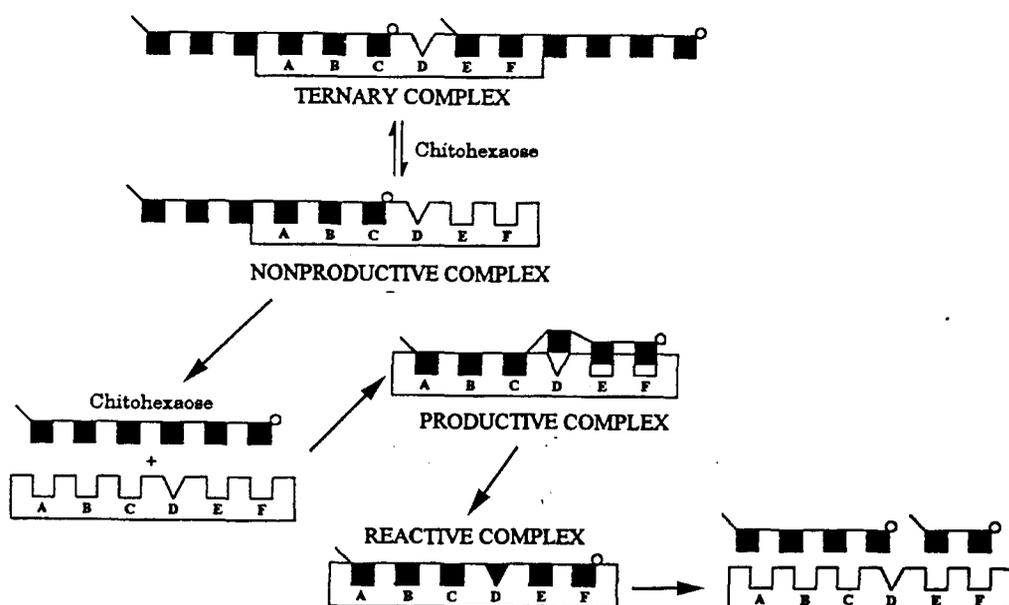
Alignment with lysozyme structures shows that the D ring in laL is located much closer to the surface of the protein than that seen in lysozyme. The lack of interactions between this sugar ring and Glu-19 suggests that the sugar ring has not penetrated deep enough into the cleft (Table 4.3). Due to the low resolution of the laL·(GlcNAc)<sub>6</sub> structure, a more elaborate interpretation of the geometry of the D ring is not possible. However, given the fact that this ring is quite remote from the cleft and that no close contacts to any protein atom is observed, it is reasonable to believe that the D ring is in a full-chair conformation in the present structure. In fact, distortion in the D ring is only observed crystallographically when this ring is penetrated deep into the cleft (Kuroki, *et al.*, 1993; Kelly, *et al.*, 1979). Similar to laL·(GlcNAc)<sub>6</sub>, the full chair conformation of the D ring has been observed in many structures where this ring is located far away from the active site (Pincus & Scheraga, 1979). For example, in the GEWL/(GlcNAc)<sub>3</sub> complex, in which subsites B,



**Figure 4.23 Superposition of mlaL and laL·(GlcNAc)<sub>6</sub> showing Tyr-132 movement upon saccharide binding.** (a). Superposition of mlaL-o and laL·(GlcNAc)<sub>6</sub>. The Tyr-132 situated at the flexible upper lip moves more than 8 Å upon inhibitor binding in mlaL-o. (b). Superposition of mlaL-c and laL·(GlcNAc)<sub>6</sub>. The displacement is not as pronounced in mlaL-c as in mlaL-o. The Tyr-132 is stabilized by hydrogen bonding with the carbonyl O atom of Asn-122 in the complex structure.

C, and D are occupied, the D ring is situated much closer to the protein surface than what has been proposed in the Phillips model (Weaver *et al.*, 1995).

The discrepancy in the crystallographic results of the D ring leads to the hypothesis of the existence of productive and non-productive lysozyme-chitosaccharide complexes (Fig. 4.24). As shown in the graphical representation, an unproductive complex results when the reducing end of the substrate binds to site C; thus, cleavage cannot occur since nothing is binding to the cleavage sites D and E. A stable productive complex forms when the saccharide binds to the active site across the whole cleft without full penetration into a productive reactive complex; this also prevents cleavage to occur. It has been proposed that the energy for transition from the stable productive complex to a reactive complex is contributed from binding energy in subsites E and F (Weaver *et al.*, 1995; Holler *et al.* 1975).



**Figure 4.24** Productive vs. non-productive lysozyme-chitosaccharide complexes (Holler, *et al.*, 1975). See text for details.

*Subsites E and F (residues 401-402)*

While sugar binding to subsites A to D has been experimentally observed in other lysozyme structures, saccharide interactions to subsites E and F have never been seen crystallographically. Although six sugar rings are observed to bind to HuL, the (GlcNAc)<sub>2</sub> does not bind directly to subsites E and F but is trapped near the sites (Song, *et. al.*, 1994). Since the O4 of ring E is in hydrogen bonding distance with Glu-19 in laL, it is likely that ring E and ring F observed in laL·(GlcNAc)<sub>6</sub> correspond closest to the physiological substrate/protein complex in that region. These rings have the lowest thermal factors compared to other rings in both monomers, suggesting strong binding to the protein. Distinguished from subsites B to D, E and F have no stacking interactions with the protein. However, ring E does have the most van der Waals contacts compared to other rings in both inhibitor molecules. Similar to rings A to C, the number and the type of polar interactions in ring E is not the same (Table 4.3). The identical interactions occur between N2 and CO of the main chain of residue 19, between O7 and NH of the main chain of residue 38, between O3 and the side chain of Gln-68 (oe1), and between O6 and CO of the main chain of residue 135. Only one different interaction is observed in this site. The side chain of the catalytic residue, Glu-19 interacts with atom O3 in GlcNAc-1-E while it interacts with atom O4 in GlcNAc-2-E.

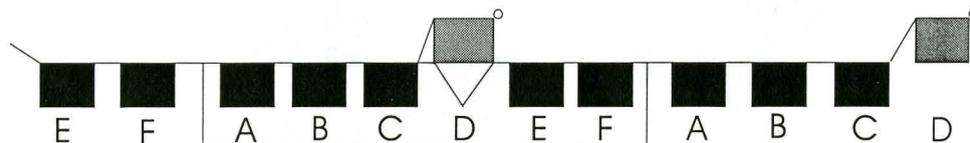
The extent of the E ring interaction to the protein is comparable to that of ring C, which has been observed in other lysozyme structures to have the most interactions with the protein (Table 4.3). Located at the opposite end of the active site, the binding mode of ring F is similar to ring A in that there are very few interactions with the protein. No polar contact to the protein is observed in this subsite for either monomer. Although ring F seems to be loosely bound to the protein as suggested by the lack of polar interactions and limited van der Waals contacts, its low thermal factor indicates otherwise. It is possible that the strong binding in ring E stabilizes ring F. Since no lysozyme structures with rings E and F occupied have been observed, it is likely that the

disaccharide does not interact with the protein too strongly. The weak interaction allows the disaccharide to diffuse away quickly, so that a water molecule can diffuse into the active site and attack the C1 atom of the oxocarbenium ion. The unexpected stable binding of rings E and F in laL may point to the possibility that the disaccharide stays in the active site longer than in other lysozymes to prevent water molecules from attacking the C1 atom of the tetrasaccharide so that transglycosylation is favourable. The observation that the cleft of laL is about 2 Å narrower than those in lysozymes further supports this hypothesis. Moreover, the surface depiction in Figure 4.17 shows that residue E is buried deep inside the cleft, which makes it more likely to reside in the active site longer as suggested. This hypothesis can be supported by future mutagenesis experiments, which will decrease the available interactions in subsite E, followed by the detection of reducible products. Finally, fluorescence studies have shown that di- and trisaccharides do not cause the blueshift observed with larger saccharides. Since sites E and F lack tryptophan residues, it is possible that smaller saccharides prefer binding to subsites E and F.

Conformational energy calculations have predicted two binding modes termed “left-sided” and “right-sided” for hexasaccharides in HEWL (Pincus, *et al.*, 1977). Experimental evidence shows that the “right-sided” conformation predominates at equilibrium (Smith-Gill, *et al.*, 1984). Basically, the binding mode is determined by which side of the protein rings E and F are interacting. Superimposing laL·(GlcNAc)<sub>6</sub> and HEWL using a similar scheme as that mentioned before shows that rings E and F in laL do not tilt to either side of the protein but are located at the centre of the active site. Except for ring D that is tilted out to the protein surface, the rest of the saccharide units bind in more of a linear manner than what is predicted in HEWL.

Figure 4.25 gives a graphical representation of what the binding mode of (GlcNAc)<sub>6</sub> in laL might be. It is possible that (GlcNAc)<sub>6</sub> is unable to bind all the way across the active site physiologically, thus acting as a competitive inhibitor but not a substrate to laL (Dewel, 1997). As will be discussed in later sections, the binding of the peptide moiety to the protein might be necessary to bring the D ring closer to the cleft to induce distortion. Without the binding energy

provided by peptide interactions, it is not surprising that the GlcNAc residue in the D site would prefer positioning closer to the surface to avoid distortion.



**Figure 4.25** Graphical representation of proposed binding mode of (GlcNAc)<sub>6</sub> in laL.

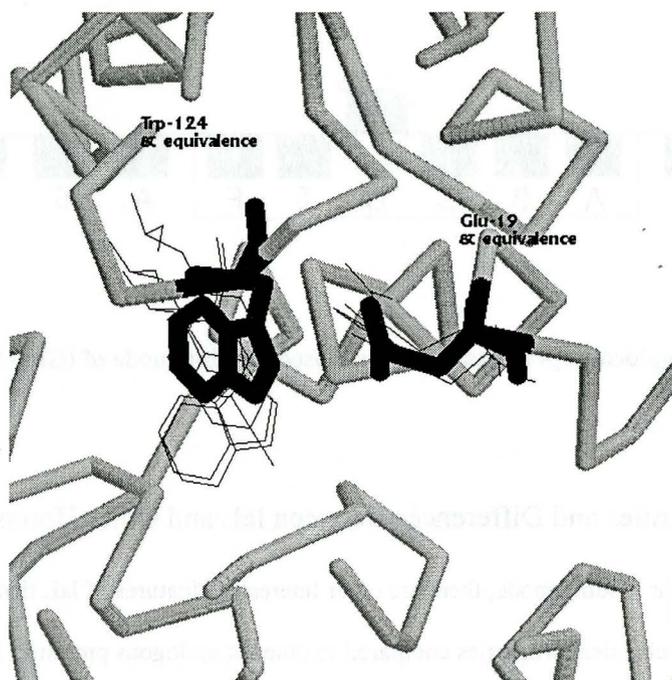
#### 4.2.8 Similarities and Differences Between laL and Other Homologous Proteins

Besides sugar binding mode, there are other interesting features of laL that contribute to its characteristic biochemical properties compared to other homologous proteins. In this section, several aspects of the laL structure that have implications on the distinctive biochemical properties of laL and/or on the mechanism of enzyme action will be discussed.

##### *Trp-124*

Fluorescence studies have shown close to identical emission maxima of laL and HEWL which suggests the average environment of the fluorescent tryptophan residues are very similar in both proteins (Duewel, 1997 thesis). This observation led to the identification of equivalent tryptophan residues in laL based on sequence comparison with HEWL (Duewel, 1997 thesis). The sequence alignment shows strong evidence that Trp73 and Trp74 in laL are equivalent to Trp62 and Trp63 in HEWL, which was subsequently confirmed by our structural data. Trp124 in laL was also proposed to be equivalent to Trp108 in HEWL (Duewel, 1997). Since the alignment was done prior to structure determination, the prediction at the time was less evident due to weak

homology in neighbouring residues. It is interesting that this residue aligns very well in all the structures compared, marking its importance in the enzymatic reaction (Fig. 4.26).



**Figure 4.26 Superposition of laL with HEWL, HuL, Slf70, T4L and GEWL.** Monomer B of laL is depicted here. Only the equivalent Trp-124 and the catalytic residue of the other structures are shown for clarity. The hydrophobic aromaticity in this position is highly conserved, indicating its possible role in maintaining the abnormally high pKa of the catalytic residue. See text for more detail. (Trp-124 equivalent residues are Trp-109 in HuL, Tyr-552-tyr in Slf70, Trp-108 in HEWL, Phe-104 in T4L, Tyr-147 in GEWL; Catalytic residues in other structures are Glu-35 in HuL, Glu-478 in Slf70, Glu-35 in HEWL, Glu-11 in T4L, Glu-73 in GEWL)

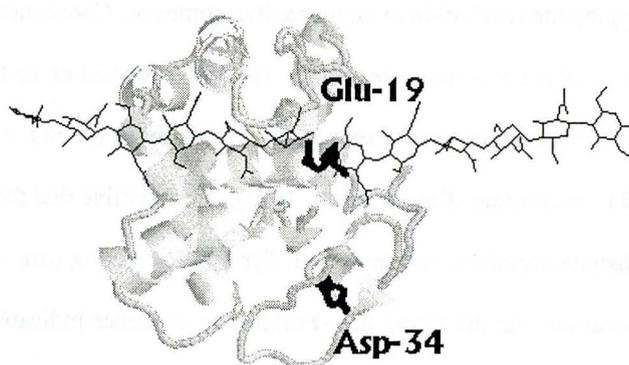
Due to the conservation of this residue in most lysozymes and Slf70, Trp-108 in HEWL has been proposed to have a role in maintaining the abnormally high pKa (pKa = 6.1) in Glu-35 compared to the normal Glu pKa of 4.4 (Inoue, *et al.*, 1992). Dequenching of Trp-108 upon (GlcNAc)<sub>n</sub> complexation has been attributed to conformational changes which involve the deprotonation and the movement of Glu-35 away from Trp-108 (Lehrer & Fassman, 1967). Since Trp-124 is in van der Waals contact with Glu-19, it might play a similar role in saccharide binding and maintenance of the high pKa of Glu-19 (Fig. 4.26).

### *Tyr-132*

Tyr-132 forming stacking interactions in ring D has no counterpart in other lysozymes. Inhibition studies have shown that (GlcNAc)<sub>4-6</sub> have comparable binding affinities to laL but are stronger than those for (GlcNAc)<sub>1-3</sub>, suggesting that a minimum of four rings are required to satisfy interactions with laL (Duewel, 1997). In contrast, maximal binding is achieved by (GlcNAc)<sub>3</sub> in HEWL. The additional van der Waals interaction in ring D might contribute to the difference. It is possible that the stacking interactions prevented ring D from penetrating deeper into the cleft, trapping the saccharide in an unreactive complex. Consequently, chitosaccharides can only inhibit but not act as a substrate of laL. Tyr-132 is located in the helical upper domain where the peptide moiety might bind as seen in the T4 mutant in which the peptide from NAM is covalently linked to the protein (Kuroki, *et al.*, 1993). It is possible that the peptide moiety in the physiological substrate somehow interferes with Tyr-132 interacting with ring D, thus allowing the saccharide to intrude into the active site. Preliminary evidence indicating that in the presence of a synthetic peptide, laL might be able to catalyze the cleavage of (GlcNAc)<sub>5</sub>-PNP gives support to this hypothesis (Duewel, 1997). Perhaps the role of the synthetic peptide is to prevent Tyr-132 from interacting with ring D, thus allowing the substrate to be cleaved. Furthermore, the adjuvant peptide N-acetylmuramyl-L-alanyl-D-isoglutamine, which closely resembles the structure of the peptide-substituted NAM component of peptidoglycan, does not inhibit the bacteriolytic activity of laL (Duewel, 1997). Could this mean that the interactions between laL, the glycan and the peptide portion of the peptidoglycan need not occur synergistically? As long as something is bound to the peptide site, the glycan can have full intrusion into the active site and be cleaved. Also, preliminary observations suggesting that the addition of synthetic peptides might be able to increase the activity of laL also supports the present hypothesis of the peptide's role in laL catalysis (Duewel, 1997). Assuming the synthetic peptide is readily bound to the peptide groove on the protein, the active site is readily opened for the glycan to bind and form a reactive complex. Therefore, it would make sense that the bacteriolytic activity increases in the presence of synthetic peptide because the activity rate no longer depends on the peptide moiety in the peptidoglycan to

prevent Tyr-132 or other residues from interacting with the D ring. Predictions in this section await more future work to be done on mutating Tyr-132 to an amino acid with a smaller side chain that is unable to interact with ring D, followed by investigating whether or not the mutant can cleave (GlcNAc)<sub>5</sub>-PNP in the absence of synthetic peptides.

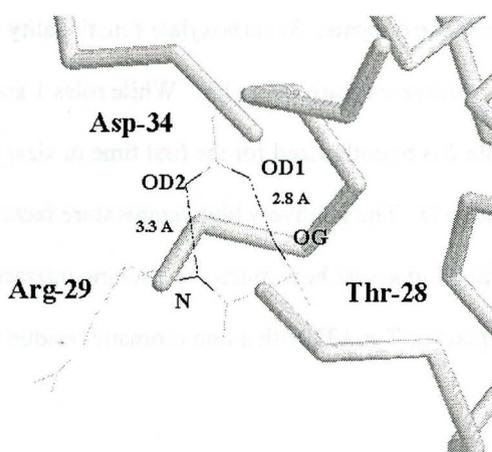
*The absence of Asp-52 counterpart in laL*



**Figure 4.27 Location of the proposed HEWL-Asp52 equivalent in laL, Asp-34.** The two hypothesized residues are highlighted here. As shown, Asp-34 is situated too far away from the active site to act as an HEWL-Asp52 equivalent, which is proposed to have stabilization effects on the transition state.

Asp-34 has been proposed to be analogous to Asp-52 in HEWL due to the observed loss of activity when this residue is mutated (Jespers, *et.al.*, 1992). The role of Asp-52 in HEWL is thought to act as an oxocarbenium ion stabilizer (Phillips, 1966). The hypothesis is disproved from the laL structures since Asp-34 is located too far away from the active site to have any effect on the oxocarbenium ion (Fig. 4.27). The absence of the second catalytic Asp residue is also observed in g-type lysozyme and in Slf70 (Weaver, *et. al.*, 1995, Thunnissen, *et. al.*, 1994, 1995a; 1995b). The requirement of the second catalytic carboxylate group has been controversial. It has been shown that the carboxylate is not essential for catalysis but its presence will increase the enzyme efficiency (Matsumura & Kirsch, 1996). Since Asp-34 is located at the disordered  $\beta$ -

sheet, it is probably important in maintaining the overall fold in the lower domain so that residues responsible for catalysis are positioned properly at the active site. The oxygen atoms of the side chain in Asp-34 form hydrogen bonds with Thr-28 and Arg-29 located at the loop following H1 where Glu-19 is situated (Figure 4.28). It is likely that these interactions contribute to the stability of the protein because it was observed that Asp34Asn and Asp34Ala mutants had lower melting temperatures (Jespers, *et al.*, 1990).



**Figure 4.28 Hydrogen bonding interactions in Asp-34.** Thr-28 and Arg-29 are located at the loop following H1 where Glu-19 is situated. Atom OD1 in Asp-34 hydrogen bonds with OG atom of Thr-28; atom OD2 in Asp-34 hydrogen bonds with the N atom in the main chain of Arg-29.

#### 4.2.9 The Role of the Peptide in laL Catalysis

The peptide requirement for laL catalysis has been mentioned several times in this thesis. In this section, a summary of the possible role of the peptide in catalysis is discussed. A superposition of the T4 mutant with a bound substrate and laL·(GlcNAc)<sub>6</sub> was done to investigate whether or not a peptide groove is present in laL. The covalently linked peptide in T4 resides in a groove between two helices, which is located on the surface of the protein at the lower domain (Figure 1.7). A similar groove formed between two helices, locating close by where a NAM

residue is supposed to bind (subsites B, D, and F), is also present in laL·(GlcNAc)<sub>6</sub>. Taken together, it is likely that the peptide moiety binds to the same groove on the laL surface. However, in the comparison of the surface depiction diagram in Figure 1.7 (T4L) and the surface of laL (not shown), the groove in laL is not as deep and well defined as that of T4L.

Throughout the discussions, three roles of the peptide have been proposed. 1) the binding energy from peptide interactions compensate for the high energy distortion in D ring necessary for the reaction mechanism; 2) the binding of the peptide in the lower domain prevents residues in the region from interacting with the glycan component of peptidoglycan, thus allowing it to penetrate into a reactive complex with the enzyme; 3) carboxylate functionality in peptide acts like Asp52 in HEWL to stabilize the positive oxocarbenium ion. While roles 1 and 3 have been proposed elsewhere in literature, role 2 is hypothesized for the first time in view of the stacking interactions between Tyr-132 and residue D. The relatively high temperature factor implies weak interactions with the protein. It is difficult to assess how much the stacking interaction contributes to the binding of residue D. Replacing Tyr-132 with a non-aromatic residue will definitely shed light on its role.

#### 4.2.10 The Implication of Ring D

Although the hypothesis that the role of Tyr-132 is to prevent ring D from penetrating into the active site explains previous observations of the synthetic peptide study, the reader should keep in mind that the inhibitor did not bind in one piece across laL·(GlcNAc)<sub>6</sub>. The stacking interaction observed between Tyr-132 and ring D could be due to crystallographic artifact. First of all, it is possible that GlcNAc without the peptide moiety cannot bind across the active site due to the distortion requirement in ring D. As a result, (GlcNAc)<sub>6</sub> prefers to bind to two proteins as four and two rings to avoid distortion. The stacking interactions might have formed then to optimize binding energy rather than for the more direct role in inhibition suggested before. Similar stacking interactions were not observed in either GEWL or Slt70, which both require the peptide moiety for

activity, making the prediction less plausible. In consideration of consistency, the role of the peptide seems more likely to be participation in catalysis by supplying the second carboxylate functionality to stabilize the oxocarbonium ion, because, similar to IaL, both GEWL and Slf70 lacking the Asp52 counterpart have strict requirement for the peptide moiety. On the other hand, the hypothesized role of Tyr-132 explains the observations of previous peptide studies well. As such, further mutagenesis studies are required to address whether or not Tyr-132 is important in inhibition.

## Chapter 5 Summary and Future Work

### 5.1 Summary

The structure of a lysozyme from bacteriophage lambda is the subject of this thesis. Since the ligand-free mutant form was solved by another group during the middle of this project and the wild type ligand-free crystal form was in poor diffraction quality, the focus was shifted to the laL·(GlcNAc)<sub>6</sub> co-crystal. The wild type laL·(GlcNAc)<sub>6</sub> complex has been solved to 2.6 Å by molecular replacement using the mutant structure as a model. The protein packs as a dimer in the crystal with the two backbones being nearly identical. The major differences are located at the upper lip loop and the lower lip loop. The overall fold of laL resembles that of common lysozyme, consisting of an upper and lower domain and a connecting helix in the middle of these domains. The (GlcNAc)<sub>6</sub> binds in the deep crevice in between the domains. All subsites are occupied simultaneously by two (GlcNAc)<sub>6</sub> molecules. Four rings of one (GlcNAc)<sub>6</sub> bind at subsites A to D, while two rings from the adjacent (GlcNAc)<sub>6</sub> molecule bind to the remaining subsites E and F. Comparison studies find that neither the Trp-azaTrp mutations nor saccharide binding changes the overall fold significantly.

Several observations are made in the sugar-binding mode in laL in section 4.2.7. All rings adopt a full chair conformation and have well-ordered density. Both rings A and F have the least contact with the enzyme, as they are located close to the solvent. It is postulated that the strong binding of ring E stabilizes ring F which has a low thermal factor despite its weak interactions with the protein. Ring B has the characteristic stacking interactions with Trp-73, which also makes contact with ring C. The distinctive hydrogen bonding network in the 2-acetamido group in ring C observed in all other lysozymes and Slf70 structures is also noted in laL. Ring D is situated far away from the active site close to the protein surface. It forms perfect stacking interactions with Tyr-132, an observation not seen in other lysozymes and Slf70 structures. It is

hypothesized that this unique interaction prevents the pyranose ring from penetrating deeper into the active site. Since the Phillips model suggests that a distortion in this ring is required for the enzyme mechanism, the failure of ring D to intrude into the active site may be the reason why GlcNAc is not a substrate but an inhibitor for laL. Since a peptide moiety is essential for laL activity, the peptide binding might provide energy to compensate for the distortion in ring D. It is suggested in this thesis that the role of the peptide moiety might be to interfere with Tyr-132, thus preventing ring D from fully entering into the active site. Ring E is likely to bind at this location during catalysis because the O4 atom is in hydrogen bonding distance with the OE atom of Glu-19. Its extensive interaction with the enzyme and low thermal factors indicates tight binding to the protein. Moreover, the surface depiction shows that this ring is buried deep inside the active site. The opening of the active site in other lysozymes is much wider. It is predicted in this thesis that rings E and F remain in the active site longer than what is necessary in other lysozyme structures due to differences in the reaction mechanism. In lysozymes, a water molecule is required to diffuse into the active site to complete the reaction; in laL, the O6 atom of ring D performs the last nucleophilic attack. The stronger binding of residues E and F prevents water molecules from attacking the tetrasaccharide, thus allowing transglycosylation to occur.

LaL has no second catalytic Asp residue (Asp-52 in HEWL), though Asp-34 was previously predicted to fulfill this role. Asp-34 is located in the  $\beta$ -sheet region of the lower domain. The observation that mutation of this residue reduces laL activity might point to its importance in maintaining the folding of the  $\beta$ -sheet region.

The peptide-binding groove is predicted in section 4.2.9 based on the structure of a T4 mutant covalently complexed with a substrate. Three possible roles of the peptide are suggested: 1) the peptide-binding energy might compensate for the high energy distortion in ring D required for catalysis; 2) the peptide might prevent Tyr-132 from forming stacking interactions with ring D, so that the sugar can penetrate deeper into the cleft; and 3) the carboxylate functionality might act as an Asp-52 counterpart to stabilize the positive oxocarbenium ion.

## **5.2 Future Work**

The peptide-binding mode in laL is worth more attention in future efforts. First of all, the function of Tyr-132 suggested in this thesis can be verified by mutating the residue to one incapable of forming stacking interactions since it is difficult to assess how much stacking interaction is involved in preventing the ring D from penetrating into the active site. If a Tyr-132 mutation can produce a protein that can cleave homopolymers of GlcNAc to generate a 1,6 anhydro compound, it would be a convenient method to produce large amounts of anhydro-muropeptides for further studies. Interest in anhydro-muropeptides has increased since they are found to have a wide variety of useful biological activities.

Duewel (1997) has made synthetic peptides that are capable of complexing with laL. In the same study, the presence of the synthetic peptides seems to enhance effects on laL activity. A suggestion for a future project is to co-crystallize the synthetic peptide with laL. Similarly, crystals of laL complexed with synthetic peptide and homopolymers of GlcNAc can be prepared. The structure of the first suggested complex will give useful information on the peptide-binding mode in laL. If the prediction that the role of the peptide is to prevent Tyr-132 from interacting with ring D, the crystal of the second suggested project might produce several interesting observations: 1) laL complexed with the synthetic peptide binding to the surface groove, one complete hexasaccharide molecule binding across the active site; 2) laL complexed with the synthetic peptide binding to the surface groove with the hexasaccharide cleaved into two fragments, with the non-reproducible tetrasaccharide trapped in subsites A to D.

## **5.3 Concluding Remark**

This project has certainly shed light on the understanding of the distinctive laL mechanism. Several hypotheses have been made based on the structural findings presented in this thesis. Such predictions await support from more structural and mutagenesis studies as suggested in section 5.2.

## References

- Arber, W. (1983). A Beginner's Guide to Lambda Biology. In *Lambda II*, 381-394, Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory.
- Arnheim, N., Inouye, M., Law, L. and Laudin, A. (1973) Chemical studies on the enzymatic specificity of goose egg white lysozyme. *J. Biol. Chem.* **248**, 233-236.
- Bacon, D. & Anderson, W.F. (1988) A fast algorithm for rendering space-filling molecule pictures *J. Molec. Graphics*, **6**, 219-220.
- Banerjee, S.K., Holler, E., Hess, G.P., & Rupley, J.A. (1975) Reaction of N-acetylglucosamine oligosaccharides with lysozyme. Temperature, pH, and solvent deuterium isotope effects; equilibrium, steady state, and pre-steady state measurements. *J. Biol. Chem.* **250**, 4355-4367.
- Belfort, M. and Wulff, D.L., 1973. An analysis of the processes of infection and induction of *E. coli hfl-1* by bacteriophage lambda. *Virology* **55**: 183.
- Bienkowska-Szewczyk, K., Lipinska, B., & Taylor, A. (1981) The R gene product of bacteriophage lambda is the murein transglycosylase. *Mol Gen. Genet.* **184**, 111-114.
- Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C., Sarma, V.R. (1965) Structure of hen egg-white lysozyme. *Nature(London)*, **206**, 757-761.
- Blake, C.C., Mair, G.A., North, A.C., Phillips, D.C. and Sarma, V.R. (1967a) On the conformation of the hen egg-white lysozyme molecule. *Proc. Roy. Soc. Lond. Series B: Biol. Sci.* **167(9)**, 365-377.
- Blake, C.C., Johnson, L.N., Mair, G.A., North, A.C., Phillips, D.C. and Sarma, V.R. (1967b) Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. Roy. Soc. Lond. Series B: Biol. Sci.* **167 (9)**, 378-388.
- Branden & Tooze, (1991) *Introduction to protein structure*, 8-9, New York and London. Garland Publishing Inc.
- Brünger, A.T. (1992) X-PLOR version 3.1 A system for X-ray Crystallography and NMR (Yale University Press, New Haven, CT).
- Brünger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
- Brünger, A.T. (1993) Assessment of phase accuracy by cross validation: the Free R value. Methods and applications. *Acta Cryst.* **D49**, 24-36.
- Cheetham, J.C., Artymiuk, P.J. and Phillips, D.C. (1992) Refinement of an enzyme complex with inhibitor bound at partial occupancy. Hen egg-white lysozyme and tri-N-acetylchitotriose at 1.75 Å resolution. *J. Mol. Biol.* **224(3)**, 613-628.
- Chipman, D.M., & Sharon, N. (1969) Mechanism of lysozyme action. *Science*, **165**, 454-465.

- Dijkstra, B.W., & Thunnissen, A.-M. W.H. (1994) 'Holy' proteins. II: The soluble lytic transglycosylase. *Curr. Opin. Struct. Biol.* **4**, 810-813.
- Duewel, H.S. (1997) Lambda lysozyme : overexpression, ligand studies, and incorporation of trifluoromethionine. Ph.D Dissertation, University of Waterloo.
- Duewel, H.S., Daub, E., Honek, J.F. (1995). Investigations of the interactions of saccharides with the lysozyme from bacteriophage lambda. *Biochim. Biophys. Acta*, **1247**, 149-158.
- Evrard, C., Fastrez, J., Declercq, J.-P. (1998) Crystal structure of the lysozyme from bacteriophage lambda and its relationship with V and C-type lysozymes. *J. Mol. Biol.* **278**, 151-164.
- Ford, L.O., Johnson, L.N., Machin, P.A., Phillips, D.C. and Tjian, R. (1974) Crystal structure of a lysozyme-tetrasaccharide lactone complex. *J. Mol. Biol.* **88**(2), 349-371.
- Friedman, D.I., Gottesman, M.(1983). Lytic Mode of Lambda Development. In *Lambda II*, 21-52, Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory.
- Garrett, J., Fusselman, R., Hise, J., Chiou, L., Smith-Grillo, D., Schulz, J. & Young, R. (1981). Cell lysis by induction of cloned lambda lysis genes. *Mol. Gen. Genet.* **182**, 326-331.
- Ghuysen, J.M. & Shockman, G.D. (1973). Biosynthesis of peptidoglycan. In *Bacterial membranes and walls*. 37-130, New York. Marcel Dekker Inc.
- Ho, Y.S., Rosenberg, M. (1988). Structure and Function of the Transcription Activator Protein cII and Its Regulatory Signals. In *The Bacteriophage*, vol. 2, 725, New York and London. Plenum Press.
- Holler, E., Rupley, J.A., & Hess, G.P. (1975) Productive and unproductive lysozyme-chitosaccharide complexes. Kinetic investigations. *Biochemistry*. **14**, 2377-2385.
- Inoue, M., Yamada, H., Yasukochi, T., Kuroki, R., Miki, T., Horiuchi, T., & Imoto, T. (1992). Multiple role of hydrophobicity of tryptophan-108 in chicken lysozyme: structural stability, saccharide binding ability, and abnormal pKa of glutamic acid-35. *Biochemistry*. **31**, 5545-5553.
- Insight II User Guide*, version 2.3.0. San Diego: Biosym Technologies, 1993.
- Jespers, L., Sonveaux, E., & Fastrez, J. (1992) Is the bacteriophage lambda lysozyme an evolutionary link or a hybrid between the C and V-type lysozymes? Homology analysis and detection of the catalytic amino acid residues. *J. Mol. Biol.* **228**, 529-538.
- Jollès, P., and Jollès, J. (1984) What's new in lysozyme research? Always a model system, today as yesterday. *Mol Cell. Bioch.* **63**, 165-189.
- Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.
- Kabsch W. & Sander C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, **22**, 2577-2637

- Kelly, J.A., Sielecki, A.R., Sykes, B.D., James, M.N.G., Phillips, D.C. (1979) X-ray crystallography of the binding of the bacterial cell wall trisaccharide NAM-NAG-NAM to lysozyme. *Nature* **282**, 875-878.
- Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography*. **24**, 946-950.
- Kuroki, R., Weaver, L.H. and Mathews, B.W. (1993) A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262**(5142), 2030-2033.
- Laskowski R A, MacArthur M W, Moss D S & Thornton J M (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.
- Lattman, E. (1985) Use of the Rotation and Translation Functions. *Methods In Enzymology*. **115**, 55-76.
- Lehrer, S.S., & Fasman, G.D. (1967) Fluorescence of lysozyme and lysozyme substrate complexes. Separation of tryptophan contributions by fluorescence difference methods. *J.Biol.Chem.* **242**, 4644-4645.
- Luzzati, V. (1952) Statistical treatment of errors in the determination of crystal structures. *Acta Cryst.* **5**, 802-810.
- Maenaka, K., Matsushima, M., Song, H., Sunada, F., Watanabe, K. and Kumagai, I. (1995) Dissection of protein-carbohydrate interactions in mutant hen egg-white lysozyme complexes. *J. Mol. Biol.* **247**, 281-293.
- Malcolm, B.A., Rosenberg, S., Corey, M.J., Allen, J.S., de Baetselier, A. & Kirsch, J.F. (1989) Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg white lysozyme. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 133-137.
- Matsumura, I., & Kirsch, J.F. (1996) Is aspartate 52 essential for catalysis by chicken egg white lysozyme? The role of natural substrate-assisted hydrolysis. *Biochemistry*, **35**, 1881-1889.
- Merritt, E.A. & Murphy, M.E.P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta. Cryst.* **D50**, 869-873.
- Morris A L, MacArthur M W, Hutchinson E G & Thornton J M (1992). Stereochemical quality of protein structure coordinates. *Proteins*, **12**, 345-364.
- Nicholls, A., Sharp, K.A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Gen.* **11**, p. 281ff.
- Otwinowski, Z., Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, p. 307-325, *In* Carter, C.W.Jr, Sweet, R. (ed.), *Methods in Enzymology* **276** Academic Press Inc.
- Phillips, D.C. (1966) The three-dimensional structure of an enzyme molecule. *Sci. Amer.* **215**, 78-90.

- Pincus, M.R., Zimmerman, S.S., Scheraga, H.A. (1977) Structures of enzyme-substrate complexes of lysozyme. *Proc. Natl. Acad. Sci.* **74**, 2629-2633.
- Quicocho, F.A., (1986) Carbohydrate-binding proteins: Tertiary structures and protein-sugar interactions. *Annu Rev Biochem* **55**, 287-315.
- Rodgers, D.W. (1997) Practical Cryocrystallography, p. 183-202, *In* Carter, C.W.Jr, Sweet, R. (ed.), *Methods in Enzymology* **276** Academic Press Inc.
- Smith-Gill, S.J., Rulpley, J.A., Pincus, M.R., Carty, R.P., & Scheraga, H.A. (1984) Experimental identification of a theoretically predicted "left-sided" binding mode for (GlcNAc)<sub>6</sub> in the active site of lysozyme. *Biochemistry*, **23**, 993-997.
- Strynadka, N.C.J. and James, M.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.
- Strynadka, N.C.J. & James, M.N.G. (1996) Lysozyme: A model enzyme in protein crystallography. *EXS* **75**: 185-224.
- Song, H., Inaka, K., Maenaka, K., Matsushima, M. (1994) Structural changes of active site cleft and different saccharide binding modes in human lysozyme co-crystallized with hexa-N-acetyl-chitohexaose at pH 4.0. *J. Mol. Biol.* **244**(5), 522-540.
- Taylor, A., Das, B.C., Heijenoort, J.V. (1975) Bacterial-cell-wall peptidoglycan fragments produced by phage  $\lambda$  or Vi II endolysin and containing 1,6-anhydro-N-acetylmuramic acid. *Eur. J. Biochem.* **53**, 47-54.
- Thunnissen, A.-M. W.H., Dijkstra, A.J., Kalk, K.H., Rozeboom, H.J., Engel, H., Keck, W., & Kijkstra, B.W. (1994) Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature*, **367**, 750-753.
- Thunnissen, A.-M. W.H., Rozeboom, H.J., Kalk, K.H., & Dijkstra, B.W. (1995a) Structure of the 70-kDa soluble lytic transglycosylase complexed with bulgecin A. Implications for the enzymatic mechanism. *Biochemistry*, **34**, 12729-12737.
- Thunnissen, A.-M. W.H., Isaacs, N.W., & Dijkstra, B.W. (1995b) The catalytic domain of a bacterial lytic transglycosylase defines a novel class of lysozymes. *Proteins: Struct. Funct. Genet.*, **22**, 245-258.
- Weaver, L.H., Grütter, M.G., & Matthews, B.W. (1995) The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.*, **245**, 54-68.