

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

X-RAY CRYSTALLOGRAPHIC
INVESTIGATIONS OF THE STRUCTURES
OF ENZYMES OF MEDICAL AND
BIOTECHNOLOGICAL IMPORTANCE

by

Richard Lawrence Kingston

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in the

Department of Biochemistry

at

MASSEY UNIVERSITY, NEW ZEALAND

November, 1996

ABSTRACT

This thesis is broadly in three parts. In the first, the problem of identifying conditions under which a protein will crystallize is considered. Then structural studies on two enzymes are reported, glucose-fructose oxidoreductase from the bacterium *Zymomonas mobilis*, and the human bile salt dependent lipase (carboxyl ester hydrolase).

The ability of protein crystals to diffract X-rays provides the experimental data required to determine their three dimensional structures at atomic resolution. However the crystallization of proteins is not always straightforward. A systematic procedure to search for protein crystallization conditions has been developed. This procedure is based on the use of orthogonal arrays (matrices whose columns possess certain balancing properties). The theoretical and practical background to the problem is discussed, and the relationship of the presented procedure to other published search methods is considered.

The anaerobic Gram-negative bacterium *Zymomonas mobilis* occurs naturally in sugar-rich growth media, and has attracted much interest because of its potential for industrial ethanol production. In this organism the periplasmic enzyme glucose-fructose oxidoreductase (GFOR) is involved in a protective mechanism to counter osmotic stress. The enzyme is unusual in that it contains tightly associated NADP which is not released during its catalytic cycle. The crystal structure of *Z. mobilis* GFOR has been determined by the method of multiple isomorphous replacement, and refined by restrained least squares methods using data extending to an effective resolution of 2.7 Å. The structure determination reveals that each subunit of the tetrameric protein is folded into two domains, one of which is the classical dinucleotide binding domain, or Rossmann fold. The C-terminal domain is a nine-stranded predominantly antiparallel β -sheet around which the tetramer is constructed. Preceding the Rossmann fold there is a 30 amino acid proline rich 'arm' which wraps around an adjacent subunit in the tetramer. The N-terminal arm buries the adenine ring of the NADP, and may also be involved in stabilization of the quaternary structure of the enzyme. The tight association of NADP is accounted for by the structure. An unsuspected structural relationship has been discovered between GFOR and the cytoplasmic enzyme glucose-6-phosphate dehydrogenase (G6PD). It is proposed that GFOR and G6PD derive from a common ancestral gene, and GFOR has evolved to allow it to function in the bacterial periplasm where it is required.

The human bile salt dependent lipase (BSDL) is secreted by the pancreas into the digestive tract, and by the lactating mammary gland into human milk, and is integral to the effective absorption of dietary lipids. It is markedly non-specific, and as its name implies is only active against water-insoluble substrates in the presence of primary bile salts. This differentiates the

enzyme from conventional lipases. Diffraction data has been collected from crystals of native BSDL (isolated from human milk), and from crystals of recombinant BSDL (including a truncated variant which lacks a C-terminal heavily glycosylated tandem repeat region found in the native enzyme). The structure of the truncated variant has been partially determined at 3.5 Å resolution, by the method of molecular replacement. The recent collection of a higher resolution (2.8 Å) data set should allow the completion of the structure. The current status of the crystallographic investigations of the human bile salt dependent lipase are reported.

ACKNOWLEDGMENTS

I thank my principal supervisor, Professor Ted Baker for his friendship, his enthusiasm for science, and for allowing me much freedom to follow my ideas. I also thank Mrs. Heather Baker for her help and constant encouragement.

I thank my assistant supervisors, Dr. Bryan Anderson (for many discussions about crystallographic computing), and Professor Sylvia Rumball (who initiated a collaboration with Umeå University on bile salt dependent lipase).

I would like to specifically acknowledge the scientific contributions of Dr. Rick Faber, and Dr. Stanley Moore to the work presented in this thesis (when they weren't fly fishing, that is).

I would like to thank all the other friends I have made during my time at Massey. Andrew, Rosemary, Isobel, Shaun, Treena, Ross, Catherine, Alain and Anne-Gael, Mark, Michelle, Jakki and Paul, Neil and Liz, Phil, and Maria to name only some. Thanks.

The work on bile salt dependent lipase was carried out in collaboration with Professor Olle Hernell and Dr. Lars Bläckberg (Umeå University, Sweden); and Dr. Kerry Loomes (Auckland University, New Zealand). The work on glucose-fructose oxidoreductase was carried out in collaboration with Professor Robert Scopes (La Trobe University, Australia). I thank these people for their scientific contributions, and especially Kerry, for believing we could solve the structure and working so hard to overcome difficult technical problems.

For financial assistance while this work was completed I thank Massey University (through the award of a doctoral scholarship), and latterly Professor Ted Baker.

Thanks to the members of my family for encouragement and support. I owe much to my parents for supporting what I do, and helping to finance my study.

Finally I would like to thank Wendy, a very special friend, and someone whom it has been difficult to be separated from while this thesis was written

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vii
ABBREVIATIONS	ix
RELATED PUBLICATIONS	xi

Chapter 1

PROTEIN CRYSTALLIZATION

1.1	INTRODUCTION	1
1.1.1	HISTORICAL BACKGROUND	1
1.1.2	THE EXPERIMENTAL PROBLEM TODAY	2
1.1.3	PHYSICAL BACKGROUND	2
1.2	SEARCH DESIGNS FOR PROTEIN CRYSTALLIZATION	4
1.2.1	TERMS ASSOCIATED WITH EXPERIMENTAL DESIGN	4
1.2.2	CURRENT APPROACHES TO SEARCHING FOR PROTEIN CRYSTALLIZATION CONDITIONS	5
1.2.3	GENERAL CRITERIA FOR INITIAL SEARCH EXPERIMENTS	6
1.2.4	ORTHOGONAL ARRAYS	7
1.2.5	UNDERLYING FACTORIAL STRUCTURE FOR SEARCH EXPERIMENTS	13
1.2.6	PRACTICAL IMPLEMENTATION OF ORTHOGONAL ARRAY-BASED SEARCH DESIGNS	14
1.2.7	EXPERIMENTAL CONSIDERATIONS	22
1.3	PRACTICAL APPLICATION TO SEVERAL PROBLEMS	24
1.3.1	BILE SALT DEPENDENT LIPASE	25
1.3.2	GLUCOSE-FRUCTOSE OXIDOREDUCTASE	27
1.3.3	$\alpha_2\epsilon_2$ EMBRYONIC HEMOGLOBIN	28
1.4	RELATIONSHIP TO PUBLISHED SEARCH PROCEDURES	29
1.5	DISCUSSION AND CONCLUSION	31
1.5.1	ANALYSIS USING LINEAR MODELS	31
1.5.2	DISTRIBUTION PROPERTIES OF ORTHOGONAL ARRAYS	32
1.5.3	DYNAMIC LIGHT SCATTERING	33
1.5.4	CRYSTALLIZATION OF OTHER BIOLOGICAL MACROMOLECULES	34
1.5.5	CONCLUSION	34

Chapter 2

GFOR: STRUCTURE DETERMINATION

2.1	INTRODUCTION.....	36
2.2	OVERVIEW OF THE STRUCTURE DETERMINATION	38
2.3	PROTEIN PURIFICATION AND CRYSTALLIZATION	38
2.3.1	CELL GROWTH AND PROTEIN PURIFICATION.....	38
2.3.2	PROTEIN CRYSTALLIZATION.....	38
2.4	X-RAY DATA COLLECTION	39
2.4.1	CHARACTERIZATION OF THE CRYSTALS.....	39
2.4.2	DATA COLLECTION AND PROCESSING	41
2.4.3	SPACE GROUP TRANSITIONS	46
2.5	MULTIPLE ISOMORPHOUS REPLACEMENT	47
2.6	DENSITY MODIFICATION.....	49
2.6.1	ENVELOPE DEFINITION	49
2.6.2	PHASE IMPROVEMENT AND EXTENSION	52
2.6.3	RESULTS OF PHASE IMPROVEMENT AND EXTENSION	56
2.7	MODEL BUILDING AND REFINEMENT.....	56
2.7.1	BUILDING THE INITIAL MODEL.....	56
2.7.2	RECOVERY OF THE MISSING STRUCTURE	58
2.7.3	MODEL REFINEMENT.....	59
2.7.4	COMBINATION OF PHASE INFORMATION	60
2.7.5	ITERATIVE CYCLES OF REBUILDING, REFINEMENT AND PHASE COMBINATION.....	61
2.7.6	FINAL REFINEMENT OF THE MODEL	64
2.8	SUMMARY	72

Chapter 3

GFOR: STRUCTURE AND FUNCTION

3.1	STRUCTURE OF THE MONOMER.....	74
3.1.1	N-TERMINAL DOMAIN	74
3.1.2	C-TERMINAL DOMAIN	75
3.1.3	N-TERMINAL ARM	77
3.2	COMPARISON WITH GLUCOSE 6-PHOSPHATE DEHYDROGENASE	77
3.2.1	STRUCTURE COMPARISON.....	77
3.2.2	EVOLUTIONARY IMPLICATIONS.....	79
3.3	STRUCTURE OF THE TETRAMER	81
3.4	DINUCLEOTIDE BINDING	82
3.4.1	NADP CONFORMATION	82
3.4.2	INTERACTIONS WITH GFOR	82

3.4.3	TIGHT ASSOCIATION WITH GFOR.....	85
3.4.4	EVOLUTIONARY IMPLICATIONS OF THE N-TERMINAL ARM.....	86
3.5	IMPLICATIONS FOR CATALYSIS	86
3.5.1	BACKGROUND.....	86
3.5.2	THE ACTIVE SITE OF GFOR.....	87
3.5.3	SEQUENCE AND STRUCTURAL SIMILARITIES	88
3.5.4	GENERAL DISCUSSION	90
3.6	GFOR AS A PERIPLASMIC ENZYME.....	91
3.7	CONCLUSION.....	92

Chapter 4

BILE SALT DEPENDENT LIPASE

4.1	INTRODUCTION.....	94
4.1.1	GENERAL BACKGROUND.....	94
4.1.2	LIPASES.....	94
4.1.3	BILE SALT DEPENDENT LIPASE	98
4.1.4	THE ROLE OF STRUCTURAL STUDIES	110
4.2	NATIVE BSDL.....	111
4.2.1	PROTEIN PURIFICATION AND CRYSTALLIZATION	111
4.2.2	CHARACTERIZATION OF THE CRYSTALS.....	111
4.2.3	ANISOTROPIC DIFFRACTION.....	114
4.2.4	DIFFUSE SCATTERING	116
4.2.5	ENZYMATIC DEGLYCOSYLATION.....	119
4.3	RECOMBINANT FULL LENGTH BSDL.....	122
4.3.1	EXPRESSION, PURIFICATION AND CRYSTALLIZATION	122
4.3.2	PRELIMINARY CRYSTALLOGRAPHIC INVESTIGATION	123
4.4	RECOMBINANT TRUNCATED BSDL	124
4.4.1	EXPRESSION AND PURIFICATION.....	124
4.4.2	CRYSTALLIZATION.....	125
4.4.3	DATA COLLECTION AND PROCESSING	127
4.4.4	STRUCTURE SOLUTION BY MOLECULAR REPLACEMENT.....	133
4.4.5	BUILDING AN INITIAL MODEL.....	136
4.4.6	REFINEMENT AT LOW RESOLUTION.....	138
4.4.7	DIFFICULTIES IN COMPLETION OF THE PARTIAL STRUCTURE.....	139
4.4.8	CURRENT STATUS OF THE STRUCTURE DETERMINATION	140
	REFERENCES.....	143

LIST OF FIGURES

Chapter 1

FIGURE 1.1	GEOMETRIC REPRESENTATION OF THE 2 X 2 X 2 FACTORIAL AND SOME POSSIBLE SUBSETS.....	9
FIGURE 1.2	CRYSTAL OF NATIVE BSDL.....	26
FIGURE 1.3	CRYSTAL OF <i>ZYMONONAS MOBILIS</i> GFOR.....	28
FIGURE 1.4	GEOMETRIC REPRESENTATION OF TWO ORTHOGONAL ARRAYS, OA(8, 3, 2X2X4, 2).....	33

Chapter 2

FIGURE 2.1	STEREOGRAPHIC PROJECTIONS OF THE SELF-ROTATION FUNCTIONS OF THE TWO CRYSTAL FORMS OF GFOR.....	40
FIGURE 2.2	PATTERSON FUNCTION CALCULATED FROM THE FORM II DATA.....	41
FIGURE 2.3	THE RELATIONSHIP BETWEEN THE TWO CRYSTAL FORMS.....	42
FIGURE 2.4	DATA COLLECTION USING CRYSTALS MOUNTED IN LIQUID-FILLED CAPILLARIES.....	44
FIGURE 2.5	CALCULATED AND OBSERVED ELECTRON DENSITY HISTOGRAMS.....	53
FIGURE 2.6	ELECTRON DENSITY MAPS FOR GFOR.....	56
FIGURE 2.7	STEREOVIEW OF ELECTRON DENSITY MAPS CALCULATED FROM AN ATOMIC AND A 'GLOBIC' REPRESENTATION OF AN α -HELIX AT 3.0 Å RESOLUTION.....	63
FIGURE 2.8	STEREOVIEW OF A DIFFERENCE FOURIER SYNTHESIS WITH RESIDUES IN THE REGION CONFLICTING WITH THE PUBLISHED SEQUENCE OMITTED.....	66
FIGURE 2.9	RAMACHANDRAN PLOT FOR THE REFINED GFOR MONOMER.....	71
FIGURE 2.10	ELECTRON DENSITY MAP CALCULATED USING THE FORM II DATA.....	73

Chapter 3

FIGURE 3.1	C α PLOT OF GFOR.....	74
FIGURE 3.2	TOPOLOGY OF GFOR.....	76
FIGURE 3.3	RIBBON DIAGRAMS OF GFOR AND G6PD.....	78
FIGURE 3.4	QUATERNARY STRUCTURE OF GFOR.....	81
FIGURE 3.5	CONFORMATION OF THE ENZYME-BOUND NADP.....	83
FIGURE 3.6	HYDROGEN-BONDING INTERACTIONS BETWEEN GFOR AND NADP.....	84
FIGURE 3.7	THE ACTIVE SITE OF GFOR.....	87
FIGURE 3.8	ALIGNMENT OF SEQUENCES WITH HOMOLOGY TO GFOR.....	89

Chapter 4

FIGURE 4.1	DIAGRAM SHOWING THE CONFORMATIONAL CHANGE ASSOCIATED WITH ACTIVATION IN <i>CANDIDA RUGOSA</i> LIPASE.....	96
FIGURE 4.2	SCHEMATIC DIAGRAM SHOWING THE CONFORMATIONAL CHANGE ASSOCIATED WITH INTERFACIAL ACTIVATION IN THE FUNGAL LIPASES.....	97
FIGURE 4.3	ALIGNMENT OF KNOWN BSDL SEQUENCES.....	102
FIGURE 4.4	TOPOLOGY DIAGRAM OF THE LIPASE/ESTERASE FAMILY FOLD.....	104
FIGURE 4.5	RIBBON DIAGRAM OF <i>T. CALIFORNICA</i> ACETYLCHOLINESTERASE.....	105
FIGURE 4.6	BILE ACID STRUCTURE.....	108
FIGURE 4.7	SPACE FILLING MODEL OF CHOLIC ACID.....	109
FIGURE 4.8	CRYSTAL OF NATIVE BSDL.....	111
FIGURE 4.9	DIFFUSE SCATTERING PATTERNS FROM NATIVE BSDL CRYSTALS (I).....	117

FIGURE 4.10	DIFFUSE SCATTERING PATTERNS FROM NATIVE BSDL CRYSTALS (II).....	118
FIGURE 4.11	ISOELECTRIC FOCUSING OF BSDL.....	120
FIGURE 4.12	CRYSTALS OF DESIALIDATED BSDL.....	121
FIGURE 4.13	CRYSTALS OF FULL-LENGTH RECOMBINANT BSDL.....	123
FIGURE 4.14	CRYSTALS OF TRUNCATED RECOMBINANT BSDL.....	125
FIGURE 4.15	GLASS-SLIDE MOUNTING DEVICE FOR CRYOCRYSTALLOGRAPHY.....	131
FIGURE 4.16	BACKGROUND SCATTER AND ABSORPTION DUE TO THE SOLID-SURFACE MOUNT.....	132
FIGURE 4.17	RESULTS OF PATTERSON CORRELATION REFINEMENT.....	135
FIGURE 4.18	ELECTRON DENSITY FOR TRUNCATED RECOMBINANT BSDL.....	141

ABBREVIATIONS

AChE	Acetylcholinesterase
AMPSO	3-[(1,1-Dimethyl-2-hydroxyethyl)amino]2-hydroxypropanesulfonic acid
BIS-TRIS PROPANE	1,3-bis[tris(Hydroxymethyl)-methylamino]propane
BSDL	Bile salt dependent lipase
BSSL	Bile salt stimulated lipase
CDL	Colipase-dependent lipase
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CRL	<i>Candida rugosa</i> lipase
DHPR	Dihydrodipicolinate reductase
DNA	Deoxyribonucleic acid
EPPS	N-[2-Hydroxyethyl]piperazine-N'[3-propanesulfonic acid]
FAD	Flavin-adenine dinucleotide
FMN	Flavin mononucleotide
G6PD	Glucose-6-phosphate dehydrogenase
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GCL	<i>Geotrichum candidum</i> lipase
GFOR	Glucose-fructose oxidoreductase
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
IEF	Isoelectric focussing
LDH	Lactate dehydrogenase
MDH	Malate dehydrogenase
MES	2-[N-Morpholino]ethanesulfonic acid
MIR	Multiple isomorphous replacement
MOPS	3-[N-Morpholino]propanesulfonic acid
NAD	Oxidized or reduced form of nicotinamide adenine dinucleotide
NADP	Oxidized or reduced form of nicotinamide adenine dinucleotide phosphate
NADP⁺	Oxidized form of nicotinamide adenine dinucleotide phosphate
NADPH	Reduced form of nicotinamide adenine dinucleotide phosphate
NAD(P)	NAD or NADP
NCBI	National Center for Biotechnology Information
NCS	Non-crystallographic symmetry
NIST	National Institute of Standards and Technology
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEG-mme	Polyethylene glycol monomethyl ether

PIPES 1,4-Piperazinediethanesulfonic acid

PQQ Pyrrolo-quinoline quinone

RMS Root mean square

SEL Sequential elimination of levels

SIR Single isomorphous replacement

TAPS N-tris[Hydroxymethyl]methyl-3-aminopropanesulfonic acid

TcAChE *Torpedo californica* acetylcholinesterase

TRIS Tris(hydroxymethyl)aminomethane

RELATED PUBLICATIONS

Some of the material presented in this thesis has already been published, or has been accepted for publication.

Kingston, R.L., Baker, H.M. & Baker, E.N. (1994) Search designs for protein crystallization based on orthogonal arrays. **Acta Crystallographica**. **D50**, 429-440.

Kingston, R.L., Scopes, R.K. & Baker, E.N. (1996) The structure of glucose fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. **Structure**. in Press.