

Structural studies on dehydrogenases



A thesis submitted in part fulfillment of the requirements for
the degree of Doctor of Philosophy

By

Abdelhamid Elbrghathi
B.Sc. (Hons) and M.Sc. University of Benghazi

Department of Molecular Biology and Biotechnology
University of Sheffield

(July2014)

Abstract

The amino acid dehydrogenase superfamily has good potential for biotechnological use for example, either as enzymes for the production of non-natural amino acids, or for use as biosensors. Despite many years work on this family, there is still a gap understanding aspects of both substrate and cofactor binding and reaction mechanism. The crystal structures of *M. smegmatis* Glutamate dehydrogenase (GluDH) have been determined in two crystal forms, one of a ternary complex (GluDH/NADP⁺/2oxoglutarate) at 1.64Å, and one of a binary complex (GluDH/NADPH) at 1.78Å. The binary complex structures of *C. symbiosum* GluDH, with NAD⁺ and L-glutamate have also been solved at 1.80 Å and 1.24Å respectively. Comparison of these structures has allowed the nature of coenzyme specificity in GluDHs to be structurally characterized. Analysis of *M. smegmatis* GluDH/NADP⁺/2oxoglutarate ternary complex showed that 2-oxoglutarate bound to the enzyme as the gem-diol in an extended conformation, which mimics the structure of the carbinolamine intermediate. This structure, together with the high resolution *C. symbiosum* GluDH/L-glutamate complex structure, where the substrate binds in a compact, curved conformation have given valuable insights into the progress of the reaction. They show that the substrate alters its conformation during the reaction, explaining how the abortive reduction of the keto acid to the hydroxy acid is prevented, and showing the stereochemistry of attack by ammonia on the keto acid substrate, aspects of the mechanism that have been poorly understood. In addition, the structure of the AADH family member *Bacillus sphaericus* phenylalanine dehydrogenase in complex with NADH has been determined. Furthermore, initial experiments towards the structure of *mycobacterium smegmatis* typeII NADH: menaquinone oxidoreductase, which is a potential anti bacterial drug target are described.

Acknowledgement

I would like to express my thanks and gratitude to my supervisor Dr. Patrick Baker for the great support, assistance, patience and encouragement during my PhD. I would also like to thank Prof David Hornby for all his help with explaining lots of molecular biology problems, and Dr Qaiser for his help with molecular biology lab. I would like to thank Dr Svetlana Sedelnikova for helping me in the protein purification experiments and crystallography group members who have assisted me during my period of study.

All thanks going out to Libyan ministry of higher education for their financial support. I would like to thank my mum for her help and support all the time. I would also like to thank my wife for her support and encouragement.

List of Abbreviations

Crystallographic

Å	Ångström
a,b,c	Real space unit cell dimension
$a^* b^* c^*$	Reciprocal space unit cell dimension
α, β, γ	Real space unit cell angles
B-factor	Crystallographic temperature factor
F(S)	Structure factor
F_{calc}	Calculated structure factors
$ F_c $	Calculated structure factor amplitudes
F_{obs}	Observed structure factor
F (hkl)	Structure factor for reflection <i>hkl</i>
$ F (hkl) $	Structure factor amplitude reflection <i>hkl</i>
<i>h, k, l</i>	Reciprocal lattice points
I	Diffraction intensity
λ	wavelength
$\rho(xyz)$	Electron density at position <i>xyz</i>
R _{free}	Free R-factor
R _{merge}	R-factor relating agreement between symmetry related reflections
R _{msd}	Root mean square deviation

Biological and chemical

DEAE	Diethylaminoethyl
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
EDTA	Ethylenediaminetetraacetic acid
IPTG	Isopropyl β - D-1-thiogalactopyranoside
MES	2-(N-morpholino)ethanesulfonic acid
mRNA	Messenger ribonucleic acid
NADPH	Nicotinamide adenine dinucleotide phosphate
PEG	Polyethylene glycol
RNase	RNA nuclease

SDS	Sodium dodecyl sulfate
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
GluDH	Glutamate dehydrogenase
PheDH	Phenylalanine dehydrogenase
LeuDH	Leucine dehydrogenase

List of organisms

<i>Mycobacterium smegmatis</i>	<i>M. smegmatis</i>
<i>Clostridium symbiosum</i>	<i>C. symbiosum</i>
<i>Bacillus sphaericus</i>	<i>B. sphaericus</i>
<i>Escherichia coli</i>	<i>E. coli</i>

Table of content

Chapter one: Introduction	1
1.1 Amino acid dehydrogenases.....	1
1.1.1 Glutamate dehydrogenase (GluDH)	2
1.1.2 Phenylalanine dehydrogenase.....	2
1.2 The biological function of amino acid dehydrogenases	3
1.2.1 Biological role of GluDHs.....	3
1.2.1.1 Ammonia assimilation	4
1.2.1.2 Amino acid degradation	6
1.3 Molecular weight and quaternary structure.....	6
1.4 Structures and conformational changes of amino acid dehydrogenases.....	8
1.4.1 Structures of glutamate dehydrogenase	8
1.4.1.1 Bacterial GluDHs.....	8
1.4.1.2 Mammalian GluDHs.....	9
1.4.1.3 Conformational change of GDHs.....	9
1.4.2 Structure of Leucine dehydrogenase	10
1.4.3 PheDH structure	13
1.5 Coenzyme specificity of amino acid dehydrogenases	15
1.5.1 Coenzyme stereospecificity	17
1.6 Substrate specificity of amino acid dehydrogenases	17
1.6.1 Glutamate dehydrogenase.....	17
1.6.2 Phenylalanine dehydrogenase.....	20
1.6.3 Leucine dehydrogenase	20
1.7 Homology and sequence alignment of amino acid dehydrogenases	21
1.8 Chemical mechanism of GluDHs.....	22
1.9 Chemical mechanism for phenylalanine dehydrogenase	27
1.10 Previous structures of amino acid dehydrogenases.....	29
1.11 Project Aim.....	29
Chapter Two: Materials and Methods	36
2.0 Overview.....	36
2.1 Gene amplification, cloning and overexpression of glutamate dehydrogenase from <i>M. smegmatis</i>	36
2.1.1 The design of primers.....	36
2.1.2 Amplification of target gene using polymerase chain reaction (PCR).....	36
2.1.2.1 Purification of PCR products from agarose gel	38

2.1.3 Preparation of pET21a plasmid	38
2.1.4 Restriction digestion	39
2.1.5 Ligation of PCR product into plasmid.....	40
2.1.5.1 Transformation of ligation mixes into DH5 α E. coli.....	40
2.1.5.2 Screening of possible clones.....	41
2.1.6 Protein Expression.....	41
2.1.6.1 BL21 (DE3) transformation protocol.....	41
2.1.6.2 Small-scale overexpression trials.....	41
2.1.6.2.1 Expression Level and Solubility Test	42
2.1.6.3 Large-scale over-expression.....	42
2.1.6.4 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE).....	44
2.2 Purification of all target proteins; <i>Myobacterium smegmatis</i> GluDH, <i>Clostridium symbiosum</i> GluDH and phenylalanine dehydrogenase (PheDH) from <i>Bacillus sphaericus</i>.	47
2.2.1 Cell Lysis.....	47
2.2.2 Ion Exchange Chromatography	47
2.2.3 Hydrophobic Chromatography.....	48
2.2.4 Gel Filtration.....	48
2.3 Protein crystallization	49
2.3.1 Robot screening	49
2.3.2 Optimization trials	49
2.4 X-ray Crystallography	49
2.4.1 Protein crystallization	51
2.4.1.1 Vapour diffusion method	52
2.4.1.1.1 Hanging drop vapour diffusion method.....	53
2.4.1.1.2 Sitting drop vapour diffusion	53
2.4.2 Cryocooling	53
2.4.2 X-ray generation	55
2.4.2.1 Rotating anode generator.....	55
2.4.2.2 Synchrotron radiation	57
2.5 Data processing.....	58
2.6 Determination of Phases by Molecular Replacement.....	60
2.6.1 Rotation Function.....	60
2.6.2 Translation Function.....	61
2.6.3 Refinement and Re-building.....	61
2.6.4 Rigid Body Refinement	62
2.6.5 Monitoring the Progress of Refinement	62

Chapter Three: Cloning, Expression, Purification, Crystallization, Data Collection and Structure Determination of <i>M. smegmatis</i> GluDH.	64
3.0 Overview	64
3.1 Identification, amplification, cloning, overexpression, purification of <i>M. smegmatis</i> GluDH	64
3.1.1 Identification of Target Gene	64
3.1.2 PCR for amplification of <i>M. smegmatis</i> GluDH gene (msmeg_5442)	64
3.1.3 Cloning of <i>M. smegmatis</i> GluDH gene (msmeg_5442)	65
3.1.4 Identification of possible clones	65
3.1.5 Expression of <i>M. smegmatis</i> GluDH	65
3.1.5.1 small-scale over-expression	65
3.1.5.2 Large-scale over-expression	68
3.1.6 Purification of GluDH from <i>M. smegmatis</i>	68
3.1.6.1 Ion exchange chromatography	68
3.1.6.2 GluDH assay	68
3.1.6.3 Hydrophobic and gel filtration chromatography	70
3.2 Crystallization of glutamate dehydrogenase from <i>Mycobacterium smegmatis</i>	70
3.3 Data collection from crystals of <i>M. smegmatis</i> GluDH in apo state	73
3.4 Data collection and processing of GluDH -NADPH-glutamate complex crystals	73
3.4.1 Data collection	73
3.4.2 Data processing	75
3.5 Molecular replacement and model building of the <i>M. smegmatis</i> GluDH in complex with NADPH	75
3.6 Structure of 2-oxoglutarate-NADP-GluDH complex	78
3.6.1 Data collection and processing for low resolution data	78
3.6.2 Molecular replacement and structure building for low resolution data	85
3.6.3 Data collection and processing for higher resolution data	87
3.6.4 molecular replacement structure building for high-resolution data	87
3.6.4.1 Refinement of the substrate	88
3.7 Glutamate dehydrogenase from <i>Clostridium symbiosum</i>	93
.....	94
3.7.1 Purification of glutamate dehydrogenase of <i>Clostridium symbiosum</i>	96
3.7.2 Crystallization of glutamate dehydrogenase from <i>C. symbiosum</i>	96
3.7.3 Data collection and processing for crystals of <i>C. symbiosum</i> GluDH with L-glutamate	100

3.7.4 Molecular replacement and model building for the binary complex of <i>C. symbiosum</i> GluDH with glutamate.....	100
3.7.5 Data collection and processing for <i>C. symbiosum</i> GluDH soaked in NAD ⁺	103
3.7.6 Molecular replacement and model building for binary complex of <i>C. symbiosum</i> GluDH with NAD ⁺	103
Chapter four: Structural analysis of glutamate dehydrogenase from <i>M. smegmatis</i> and <i>C. symbiosum</i>	107
4.0 Introduction	107
4.1 Subunit structure of <i>M. smegmatis</i> GluDH	107
4.2 Subunit assembly	108
4.3 Alternative conformations of Serine residues in a ternary complex model of <i>M. smegmatis</i> GluDH	108
4.4 Binding of the dinucleotide with the enzyme	113
4.4.1 Comparisons between dinucleotide binding dehydrogenases.....	113
4.4.2 Recognition of the adenine ribose 3'OH	113
4.4.3 Binding the nicotinamide ring and associated ribose moiety and pyrophosphate	114
4.4.4 Comparison between the binding of dinucleotide in the NADPH binary complex and NADP ⁺ and 2oxoglutarate ternary complex of <i>M. smegmatis</i> GluDH.....	117
4.4.5 Coenzyme specificity.....	117
4.5 Conformational change of GluDH from <i>C. symbiosum</i> and <i>M. smegmatis</i>	126
4.5.1 Open and closed conformations in <i>C. symbiosum</i> GluDH	126
4.5.2 K89L mutant of apo-enzyme of <i>C. symbiosum</i> GluDH present in a closed form...	128
4.6 Catalytic pocket.....	131
4.6.1 Substrate binding in GluDH.....	131
Chapter five: Purification, crystallization, data collection, structural determination and structural analysis of phenylalanine dehydrogenase from <i>Bacillus sphaericus</i>.	141
5.0 Introduction	141
5.1 Purification of phenylalanine dehydrogenase (PheDH)	141
5.1.1 Purification of single mutant (N145A) PheDH	141
5.1.2 Purification of wild type protein.....	143
5.2 Crystallization of <i>Bacillus sphaericus</i> PheDH.....	143
5.2.1 Crystallization of single mutant (N145A) PheDH with NAD ⁺	143

5.2.2 Crystallization of wild-type PheDH with NADH.....	146
5.3 Data collection and processing of Phenylalanine Dehydrogenase from <i>Bacillus Sphaericus</i>	146
5.3.1 Single mutant (N145A) of phenylalanine dehydrogenase	146
5.3.2 Binary complex of wild type PheDH with NADH.....	149
5.4 Molecular replacement and model building.....	149
5.5 Analysis of the <i>B. sphaericus</i> PheDH structure.....	152
5.5.1 Tertiary structure of PheDH subunit.....	152
5.5.2 Quaternary structure	152
5.5.3 Dinucleotide binding site	155
5.5.3.1 Mode of binding of NADH with <i>B. sphaericus</i> PheDH enzyme	155
5.5.4 Domain motion.....	157
5.5.5 Active site.....	161
5.5.6.1 Single mutant PheDH (N145A).....	164
5.5.6.2 Triple mutant proteins N145A, G124A and L307V	165
Chapter Six: Cloning, Over-Expression and Purification of NDH-2 from <i>M. smegmatis</i>, a Model of <i>M. tuberculosis</i>	168
6.0 Introduction	168
6.1 Tuberculosis:	168
6.2 Epidemiology of tuberculosis:.....	168
6.3 Treatment of Tuberculosis:	170
6.4 Mechanism of action of anti-tuberculosis agents:.....	170
6.4.1 Mechanism of action of isoniazid and rifampin:	170
6.5 Drug resistant tuberculosis:	171
6.6 Development of new antibiotics:	172
6.6.1 Structure-based drug design:.....	172
6.7 Organization of respiratory chain of <i>Mycobacterium tuberculosis</i>:	173
6.7.1 Respiratory chain (Electron Transport Chain):	173
6.7.2 Electron transport chain of <i>M. tuberculosis</i> :	173
6.7.3 Type II NADH: Menaquinone oxidoreductase (NDH-2) of <i>M. tuberculosis</i> :.....	174
6.8 Aims of project	175
6.9 Identification of <i>Mycobacterium tuberculosis</i> <i>ndh</i> gene, which encodes NDH-2 protein:	177
Chapter seven: Conclusion and future work	186
7.1 Conclusion.....	186

7.1.1 Introduction.....	186
7.1.2 Conformational changes in GluDH.....	186
7.1.3 Catalytic pocket of GluDH.....	187
7.2 Phenylalanine Dehydrogenase.....	193
7.3 Cofactor specificity.....	194
7.4 Further Work.....	196
References.....	198
AppendixI: X-ray crystallography Theory.....	208

List of Figures

Chapter One

1.1 Ammonium assimilation by glutamine synthetase (GS) with glutamate synthase (GOGAT) pathway and glutamate dehydrogenase-NADP ⁺ -linked (GDH1) with glutamate dehydrogenase-NAD ⁺ -linked (GDH2).....	5
1.2 Amino acid metabolism fate of their carbon skeletons.....	7
1.3 Schematic diagram showing one subunit for GluDH from <i>C. symbiosum</i> and bovine.....	11
1.4 The overall organization of <i>bovine</i> GluDH hexamer.....	12
1.5 A diagram displaying the domain motion in GDH from.....	14
1.6 Schematic diagram representing one subunit for LeuDH from <i>B. sphaericus</i>	16
1.7 The overall organization of <i>B. sphaericus</i> LeuDH octamer with each subunit individually coloured.....	16
1.8 Schematic diagram representing a single subunit for PheDH from <i>Rhodococcus sp. M</i>	18
1.9 Two subunits come together to form the dimer around the two-fold axis, mediated via β strands of N-terminal domain. a) <i>C. symbiosum</i> GluDH, b) <i>Rhodococcus sp. M</i> . PheDH and c) <i>B. sphaericus</i> LeuDH.....	19
1.10 Sequence alignment of <i>C. symbiosum</i> GDH with <i>B. stearothermophilus</i> LeuDH and <i>B. sphaericus</i> PheDH and <i>T. intermedius</i> PheDH.....	23
1.11 Glutamate dehydrogenase chemical mechanism proposed by Rife and Cleland.....	26
1.12 Modified chemical mechanism proposed by Srinivasan and Fisher.....	28
1.13 Glutamate dehydrogenase chemical mechanism proposed by Stillman <i>et al.</i>	30
1.14 Modified chemical mechanism for phenylalanine dehydrogenase.....	32

Chapter Two

2.1 pET21a vector map.....	43
2.2 pET14b vector map.....	43
2.3 Ion-exchange chromatography.....	50
2.4 Gel filtration chromatography.....	50
2.5 Phase diagram of protein crystallization.....	54
2.6 Vapour diffusion methods.....	56
2.7 X-ray emission spectra generated using a copper target.....	59

2.8 Schematic of a synchrotron radiation source.....	59
--	----

Chapter Three

3.1 Electrophoretic pattern of agarose gel showing PCR products of GluDH gene from <i>M. smegmatis</i>	67
3.2 A picture of 1% agarose gel electrophoresis showing the screening results for the possible constructs of pET21a with <i>M. smegmatis</i> gene after digestion with NdeI and Hindiii.....	67
3.3 SDS gel showing the level of <i>M. Smegmatis</i> GluDH overexpression and solubility at different temperatures.....	69
3.4 An SDS gel displaying the level of <i>M. smegmatis</i> GluDH expression during large-scale growth.....	69
3.5 An SDS gel displaying the level of <i>M. smegmatis</i> GDH expression during large-scale growth.....	71
3.6 SDS-PAGE illustrates the purification of GluDH using Phe-HP and gel filtration columns.....	72
3.7 Chromatogram pictures for the purification of <i>M. smegmatis</i> GluDH.....	72
3.8 Photographs of <i>M. Smegmatis</i> GluDH crystals.....	74
3.9 Diffraction picture for binary complex <i>M. smegmatis</i> GluDH with NADPH and L-glutamate.....	76
3.10 Ramachandran plot for the final structure of <i>M. smegmatis</i> GluDH + NADPH	80
3.11 Main chain and side chain parameters for the final structure of <i>M. smegmatis</i> GDH + NADPH.....	81
3.12 Electron density map at different regions in the final model of <i>M. smegmatis</i> GluDH + NADPH	82
3.13 Diffraction images for 2-oxo-glutarate-NADP GluDH complex dataset at 2.63Å	83
3.14 Diffraction images for 2-oxo-glutarate-NADP GDH complex dataset at 1.64Å	84
3.15 Side-chain and main-chain parameters for the final model of GluDH-NADP-2-oxoglutarate from <i>M. smegmatis</i>	90
3.16 Ramachandran plot for the final model of ternary complex model of <i>M. smegmatis</i> GluDH + NADP ⁺ + 2oxoglutarate	91
3.17 shows the electron density maps around the cofactor (NADP ⁺)	92

3.18 Electron density for the substrate of the ternary complex of <i>M. smegmatis</i> GDH + NADP ⁺ + 2oxoglutarate.....	94
3.19 SDS-PAGE analysis and chromatogram trace for ion exchange column of <i>C. symbiosum</i> GluDH.....	97
3.20 SDS-PAGE analysis and chromatogram trace for hydrophobic column of <i>C. symbiosum</i> GluDH.....	98
3.21 SDS-PAGE analysis and chromatogram trace for gel filtration of <i>C. symbiosum</i> GluDH.....	99
3.22 Ramachandran plot for <i>C. symbiosum</i> GluDH in complex with L-glutamate.....	102
3.23 The electron density map around Val204.....	102
3.24 Ramachandran plot for binary complex of <i>C. symbiosum</i> GluDH with NAD ⁺	106

Chapter Four

4.1 Schematic illustrating the assembly of <i>M. smegmatis</i> GluDH into the active hexamer with 32-symmetry.....	109
4.2 The overall organization of <i>M. smegmatis</i> GluDH hexamer with each subunit individually coloured	110
4.3 Schematic diagram illustrates the overall organization of single subunit of <i>M. smegmatis</i> GluDH.....	111
4.4 A schematic diagram showing the connection of the secondary structure elements.....	111
4.5 The two conformations of the side chain of Ser16 and Ser339 in the ternary complex <i>M. smegmatis</i> GluDH.....	112
4.6 A schematic diagram illustrating the interaction of the $\beta\alpha\beta$ unit with the cofactor.....	115
4.7 A schematic of the interaction of the nicotinamide ring in the binary complex structure (GluDH-NADPH) and ternary complex model (GluDH-NADP ⁺ -2ox-glutarate) for <i>M. smegmatis</i> GluDH.....	118
4.8 Schematic diagram of the interactions made by the carboxamide of the nicotinamide ring with neighbouring groups <i>M. smegmatis</i> GluDH/NADP ⁺ /gem-diol and <i>C. symbiosum</i> GluDH/NAD ⁺	119
4.9 Schematic diagram of the 2' and 3'OH of the nicotinamide ribose forming hydrogen bonding interactions with adjacent residues for <i>M. smegmatis</i> GluDH/NADP ⁺ /gem-diol and <i>C. symbiosum</i> GluDH/NAD ⁺	119

4.10 Schematic diagram illustrating the pyrophosphate binding	120
4.11 Sequence alignment of <i>M. smegmatis</i> GluDH and <i>C. symbiosum</i> GluDH	122
4.12 Shows the three positively charged residues (Lys137, Lys285 and Arg291) and the Ser266, which are adjacent to 2'phosphate group in the GluDH-NADP ⁺ - 2-oxoglutarate from <i>M. smegmatis</i>	124
4.13 Sequence alignment of a number of NAD ⁺ -dependent GluDHs and NADP ⁺ -dependent GluDHs	125
4.14 The binding site of the NADP ⁺ of <i>M. smegmatis</i> GluDH with the residues implicated in interacting with the 2'phosphate group	127
4.15 Illustrates the orientation of the arginine residue (R205) from <i>C. symbiosum</i> GluDH.....	133
4.16 Electron density for the L-glutamate of 1.2Å binary complex of <i>C. symbiosum</i> GluDH with L-glutamate.....	135
4.17 Shows the binding of the side chain of R209 with γ-carboxyl group of the gem-diol and the carbonyl group of the carboxamide of the nicotinamide of the NADP ⁺	137
4.18 The two conformations of γ-carboxyl of the L-glutamate in the <i>C. symbiosum</i> GluDH binary complex	137
4.19 Shows the distance between the C4 of the nicotinamide and the alpha carbon for the two substrates, L-glutamate and gem-diol	138
4.20 Interactions of possible diol with the adjacent groups.....	139
4.21 Revised chemical mechanism of GluDH in the direction of oxidative deamination of L-glutamate.....	140

Chapter Five

5.1 SDS-PAGE gel showing the purification progress of single mutant PheDH protein	142
5.2 Chromatogram pictures for the purification of single mutant PheDH protein	142
5.3 SDS-PAGE gel showing the purification steps of wild type PheDH	144
5.4 Chromatogram traces for the purification of wild type PheDH	144
5.5 Photographs of Robot Screen crystals of single mutant <i>B. sphaericus</i> PheDH with NAD ⁺	145
5.6 Pictures of robot screen crystals of wild-type <i>B. sphaericus</i> PheDH with NADH	147
5.7 Photographs of crystals for wild-type <i>B. sphaericus</i> PheDH with NADH.....	148

5.8 Structural sequence alignment of the <i>B. sphaericus</i> PheDH, <i>B. sphaericus</i> LeuDh and <i>M. smegmatis</i> GluDH	153
5.9 Schematic diagram shows the overall organization of one chain of <i>B. sphaericus</i> PheDH and <i>M. smegmatis</i> GluDH	154
5.10 Space filling illustration of the octameric structures, of PheDH (a and b) and LeuDh (c and d) from <i>B. sphaericus</i>	156
5.11 A schematic diagram of one subunit of <i>B. sphaericus</i> PheDH showing the cofactor location	158
5.12 A schematic depicts the $\beta\alpha\beta$ motif ($\beta\epsilon$ - $\alpha 6$ - βf) of <i>B. sphaericus</i> PheDH.....	158
5.13 A stereoview of an interaction of adenine ring in the binary complex structure of <i>B. sphaericus</i> PheDH	159
5.14 Schematic illustrating the pyrophosphate binding with the adjacent groups	159
5.15 The interaction of nicotinamide ring and associated ribose to the neighboring groups in the binary complex structure of <i>B. sphaericus</i> PheDH with NADH.....	160
5.16 Shows the distance between the C4 of the nicotinamide and the alpha carbon of the modelled L-phenylalanine	162
5.17 Structure-based sequence alignment of PheDH and LeuDh from <i>B. sphaericus</i> , <i>Rhodococcus sp.</i> M4 PheDH and <i>C. symbiosum</i> GluDH	166
5.18 A tyrosine substrate modeled into the active site of <i>B.sphaericus</i> PheDH.....	167

Chapter Six

6.1 The cell wall of <i>Mycobacterium tuberculosis</i>	169
6.2 Putative organization of the <i>M. tuberculosis</i> respiratory chain	176
6.3 Sequence alignment of <i>M. tuberculosis</i> NDH-2 and its homologues from <i>M. smegmatis</i> NDH-2	180
6.4 Electrophoretic pattern of agarose gel showing PCR products of <i>ndh</i> gene from <i>M. smegmatis</i>	180
6.5 A picture of a 1% AGE gel showing the screening results for the recombinant pET14b with the <i>M. smegmatis</i> <i>ndh</i> gene that digested by NdeI and BamHI.....	181
6.6 Sequence alignment of <i>M. tuberculosis</i> NDH-2 with <i>Caldalkalibacillus Thermarum</i> NDH-2	183
6.7 10% SDS.PAGE analysis of purification of <i>M. smegmatis</i> NDH-2	184
6.8 Chromatogram picture for the purification of NDH-2 from <i>M. smegmatis</i> using Ni-NTA column.....	184

6.9 10% SDS.PAGE analysis of purification of <i>M. smegmatis</i> NDH-2 protein.....	185
6.10 Chromatogram picture for the purification of NDH-2 from <i>M. smegmatis</i> using gel filtration column	185

Chapter Seven

7.1 A schematic diagram shows the interaction of the side chain carboxyl group of L-glutamate in binary complex (1.24Å) of the <i>C. symbiosum</i> GluDH with glutamate and 1BGV	188
7.2 Shows the different position of the α -carboxyl group of L-glutamate in <i>C. symbiosum</i> GluDH/L-glutamate (1.24Å) and 1BGV	188
7.3 Illustrates the different orientations of arginine residue in GluDHs	190
7.4 Sequence alignment of a number of GluDHs	190
7.5 Shows the binding of the side chain of R209 with γ -carboxyl group of the gem-diol and carbonyl group of the carboxamide of the nicotinamide of the NADP ⁺	191
7.6 Shows the distance between the C4 of the nicotinamide and the alpha carbon for the two substrates, L-glutamate (1BGV) and gem-diol	191
7.7 A schematic diagram shows the D169 in <i>M. smegmatis</i> GluDH ternary complex	192
7.8 Sequence alignment of a number of NAD ⁺ -dependent GluDHs and NADP ⁺ -dependent GluDHs	195

List of Tables

Chapter one

1.1 Structures of some amino acid dehydrogenases in the PDB.....	35
--	----

Chapter two

2.1 Minimal Seleno-L-methionine (SeMet) media composition	45
---	----

Chapter Three

3.1 Data collection statistics for <i>M. smegmatis</i> GDH in complex with NADPH and L-glutamate	76
--	----

3.2 The final refinement parameters for binary complex of <i>M. smegmatis</i> GluDH with NADPH	79
--	----

3.3 Data-collection and processing statistics for 2.63Å resolution date	83
---	----

3.4 Data-collection and processing statistics for 1.6Å resolution date	84
--	----

3.5 The final refinement parameters for ternary complex of <i>M. smegmatis</i> GDH with NADP ⁺ and 2-oxo-glutarate, 2.63 Å resolution	86
--	----

3.6 The final refinement parameters for ternary complex of <i>M. smegmatis</i> GluDH with NADP ⁺ and 2-oxo-glutarate, 1.64Å resolution	89
---	----

3.7 B factor values for the 2-oxoglutarate and gem-diol	95
---	----

3.8 Data-collection and processing statistics for <i>C. symbiosum</i> GluDH in complex with L-glutamate	101
---	-----

3.9 The final refinement parameters for binary complex of <i>C. symbiosum</i> GluDH with L-glutamate	101
--	-----

3.10 Data-collection and processing statistics for <i>C. symbiosum</i> GluDH with NAD ⁺	105
--	-----

3.11 The final refinement parameters for binary complex of <i>C. symbiosum</i> GluDH with NAD ⁺	105
--	-----

Chapter Four

4.1 Hydrogen bonds between NAD ⁺ (P ⁺) and GluDH enzymes	121
---	-----

4.2 The domain motion in GluDHs	129
---------------------------------------	-----

4.3 The conserved and equivalent residues that are responsible for the substrate binding for GluDHs	135
---	-----

4.4 The hydrogen bonds for the substrate interaction to the ternary complex of <i>M. smegmatis</i> GluDH and <i>C. symbiosum</i> GluDH binary complex (IBVG)	138
--	-----

Chapter Five

5.1 Data-collection and processing statistics for the single mutant PheDH crystallized with the cofactor	150
5.2 Data-collection and processing statistics for the binary complex of the wild-type <i>B. sphaericus</i> PheDH with NADH	150
5.3 Hydrogen bonds between NADH and <i>B. sphaericus</i> PheDH	160
5.4 The residues responsible for the substrate binding of GluDHs, PheDH and LeuDH.....	167

Chapter one: Introduction

1.1 Amino acid dehydrogenases

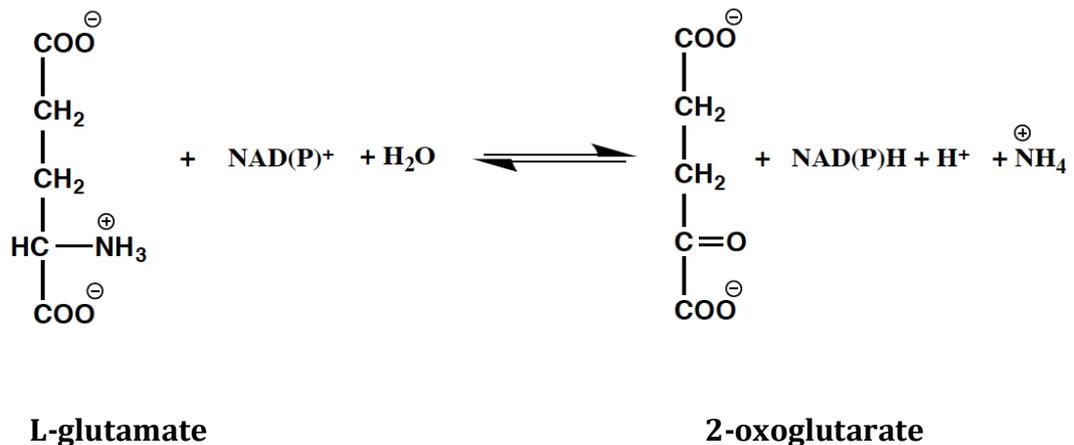
Amino acid dehydrogenase enzymes catalyse the oxidative deamination of amino acids to their corresponding 2-oxo acids with associated reduction of NAD(P⁺) to NAD(P)H. This family of enzymes belongs to the oxidoreductases, that catalyze the transfer of electrons from one molecule (electron donor) to another (electron acceptor).



Amino acid dehydrogenases have a substantial industrial potential, due to their applications in diagnostic kits to screen serum levels of amino acids that accumulate in a range of metabolic diseases (Nakamura et al, 1996). Also these enzymes are used in the synthesis of pure form of amino acids that have been used in pharmaceutical and food industries (Baker et al, 1997; Hummel & Kula, 1989) (Paradisi et al, 2007). The sequence homology between all members of Glu, Leu, Phe and Val dehydrogenase, obviously indicates the existence of an enzyme superfamily that are related, but with some differences (Britton et al, 1993b; Takada et al, 1991). For example; phenylalanine dehydrogenase (PheDH) and leucine dehydrogenase (LeuDH) from *Bacillus sphaericus* show significant amino acid sequence homology to each other, where the identity between them is about 50%. However, there is a low but important degree of homology between LeuDH and PheDH on the one hand and GluDH on the other (Britton *et al*, 1993b; Britton et al, 1992). Despite this homology, amino acid dehydrogenases show different substrate specificities. In general, glutamate dehydrogenase prefers the glutamate as a substrate to all other amino acids (Syed et al, 1991), while LeuDHs catalyses the oxidation of aliphatic amino acids only (Ohshima et al, 1978; Ohshima et al, 1994). Phenylalanine dehydrogenases have a preference for aromatic amino acids as a substrate, but they also show some activity with nonpolar aliphatic amino acids (Asano et al, 1987a; Baker et al, 1997).

1.1.1 Glutamate dehydrogenase (GluDH)

Glutamate dehydrogenases (GluDHs) are abundant enzymes that occupy a critical branch point between carbon and nitrogen metabolism. These enzymes catalyze the reversible oxidative deamination of L-glutamate to 2-oxoglutarate and ammonia with concomitant reduction of NAD⁺ or NADP⁺. Glutamate dehydrogenases are significant enzymes, due to the vital position in metabolism occupied by glutamate and 2-oxoglutarate, and the ability of these compounds to take part in several kinds of metabolism pathways, for example, the citric acid cycle and urea cycle.



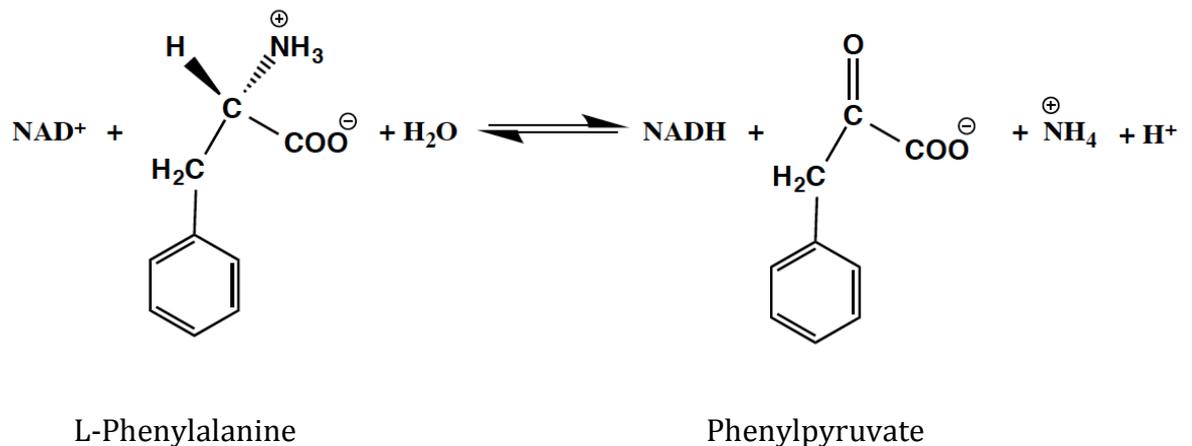
The above reaction allows the inter-conversion of L-glutamate to 2-oxoglutarate and ammonia and in coupling with other aminotransferases GluDH acts as a clearing house for α -amino groups (Braunstein, 1957).

1.1.2 Phenylalanine dehydrogenase

Phenylalanine dehydrogenase was discovered by Hummel in 1984 in a strain of *Brevibacterium* sp. That was isolated from soil (Hummel et al, 1984). Thereafter, it was found in a number of bacterial strains including *Bacillus* (Asano et al, 1987), *Nocardia* (Pasquo et al, 1998), *Rhodococcus* (Misono et al, 1989), and *Mycobacterium* (Asano et al, 1998). Phenylalanine dehydrogenase has been used as a biocatalyst, for the production of enantiomerically pure non-natural amino acids, which have been used as drug precursors (Asano et al, 1990), and in the production of L-phenylalanine a component of the

artificial sweetener aspartame (Brunhuber et al, 2000). Moreover, this enzyme also has been used as a biosensor to monitor serum levels of L-phenylalanine in genetic disorders, such as phenylketonuria (PKU)(Huang et al, 1998).

Phenylalanine dehydrogenase catalyzes the reversible oxidative deamination of L-phenylalanine, and some of its analogues, to the corresponding keto acids in the presence of NAD⁺ as a coenzyme, and it acts in the first catabolic step of phenylalanine degradation in bacteria.



1.2 The biological function of amino acid dehydrogenases

Amino acid dehydrogenases play an important role in nitrogen and carbon metabolism. These enzymes link the metabolism of amino acids to the Krebs cycle by an oxidative deamination reaction that converts the amino acids to 2-keto acids, which act as metabolites for tricarboxylic acid (TCA) cycle (Brunhuber & Blanchard, 1994).

1.2.1 Biological role of GluDHs

GluDHs are widely distributed over the eukaryotic, bacterial and archaeobacterial kingdoms, where the function of GluDH in metabolic processes varies from organism to organism. In lower organisms, the GluDH role is commonly unidirectional with either a catabolic or anabolic function. The particular function of GluDHs in vertebrates is a debatable issue, and these enzymes are implicated in several pathways of metabolism (Goldin & Frieden, 1971).

- 1- GluDH from mammalian sources is involved in the reaction that incorporates nitrogen into an organic compound.
- 2- L-glutamate acts as a nitrogen donor for the transamination reaction that converts oxaloacetate to aspartate. Because the α -amino group at aspartate is lost through the urea cycle, GluDH also provides a link to nitrogen excretion.
- 3- The conversion of L-glutamate and 2-oxoglutarate leads to a flux of carbon skeletons between the Krebs cycle and protein biosynthesis.

Generally, glutamate dehydrogenases can be divided into one of the metabolic categories; glutamate catabolism or glutamate biosynthesis (ammonia assimilation)

1.2.1.1 Ammonia assimilation

In most cells, ammonia enters into an organic compound by one or more of three main reactions. In eukaryotes, nitrogen assimilation can be accomplished by carbamoyl-phosphate synthetase, using ammonia, carbon dioxide and ATP, to produce carbamoyl-phosphate. In this reaction, nitrogen can be added into pyrimidines and by the urea cycle lead, via citrulline, to the incorporation of nitrogen into arginine (Berg et al, 2002).

In prokaryotes, there are two pathways for ammonium assimilation: under limiting nitrogen conditions the glutamine synthetase (GS) and glutamate synthase (GOGAT) cyclic pathway (figure1.1a) is activated because of the affinity of GS for ammonium is quite high (Harper et al, 2010). Glutamine synthetase catalyzes the formation of glutamine by condensation of ammonia and glutamate. Glutamine is a main nitrogen donor in the biosynthesis of several of organic compounds, such as purines, pyrimidines and a number of amino acids. The glutamine also can pass on its nitrogen in the transamination reaction catalyzed by glutamate synthetase.

The glutamate dehydrogenase route (figure1.1b) is activated under abundant nitrogen conditions and has a main assimilatory function due to the low affinity of GluDH for ammonium. Glutamate dehydrogenase catalyzes the reductive amination of 2-oxoglutarate to produce glutamate with oxidation of pyridine nucleotides NAD(P)H (Harper et al, 2010).

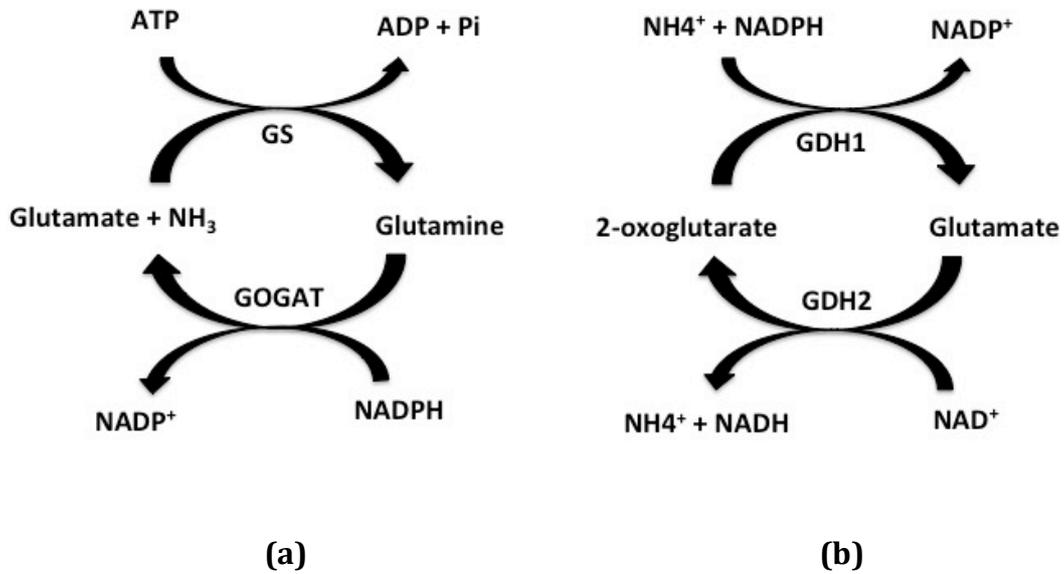


Figure 1.1. Illustrates ammonium assimilation by two routes **(a)** glutamine synthetase (GS) and glutamate synthase (GOGAT) pathway; **(b)** glutamate dehydrogenase-NADP⁺-linked (GDH1) and glutamate dehydrogenase-NAD⁺-linked (GDH2). This diagram is adapted from (Harper et al, 2010).

1.2.1.2 Amino acid degradation

Amino acids that are extra to those needed for protein synthesis are used as metabolic intermediates for energy, and are also utilized as building blocks of several biologically vital nitrogen-containing compounds, such as nucleotides, nucleotide coenzymes, active amines and glutathione (Voet et al, 2011).

The first step of amino acid catabolism is the transfer of the α -amino group by transaminase (or aminotransferase) enzymes to 2-oxoglutarate to form glutamate and the α -keto acid. The resultant glutamate can be oxidized by GluDH to yield ammonium ion and 2-oxoglutarate that is then ready to accept another α -amino group. The ammonium ion produced from this reaction is used for the biosynthesis of nitrogen compounds. In mammals, excess NH_4^+ is converted to urea for excretion. The fates of the carbon skeletons that are produced from amino acid catabolism reaction vary according to which one of the twenty amino acids has been degraded. The end result of the breakdown of the twenty fundamental amino acids is only seven molecules: pyruvate, acetyl CoA, acetoacetyl CoA, 2-oxoglutarate, succinylCoA, fumarate and oxaloacetate. All of these metabolites are able to enter to the citric acid cycle and generate energy. The metabolic pathways of the carbon skeletons are shown in figure 1.4. The amino acids can be sorted into three groups based on their metabolic fate. A ketogenic amino acid is the term used for those amino acids that break down to acetyl CoA or acetoacetyl CoA and can thus produce ketone bodies or fatty acids. Amino acids that are catabolized to pyruvate, 2-oxoglutarate, succinylCoA, fumarate or oxaloacetate are named glucogenic amino acids, because they are able to ultimately produce glucose (Berg et al, 2002).

1.3 Molecular weight and quaternary structure

GluDHs characterized so far have an oligomeric structure, and can be classified into two subgroups based on the subunit composition as either hexameric or tetrameric.

Hexameric GluDHs described to date are homo-hexamers, with subunit molecular weights between 48 kDa and 55 kDa. These enzymes from fungi and bacteria have a smaller molecular weight; for example, the subunit molecular weights of GluDHs from *Mycobacterium smegmatis* and *Peptoniphilus*

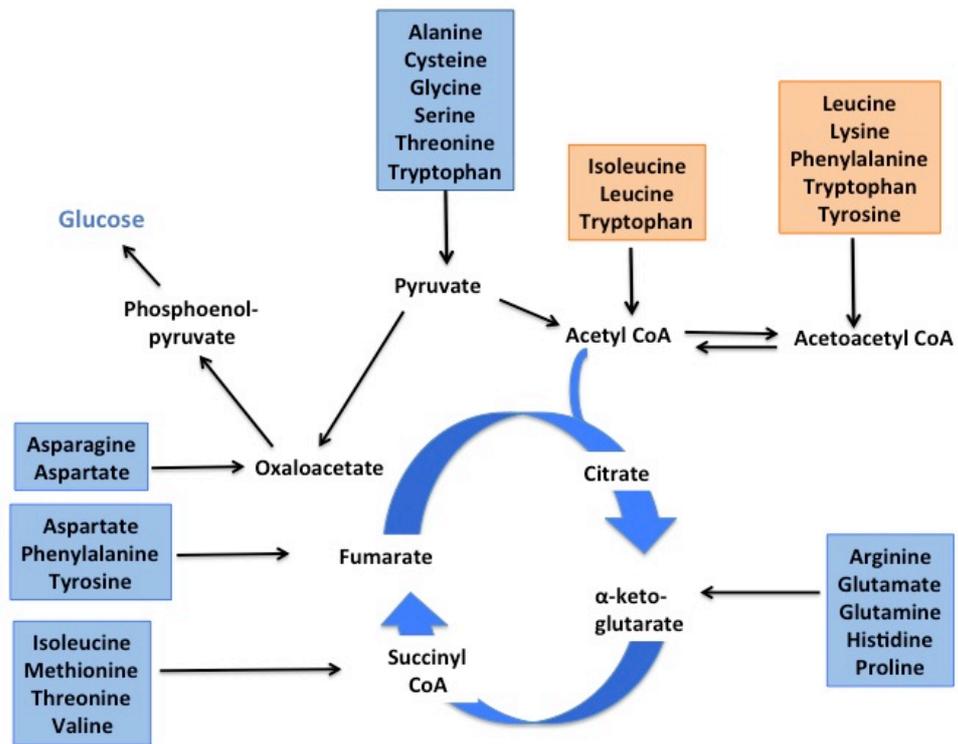


Figure 1.2. Amino acid metabolism and the fate of their carbon skeletons. Blue boxes represent glucogenic amino acids and orange boxes represent ketogenic amino acids. This figure is taken from (Berg et al, 2002).

asaccharolyticus are approximately 49 kDa (Sarada et al, 1980) (Hornby & Engel, 1984). On the other hand, vertebrate GluDHs such as bovine liver GluDHs have larger subunit molecular weights, which are about 55 kDa (Moon & Smith, 1973). GluDHs with tetrameric structure are considered as uncommon enzymes among GluDHs family. Only a few enzymes fall within this group, all of which are NAD⁺-linked, and all of those found to date have been isolated from fungal and protozoan parasite sources. The subunit molecular weight of the tetrameric GluDHs is around 115 kDa (Veronese et al, 1974).

The quaternary structure of PheDHs has substantial variation, and there is no clear majority oligomer. For example, the PheDHs that are characterized from *S. ureae*, *B. sphaericus* and *B. badius*, have been shown to be octamers, with molecular weights of 305 kDa, 340 kDa and 335 kDa respectively (Asano & Nakazawa, 1985; Asano et al, 1987a; Asano et al, 1987b). The structures of PheDHs from *Thermoactinomyces*, *Rhodococcus. sp M4*, *R. maris* and *Nocardia* are hexameric (270 kDa), tetrameric (150 kDa) (Ohshima et al, 1991), dimeric (70 kDa) (Misono et al, 1989) and monomeric (42 kDa)(Deboer et al, 1989), respectively.

The leucine dehydrogenases described to date, from a number of organisms, can be categorized into two oligomeric classes; either hexameric or octameric. For example, the quaternary structure of *B. stearothermophilus* enzyme is a hexamer of subunit molecular weight 245 kDa (Ohshima et al, 1985), whereas the LeuDH from *B. sphaericus* has an octomeric structure of 340 kDa (Turnbull et al, 1994).

1.4 Structures and conformational changes of amino acid dehydrogenases

1.4.1 Structures of glutamate dehydrogenase

1.4.1.1 Bacterial GluDHs

The three-dimensional crystal structure of the NAD⁺-linked GDH from *Clostridium symbiosum* was the first protein structure of an amino acid dehydrogenase to be determined. It was solved in both a free enzyme form, and with NAD⁺ using the X-ray diffraction (Baker et al, 1992b). The *C. symbiosum* GluDH is a hexameric protein made up from six identical subunits of 449

residues arranged in D3 point group symmetry. Each subunit is organized into two globular domains separated by a deep cleft. Domain I the N-terminal substrate binding domain is slightly larger than domain II and is involved in subunit assembly into the hexameric oligomer, and consists of residues 1-200 and 424 to 449. The polypeptide chain of domain I folds into a sequence of successive α/β secondary structure units. Mixed β sheets of parallel and anti-parallel beta-strands form the core of this domain and are flanked by α -helices on both sides. Domain I makes the main intersubunit interactions along the three-fold and two-fold axes, and the substrate binding pocket is located in a deep cleft that separates the two domains. The second domain (domain II) or C-terminal domain contains seven β -strands surrounded by α -helices and is analogous to the classical dinucleotide-binding fold that is known as Rossmann fold (Rossmann et al, 1974), with one an exception, namely that one beta strand is antiparallel to the others. This domain contains the NAD⁺-binding site and comprises the residues 201 to 423 (Baker et al, 1992b).

1.4.1.2 Mammalian GluDHs

The hexameric GluDHs from mammalian sources are generally similar to the bacterial protein, but they differ from each other due to the presence of an alpha-helical domain containing 48 residues in the mammalian GluDHs. This domain makes an antenna structure elongating from the top of the nucleotide binding domain and is near to the 3-fold axis of the hexamer. This antenna domain has been suggested to be involved in the allosteric control of the catalytic activity that is unique to the mammalian GluDH (Peterson & Smith, 1999; Smith et al, 2001). Studies on the allosteric regulation of the activity of GluDH from mammals indicate that GTP is a negative and ADP is a positive effector, respectively. (Bailey et al, 1982; Hudson & Daniel, 1993). This regulation might be assisted by the interactions of the subunit in the antenna domain (Peterson & Smith, 1999).

1.4.1.3 Conformational change of GDHs

The structures of the different GluDHs, in different complexes have shown that the two domains in the GluDH subunit exhibit conformational flexibility and vary between an open and a closed form. Although a closed form is required for

catalysis, the presence of substrate does not appear to be required for domain closure, indicating that the domains move more freely in solution (Figure 1.3). For example, in the structure of substrate-free GluDH from *C. symbiosum* and also its binary complex with cofactor, the two domains are organized in a relatively open orientation with respect to each other, which is essential for substrate binding (Baker et al, 1992a). The structure of the *Peptoniphilus asaccharolyticus* GluDH adopts the open conformation in the apo state, with the relative orientation of the two domains resembling that of the apo *C. symbiosum* GluDH (Oliveira et al, 2012). In contrast, the structure of GluDH from *E. coli*, in the free enzyme state has a closed form, but the conformation of this structure is not completely closed, when compared to the NADP⁺-binary complex of bovine GluDH (Sharkey et al, 2013). The crystal structure of *C. symbiosum* GluDH co-crystallized with L-glutamate is in a closed form, and this closed conformation is thought to be required for catalytic activity, in which the C4 in nicotinamide and the L-glutamate are in the appropriate orientation for hydride transfer and thus the oxidation/reduction reaction could occur (Burgi & Dunitz, 1983). The structural rearrangement between the two forms (open and closed form) can be described as a rigid body movement of roughly 14° about an axis being between the two domains. (Stillman et al, 1993). Furthermore, the motion of these domains with respect to each other is accompanied by substantial changes to the torsion angles of some crucial side-chains in the active site (residues 200-379 and 391-434) which enhance the interactions between the GluDH and its substrate (Stillman et al, 1993).

1.4.2 Structure of Leucine dehydrogenase

Structures have also been determined of other amino acid dehydrogenases, such as the three-dimensional structure of LeuDH from *Bacillus sphaericus* (Baker et al, 1995). There is no large difference between the structure of LeuDH from *B. sphaericus* and other superfamily members, such as *C. symbiosum* GluDH, especially in the folding arrangement of the polypeptide chain (Stillman et al, 1999). The structure determination of GluDH and LeuDH has shown that these enzymes are different in the quaternary structures, since

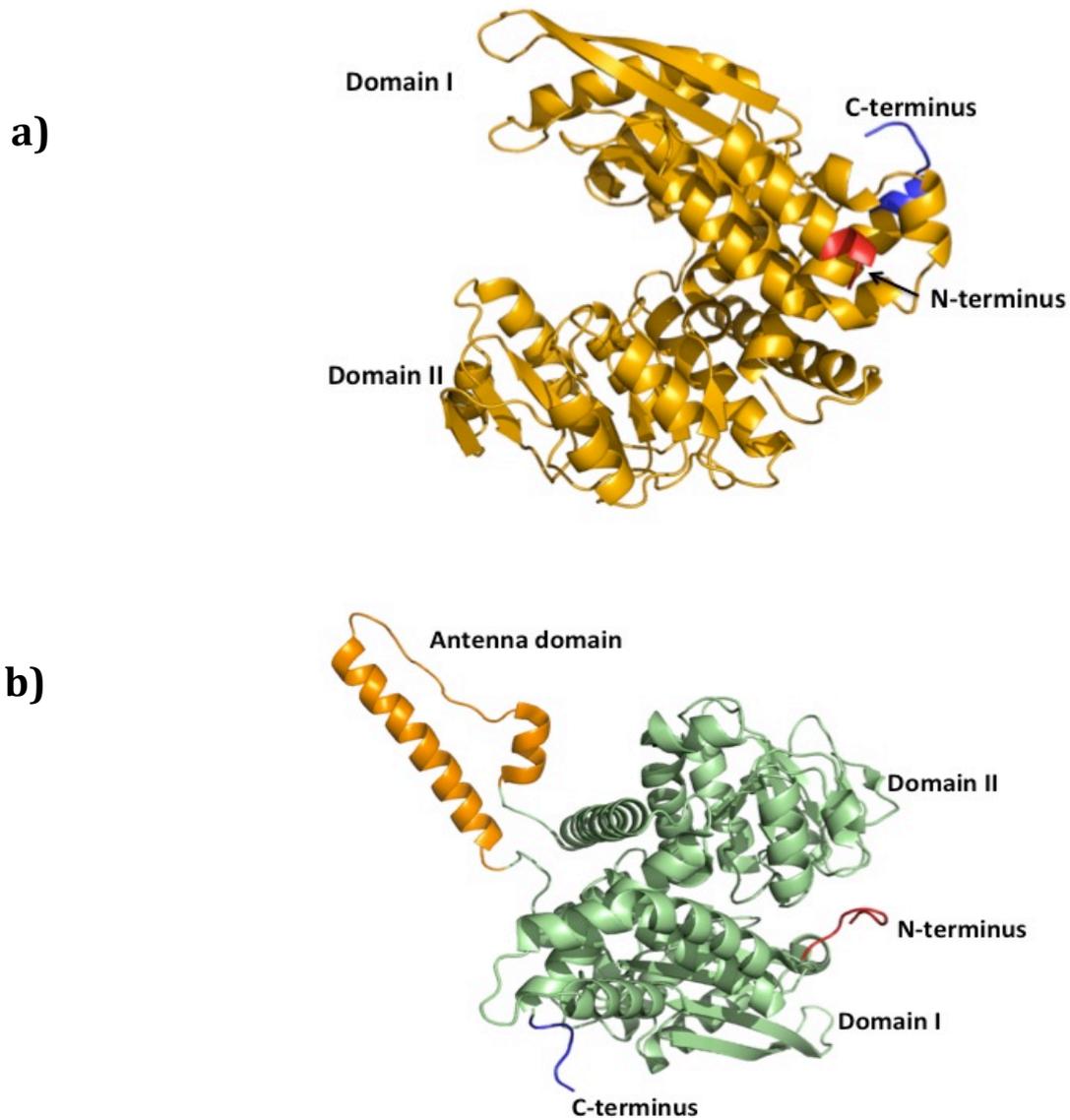


Figure 1.3. Schematic diagram showing the domain I and domain II with N-terminus in red color and C-terminus in blue color for glutamate dehydrogenase subunit of **a)** *C. symbiosum* (Accession number in PDB IHRD) and **b)** bovine enzyme, antenna domain in bovine GluDH is shown in orange color (Accession number in PDB 1HWZ).

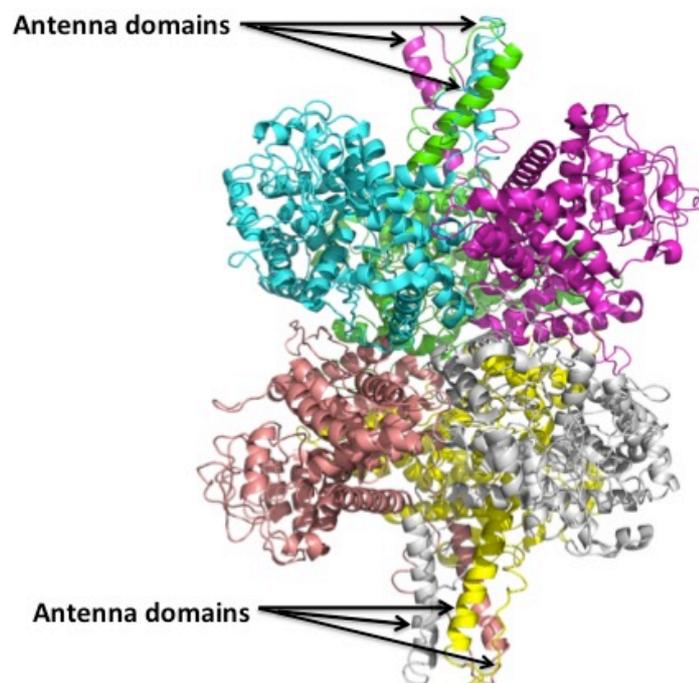


Figure1.4. The overall organization of *bovine* GluDH hexamer with each subunit individually coloured. Subunits assemble into the hexamer by the three-fold axis lying along the perpendicular axis. The antenna regions extending from the top of the domain II can be seen. Accession number in PDB 1HWZ.

GluDH adopts a hexameric structure, as mentioned above (Baker et al, 1992a; Stillman et al, 1993) while the quaternary structure of LeuDh is an octamer (Baker et al, 1995). Despite these differences, the dimeric structure of the LeuDh is very similar to that seen in GluDH, but in the case of the octameric LeuDh four dimers assemble to an octamer in D₄ symmetry, compared to the three dimers in the D₃ hexamer of GluDH.

The structure of *B. sphaericus* LeuDh (at resolution of 2.2Å) has shown that like GluDH, the subunit has two domains separated by a deep cleft. Domain I contains residues 1-136 and 332-364 and is comprised of six strands of a mixed beta-sheet, surrounded by α -helices on each side. Domain II (residues 137-331) is similar to the classical NAD⁺-binding domain (Rossmann fold), and includes six strands of parallel beta-sheet positioned in the centre and flanked by two helices on both sides (Baker et al, 1995). This is unlike the dinucleotide domain in *C. symbiosum* GDH that involves one beta-strand in the reverse direction (Baker et al, 1992a).

Structure analysis of *B. sphaericus* LeuDh (Baker et al, 1995), has shown that the enzyme's two domains can exist in different conformations, which offers more evidence in support the necessity of a conformational rearrangement in the catalytic activity of this enzyme (Baker et al, 1995).

Domain movement in this family is discussed in more detail in chapter 4 (section 4.5).

1.4.3 PheDH structure

For phenylalanine dehydrogenase, only that from *Rhodococcus sp. M4* has a published structure (Vanhook et al, 1999). This enzyme has been solved in ternary complexes to a resolution 1.5 Å and 2.3 Å for enzyme-NAD⁺-phenylpyruvate and enzyme-NAD⁺- β -phenylpropionate respectively. Analysis of both ternary complex structures has shown that this PheDH is a dimeric enzyme made up from two homo-subunits, again each with the standard amino acid dehydrogenase structure. The domains are separated by a deep groove containing the catalytic pocket (Vanhook et al, 1999). Although this structure is dimeric, the dimer is constructed in the same way as the dimer from the other family members (Vanhook et al, 1999).

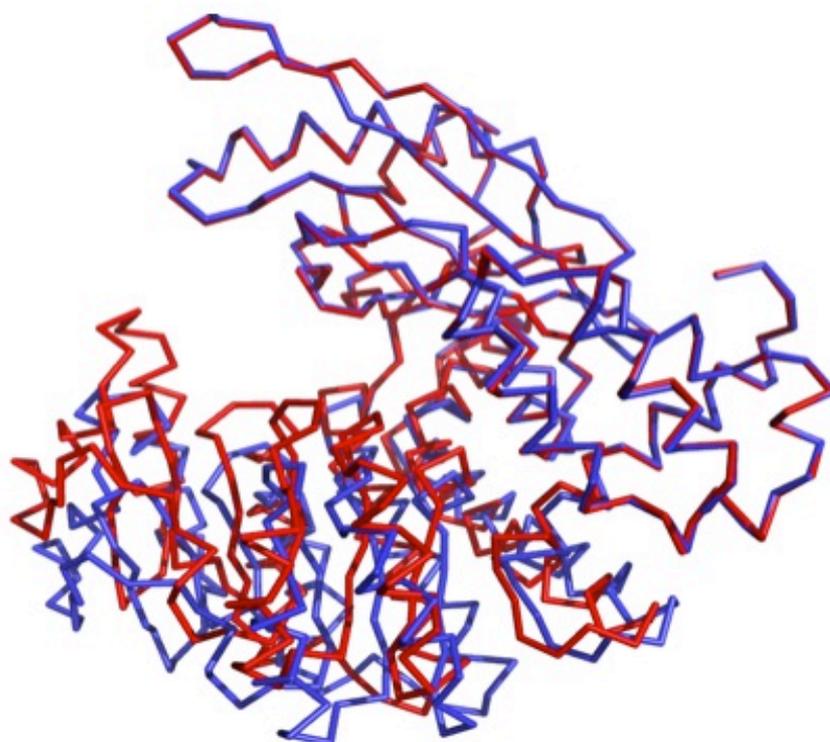


Figure 1.5. A diagram displaying the domain motion in GDH from *C. symbiosum*. The open form (apo-enzyme) chain is colored blue and the closed form (glutamate bound) chain is red. The subunits were superimposed on the basis of domain I residues using LSQ superpose in COOT.

1.5 Coenzyme specificity of amino acid dehydrogenases

One main characteristic of glutamate dehydrogenase enzymes is the distinction in cofactor specificity. Glutamate dehydrogenases from prokaryotes commonly utilize either NAD⁺ or NADP⁺ as coenzymes, whereas eukaryotic GluDHs have dual specificity and are able to act with both cofactors. The structural basis of coenzyme specificity has been proposed in a number of studies (Oliveira et al, 2012; Sharkey et al, 2013) and is addressed, later in this thesis (chapter 4, section 4.4.5). NADP⁺ dependent GluDHs are usually involved in processes of ammonia assimilation, by incorporating the ammonia into 2-oxoglutarate to produce glutamate (anabolic function) (Ohshima & Soda, 1990). The gene transcription of bacterial NADP⁺ dependent GluDH can be controlled by a change of growth conditions, such as presence of carbon and nitrogen (Hansler et al, 2009; Schwacha & Bender, 1993) In contrast, NAD⁺ dependent GluDHs are generally involved in processes of oxidative degradation of glutamate (catabolic function) (Ohshima & Soda, 1990), and do not seem to be regulated in response to nitrogen availability (Chavez & Candau, 1991; Harper et al, 2010). However, dual specificity GluDHs utilize the two cofactors and are involved in both catabolic and anabolic functions. These enzymes show a similar affinity for NAD⁺ and NADP⁺ in some species (such as chicken GluDH), whereas with both dogfish liver and frog liver GluDHs the affinity for NAD⁺ is much greater than NADP⁺ (Corman & Kaplan, 1967; Fahien et al, 1965).

All phenylalanine dehydrogenases and leucine dehydrogenases characterized so far require NAD⁺ as a coenzyme for the oxidative deamination reaction. However, a number of alternative nucleotide substrates have been analysed with PheDH from *Rh. maris* (Misono et al, 1989) and LeuDH from *B. sphaericus* enzyme (Ohshima et al, 1978). The reactivity of NAD⁺ analogues with PheDH was as following: the activity of the PheDH with NADP⁺ is very low (about 6.6% of the NAD⁺ reactivity), this is contrast to the relative rate of 3-acetylpyridine-NAD⁺ that was much higher (241%) compared to that with NAD⁺ (100%), while the reactivity of deamino-NAD⁺ and thionicotinamide-NAD⁺ were 86% and 101% respectively, which are similar to NAD⁺ (Misono et al, 1989). In the

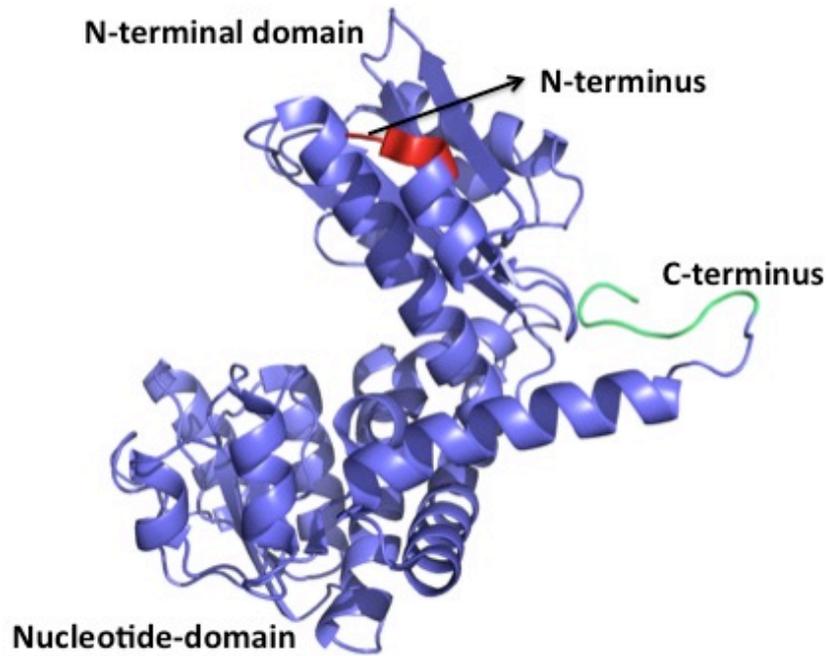


Figure 1.6. Schematic diagram representing one subunit for LeuDH from *B. sphaericus* (Accession number in PDB 1LEH).

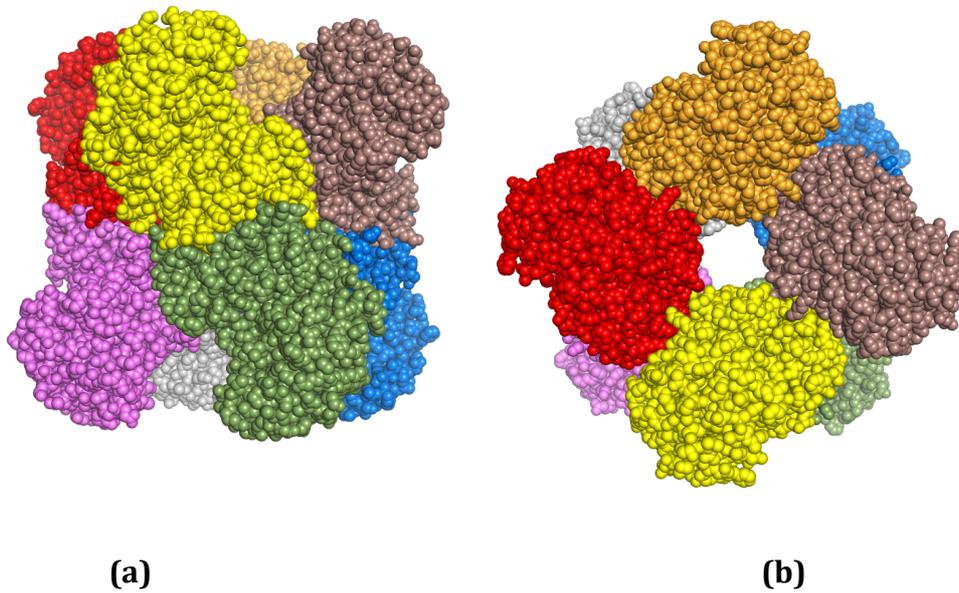


Figure 1.7. The overall organization of *B. sphaericus* LeuDH octamer with each subunit individually coloured. a) view down a molecule two-fold axis used to form the dimer, with the four-fold axis vertical. b) view down the four-fold axis (Accession number in PDB 1LEH).

case of LeuDH, the reactivity of a pyridinodeamino-NAD⁺ (100%) is similar to NAD⁺, but the relative rate of acetylpyridine-NAD⁺ (166%) is higher than NAD⁺, conversely to the reactivity of pyridinaldehyde-NAD⁺ (19%) and thio-NAD⁺ (21%) which are much lower than NAD⁺ (Ohshima et al, 1978).

1.5.1 Coenzyme stereospecificity

NAD(H) and NADP(H) act as electron carriers in all organisms by reversibly receiving or releasing a hydride ion at the C-4 position of the nicotinamide/dihyronicotinamide ring. When the coenzyme interacts with the protein, just one face of the nicotinamide ring is exposed to the substrate and the second face is packed against the protein surface. Fisher and coworkers showed that at the C-4 position, the two hydrogen atoms are enzymatically non-equivalent (prochiral) and thus the coenzyme transfers the hydride ion stereospecifically to the substrates (Fisher et al, 1953). According to the Fisher findings the NAD(P⁺)-dependent enzymes can be divided into two classes; class A enzymes eliminate the hydrogen from the 4-pro-R (A side/*re*-face) with those that transfer the hydrogen from the 4-pro-S (B side/*si*-face) are called class B enzymes. The GluDHs that have been characterized so far are all B-stereospecific enzymes (You et al, 1978). The LeuDHs and PheDH are also B-stereospecific proteins (Ohshima et al, 1978; Ohshima et al, 1991).

1.6 Substrate specificity of amino acid dehydrogenases

1.6.1 Glutamate dehydrogenase

Although having a marked preference for glutamate as substrate, GluDHs can catalyse oxidation of a number of carboxylic compounds that are analogues of glutamate. Struck and Sizer in 1960 and Hornby *et al* in 1983 observed that GluDHs from bovine liver can use monocarboxylic amino acids such as leucine and alanine in addition to glutamate as a substrate. However, the catalytic activity of these analogues is very low compared to L-glutamate, and it is not more than 1% of that of glutamate (Hornby et al, 1983; Struck & Sizer, 1960). In addition, *Mycobacterium smegmatis* GluDH was isolated and characterized by Sarada *et al*, 1980. The findings from this study showed that this enzymecatalyzes the reductive amination of 2-oxoglutarate at pH=8.0 and the

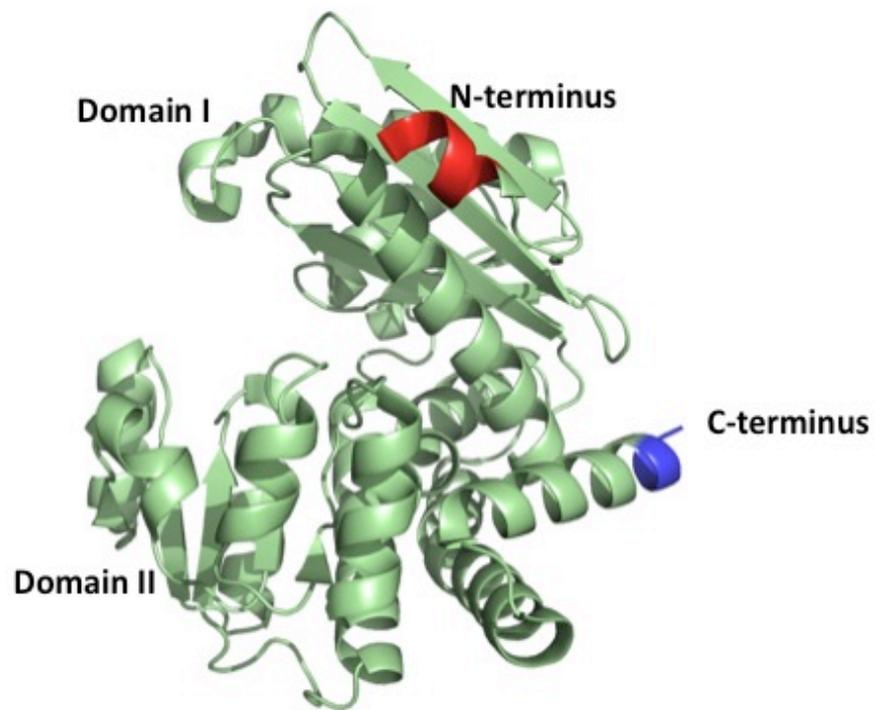


Figure 1.8. Schematic diagram representing a single subunit for PheDH from *Rhodococcus sp. M* (Accession number in PDB 1C1D).

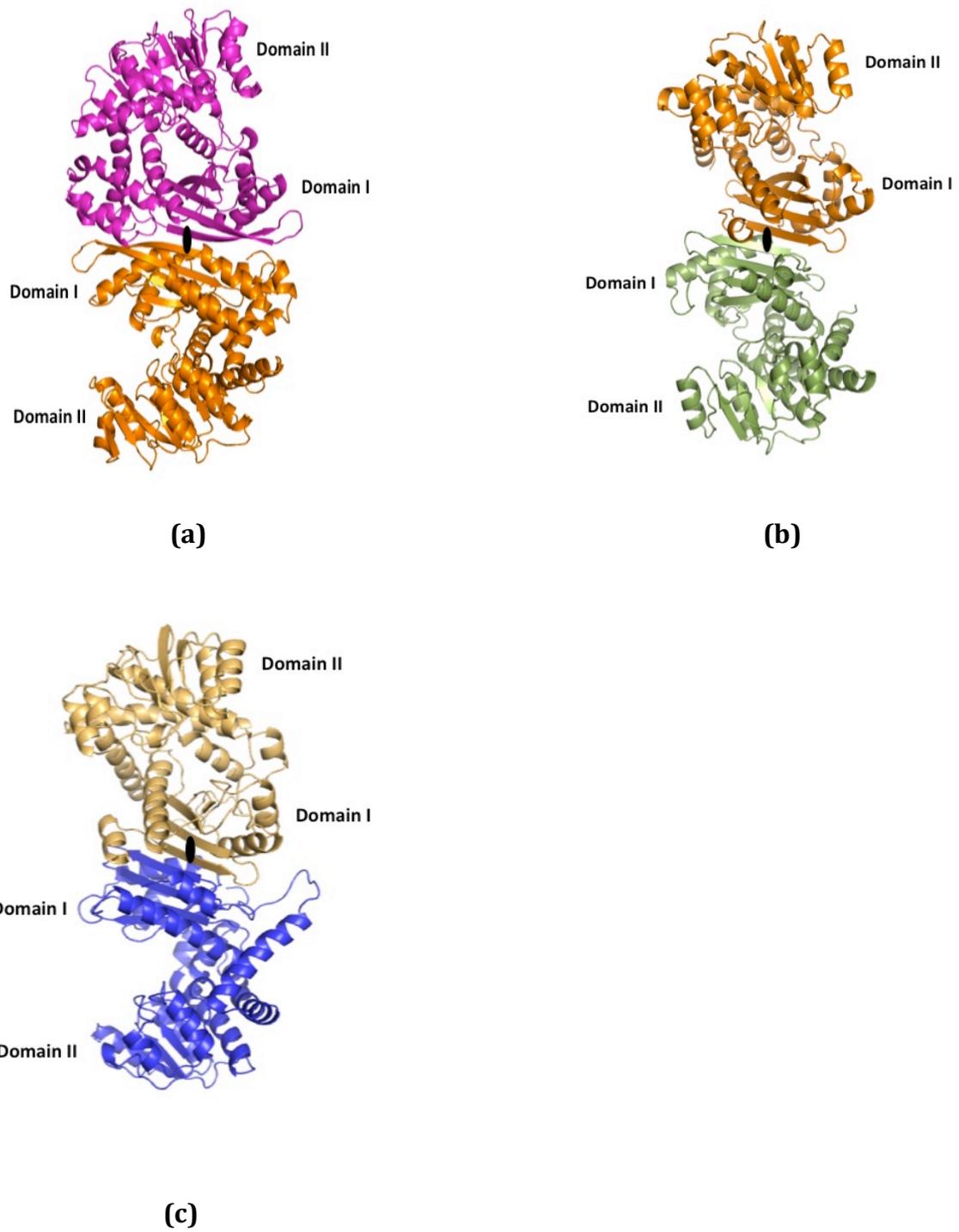


Figure1.9. Two subunits come together to form the dimer around the two-fold axis, mediated via β strands of N-terminal domain (here the two-fold axis comes out of the paper toward you). a) *C. symbiosum* GluDH, b) *Rhodococcus sp. M.* PheDH and c) *B. sphaericus* LeuDH.

optimal pH for oxidative deamination of glutamate was between 8.8-9.8. Moreover, substrate specificity of this enzyme was investigated for several amino and keto acids and the results illustrated that aspartate and leucine showed a minor activity of 10% of the activity with glutamate, while the activity of isoleucine, valine, and lysine was 5%. Insignificant activity was seen with phenylalanine, serine, glycine and norvaline. In this study 2-oxoglutarate, 2-oxobutyrate, 2-oxoisovalerate, oxovalerate were classified as poor substrates. (Sarada et al, 1980).

1.6.2 Phenylalanine dehydrogenase

The specificity of phenylalanine dehydrogenases for amino acids and keto acids varies between the enzyme sources. For example; the PheDH from *T. intermedius* is very selective for L-phenylalanine (Ohshima et al, 1991), whereas *B. sphaericus* PheDH catalyzes the oxidation of L-phenylalanine and L-tyrosine at nearly a similar rate (Asano et al, 1987a), since the relative rates of oxidative deamination reaction of L-phenylalanine and L-tyrosine were 100% and 72% respectively. In the same reaction direction, PheDH from *B. sphaericus* can utilize other aromatic amino acids, such as L-phenylalanine methyl ester, (reactivity 10%) and L-tyrosine methyl ester, (reactivity 7.3%). In addition, some aliphatic L-amino acids also showed some slight activity as a substrate, for instance the activity with L-norleucine was 3.9%. Substrate specificities for *B. sphaericus* PheDH have been tested in the reductive amination reaction. A number of α -keto acids have been shown to be active as substrates for this enzyme, with different relative activities. For example, if the reactivity with phenylpyruvate is 100%, *p*-hydroxyphenylpyruvate was 136%, similarly activities were seen with 2-oxovalerate (6.2%), 2-oxoisovalerate (5.5%) and 2-oxoisocaproate (7.8%) (Asano et al, 1987a).

1.6.3 Leucine dehydrogenase

Leucine dehydrogenases catalyze the reversible oxidative deamination of branched chain aliphatic L-amino acids. For instance; the relative activity of LeuDH from *B. sphaericus* in the direction of oxidative deamination was 100% with L-leucine; this activity was decreased with valine (74%) and isoleucine (58%). Besides these L-amino acids, which are the favourite substrates for

LeuDH, the unbranched chain aliphatic L-amino acids such as norvaline and L- α -aminobutyrate also act as substrate with LeuDH, where their relative activities were 41% and 14% respectively. D-amino acids, L-alanine, L-glutamate, L-threonine, L-phenylalanine and ω -amino acids are not substrates (Ohshima et al, 1978).

1.7 Homology and sequence alignment of amino acid dehydrogenases

A multiple alignment of the amino acid sequences of PheDH, LeuDH and GDH, shows a significant homology between these enzymes, but with some differences (Britton et al, 1993b). There are higher levels of identities between PheDH from *Bacillus sphaericus* and LeuDH from *B. stearothermophilus* (around 50%) compared to those seen between the GluDH from *C. symbiosum* to PheDH or LeuDH, which are about 20%. Although the general level of sequence identity between all these enzymes is low, the main features of secondary and tertiary structure can be recognized between these members of this homologous superfamily. Furthermore, a comparison of the structures of these enzymes has revealed that key residues responsible for catalysis, and for maintenance of the three-dimensional structure are conserved across these enzymes (Britton et al, 1993b). For example, the residues involved in the overall design of the catalytic pocket are conserved, including five glycine residues that are responsible for the shape of the active site (G122 and G123), orientation of functional groups at the catalytic centre (G90 and G91) and lie close to the binding site for the nicotinamide ring (G376) (figure 1.10). Additionally, lysine residue 113 (K113) that interacts with the α -carboxyl group of L-glutamate in GluDH, lysine 125 (K125) that enhances the nucleophilicity of the water molecule and aspartate 165 (D165), which is involved in proton transfer in the catalytic mechanism (Stillman et al, 1993), are also conserved (figure 1.10) (Britton et al, 1993b). As might be expected, the residues involved in stabilizing the common elements of the amino acid substrate are conserved, whereas those residues at the base of the binding pocket which are responsible for the main interactions that determine substrate specificity vary across the superfamily (Britton et al, 1993b). For example, in the *C. symbiosum* GluDH (as mentioned above) the essential

interactions that are implicated in substrate specificity are formed between the γ -carboxyl group of L-glutamate and the amino group of K89 and the side-chain hydroxyl group of S380 residues conserved across all GluDH (Stillman et al, 1993). With regard to LeuDH and PheDH, these two residues (K89 and S380) are different, being a leucine and valine, respectively, making the active site more hydrophobic.

In addition to the differences in the active site, there are a number of structural differences located at the N and C termini, and also a number of insertions and deletions between secondary structure elements. These variances play a principal role for the determination of the quaternary structure of these enzymes. In GluDHs, amino acid residues that are involved in the trimer interface, especially residues responsible for formation the loops which lie nearby to the three-fold axis (residues 388 to 390, 185 to 188 and 157 to 159) are deleted in PheDHs and LeuDHs. Residues that are implicated in the interactions of the two trimers are also deleted in PheDHs and LeuDHs (figure 1.10)(Britton et al, 1993b). However, there are some resemblances at the dimer interface between these enzymes. In GluDHs the two-fold interaction is centred on the binding between the extended antiparallel β a in domain I and its 2-fold symmetry related strand β a, with hydrophobic interaction formed between the side chains of Val57 and Ile58 (figure 1.9a), which is similar to that observed in LeuDHs and PheDHs (figure 1.9b,c). Figure 1.10 shows the sequence alignment between the GluDH, LeuDH and PheDH.

1.8 Chemical mechanism of GluDHs

Several mechanisms have been proposed to explain the catalytic activity of GluDHs. Smith et al (1975), have suggested that a lysine residue (126 in bovine liver and 125 in *Clostridium*) plays an important role in GluDHs catalysis, with this residue implicated in the reductive amination step, where it is thought to form a Schiff base intermediate by an interaction between the amino group of the Lys side-chain and the carbonyl group of the 2-oxoglutarate. The resultant ϵ -imino intermediate liberates a water molecule, before the ammonium ion attacks the Schiff base to generate 2-iminoglutarate moiety. The last step in this

supposed mechanism is the reduction of the 2-iminoglutarate by the transfer of the hydride ion from the cofactor, to yield L-glutamate. However, another mechanism for GluDH catalysis has been proposed, based on studies of the pH profiles for binding oxo-acids and analogues to bovine GluDH (Rife & Cleland, 1980a). This mechanism (shown in figure 1.7) suggests that, in the direction of reductive amination of 2-ketoglutarate to yield L-glutamate, the first step is the binding between the substrate and the enzyme-NADPH complex, where 2-oxoglutarate uses its γ -carbonyl group to bind to a protonated amino group of the enzymatic lysine residue. Ammonium that is present in the active site then reacts with the α -keto group, which, in turn, receives a proton from the active site Lys residue, producing a carbinolamine intermediate. The hydroxyl oxygen of the resultant carbinolamine, accepts a proton from an enzymatic carboxyl group, which concurrently deprotonates the amino group of this intermediate. The carbinolamine is thus converted to the imine with the liberation of a water molecule. The imine intermediate then accepts a hydride ion from NADPH and is converted to unprotonated L-glutamate, that takes a proton from the enzymatic carboxyl group, to produce protonated L-glutamate (Rife & Cleland, 1980a).

The suggested chemical mechanism of glutamate dehydrogenase by Rife and Cleland was the first mechanism to propose the formation of a carbinolamine, instead of a Schiff base intermediate, however, later studies have shown a number of inconsistencies in some of the details of this mechanism. Srinivasan and Fisher checked the redox portion of Rife and Cleland's suggested model, using proline and its keto acids as substrates (Srinivasan & Fisher, 1985). In this study the authors concluded that protonation of an α -iminoglutarate intermediate, after it is formed in the active site, is necessary to undergo oxidation by the cofactor. This result corresponds with the proposed mechanism by Rife and Cleland (Rife & Cleland, 1980a). Another conclusion in this study of Srinivasan and Fisher is that the amino acid substrate (L-glutamate) binds and reacts in the deprotonated state, which is opposite to Rife and Cleland model. Srinivasan and Fisher proposed a modified mechanism, which focuses on the conversion of the carbinolamine to an imine. In this revised mechanism, the authors suggested that once the carbinolamine is

formed, a proton from its amino group is removed by a lysine residue in the active site leading to breakdown of the carbinolamine into imine with the cleavage of the hydroxyl α -carbon bond with the removal of a proton from the active site carboxyl group. At this step the NADPH reduces the protonated imine, and the neutral form of the amino acid is released as a product (Srinivasan & Fisher, 1985).

In 1993 Stillman *et al* determined the crystal structure of the binary complex of glutamate dehydrogenase with L-glutamate, at 1.96Å resolution. This structure provided the crucial breakthrough in the catalytic mechanism of glutamate dehydrogenases. In general, the suggested mechanism by Stillman *et al* is similar to the Rife and Cleland model (see above). In this binary complex structure, electron density showed that a bound water molecule was present behind the 2-carbon of the glutamate substrate and the amino group of the side chain of the enzymic Lys125. In addition, the carboxyl side chain of Asp165 was in an appropriate position, due to a conformational rearrangement of the two domains, to act as a general base, to deprotonate the amino group of the substrate. This agreed with the Rife and Cleland model of the function of a catalytic carboxyl as a general base. The position of the NAD⁺ cofactor was modeled into the GluDH/L-glutamate binary complex, using the GluDH/NAD⁺ complex as a guide. From these models of the open and closed conformations of *C. symbiosum* GluDH, together with biochemical findings on the enzyme mechanism (Rife & Cleland, 1980b), Stillman *et al* proposed a chemical mechanism for GluDH catalysis.

Firstly, the α -amino group of the L-glutamate substrate is deprotonated by the side chain carboxyl group of Asp165 working as a general base, which causes the transfer of a hydride ion to the *Si* face of the NAD⁺, and produces aminoglutarate intermediate. In the second step of this mechanism, an enzyme bound water molecule whose nucleophilic character is activated by Lys 125, is proposed to attack the iminoglutarate intermediate to form a carbinolamine. In the production of carbinolamine intermediate and its subsequent breakdown to the 2-oxoacid, Asp165 is essential for proton transfer to and from the substrate. In the final step of this proposed reaction scheme, losing a proton from Lys125 and Asp165 to bulk solution is required

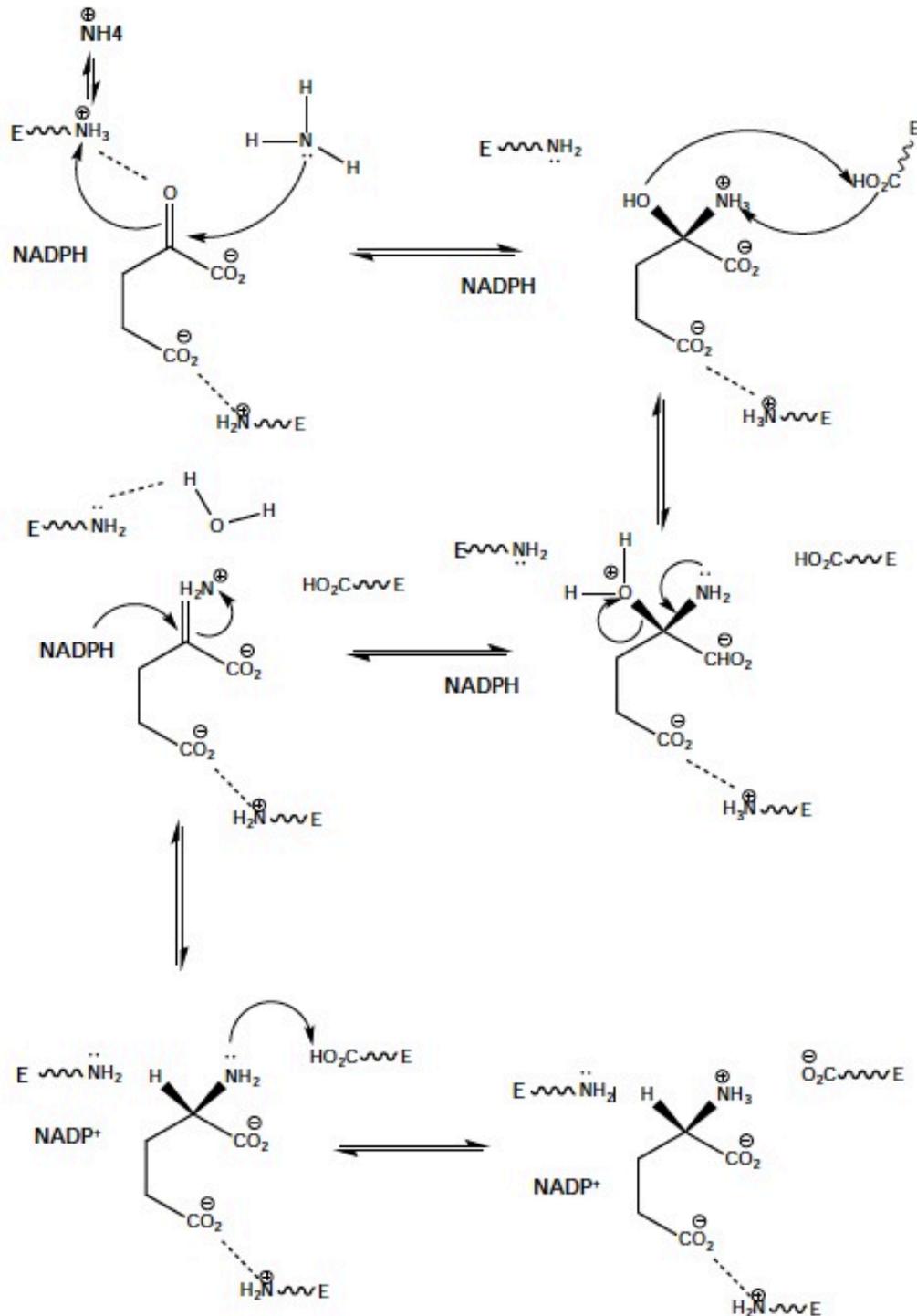


Figure 1.11 Glutamate dehydrogenase chemical mechanism proposed by Rife and Cleland (adapted from Brunhuber and Blanchard, 1994).

for the completion of the catalytic cycle (Stillman et al, 1993).

1.9 Chemical mechanism for phenylalanine dehydrogenase

In 1999 Vanhooke. J. L. and co-workers have proposed a chemical mechanism for phenylalanine dehydrogenase based on X-ray crystal structures of ternary complexes of *Rhodococcus sp. M4* PheDH with NAD⁺ and phenylpyruvate and NAD⁺ plus β -phenylpropionate (Vanhooke et al, 1999). In this study, the authors tried to answer questions that had not been explained in previous mechanisms (Rife & Cleland, 1980b) of amino acid dehydrogenases such as why these enzymes do not catalyze the reduction of bound 2-oxo acids to 2-hydroxy acids. Analysis of the structures of the two PheDH ternary complexes has provided a logical basis for the differentiation between the keto acid and the imino acid as substrate for reduction by these enzymes. Findings from this study were used to suggest modified chemical mechanism of amino acid dehydrogenases. As shown in figure 1.10, the first three steps of this chemical mechanism are similar to that of proposed by Rife and Cleland for glutamate dehydrogenase (Rife & Cleland, 1980b). At the start of the reaction the phenylpyruvate is interacting by hydrogen bonding through its keto oxygen, to the side chain amino group of the lysine residue 78 in the active site. This bonding pattern is proposed to lead to the prohibition of phenylpyruvate reduction via NADH. In the first step, ammonia that exists in the active site, attacks the carbonyl group of phenylpyruvate and lys78 protonates the keto oxygen, producing a carbinolamine intermediate with its amino group in the protonated form. In step 2, transfer a proton is facilitated by the side chain carboxyl group of Asp118 in the active site, which makes the charge of the carbinolamine hydroxyl group positive. Consequently, a water molecule is removed from the carbinolamine in the third step, producing another intermediate and the iminophenylpyruvate. Following this step (step4) in the Rife and Cleland mechanism, NAD(P)H reduces the imino group to an amino group, which then undergoes protonation by a carboxyl group in the active site. In step 4 of the PheDH mechanism, the side chain carboxyl group of the enzymic Asp118 donates a proton to Lys78 after the removal of water. This transfer of the proton generates a negative charge on the carboxyl group of

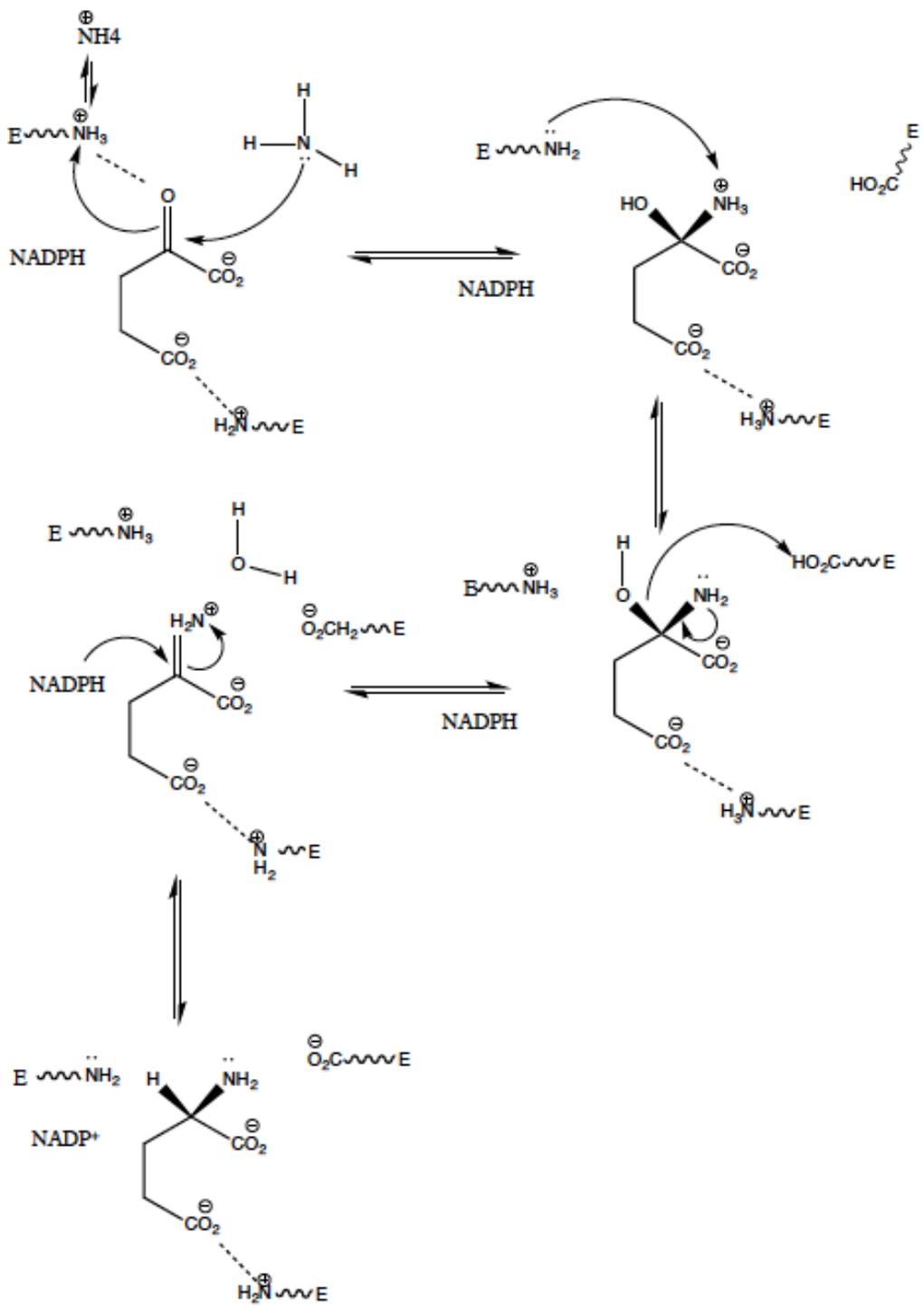


Figure 1.12. Modified chemical mechanism proposed by Srinivasan and Fisher, 1985 (adapted from Brunhuber and Blanchard, 1994).

Asp118, which stabilizes the positive charge on the imine, and positions this imine for hydride transfer. NADH then reduces the iminophenylpyruvate in step 5, producing phenylalanine that is subject to protonation by Lys78, prior to liberation from the active site (Vanhooke et al, 1999).

1.10 Previous structures of amino acid dehydrogenases

A number of three-dimensional structures have been determined of members of the amino acid dehydrogenase superfamily from both prokaryotic and eukaryotic sources (table 1.1). Although they have different quaternary structure, the basic structure of the subunit is the same, the two domains separated by a deep cleft. (For details see sections; 1.4.1, 1.4.2 and 1.4.3).

1.11 Project Aim

Despite the wealth of studies on the amino acid dehydrogenase superfamily over the last 50 years, there are a number of questions that remain unanswered, in coenzyme specificity, substrate specificity, and catalytic mechanism. The promising use of bacterial amino acid dehydrogenases in many biotechnological applications, requires a full understanding of the structure/function relationships in this enzyme superfamily, to aid in the design of mutant enzymes with added commercial potential. Although many structures of GluDH from bacteria, archaea and mammals have been determined, a GluDH/cofactor/substrate structure of a bacterial enzyme has not been deposited in the Protein Data Bank. In addition, the ternary complexes of the eukaryotic GluDHs that have been determined are only at medium resolution, precluding precise identification of the orientations of the various groups in the active site. Furthermore, only two leucine dehydrogenase structures are known and one phenylalanine dehydrogenase (from *Rhodococcus. sp*) which is a dimer. Structures that are not present in the database, and preclude a full understanding of the enzyme superfamily are a ternary complex of a bacterial GluDH, a binary complex of a bacterial GluDH with cofactor, a PheDH structure of the octameric form of the enzyme, and any truly high-resolution structure of binary or ternary complexes of GluDH. The aims of this thesis were therefore to:

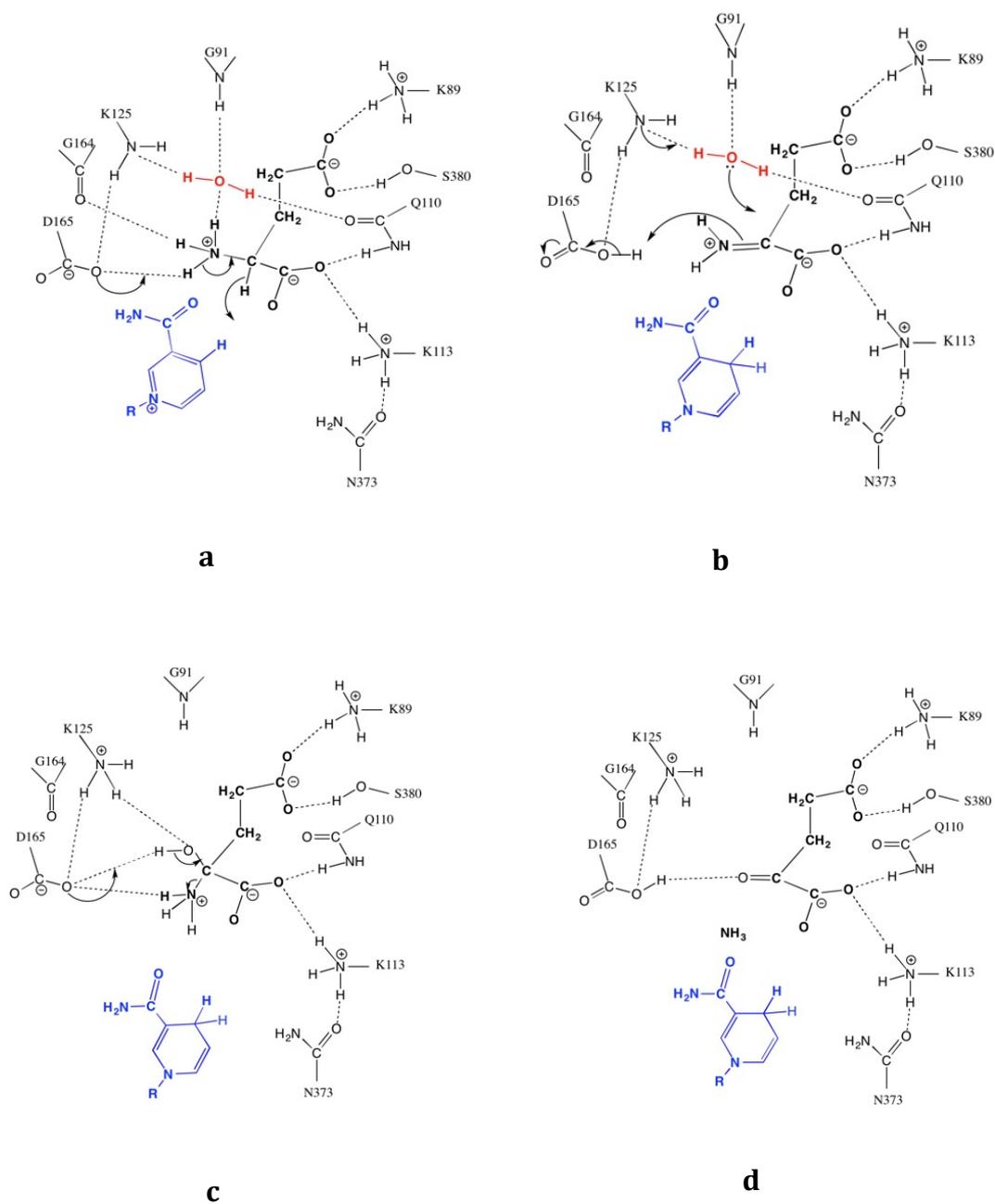


Figure 1.13. The main steps (a) to (d) of the proposed mechanism of GluDHs. This scheme illustrates the reaction pathway in the direction of the oxidative deamination of glutamate to 2-oxoglutarate. This diagram is based on that from Stillman and coworkers (Stillman et al, 1993).

- i) Determine the structure of an octameric bacterial phenylalanine dehydrogenase from *B. sphaericus* in complex with substrate and cofactor.
- ii) Determine the structure of a ternary complex of GluDH from a bacterial source (*M. smegmatis*)
- iii) Determine a high-resolution structure of the *C. symbiosum*/ NAD⁺ binary complex.

Determination of the three-dimensional structure of these amino acid dehydrogenases will offer insights into the catalytic mechanism and aid in increased understanding of the molecular basis of how these enzymes differentiate between amino acid substrates, which will help in the design of new mutant enzymes active with non-natural amino acids, which can be used in the pharmaceutical industry.

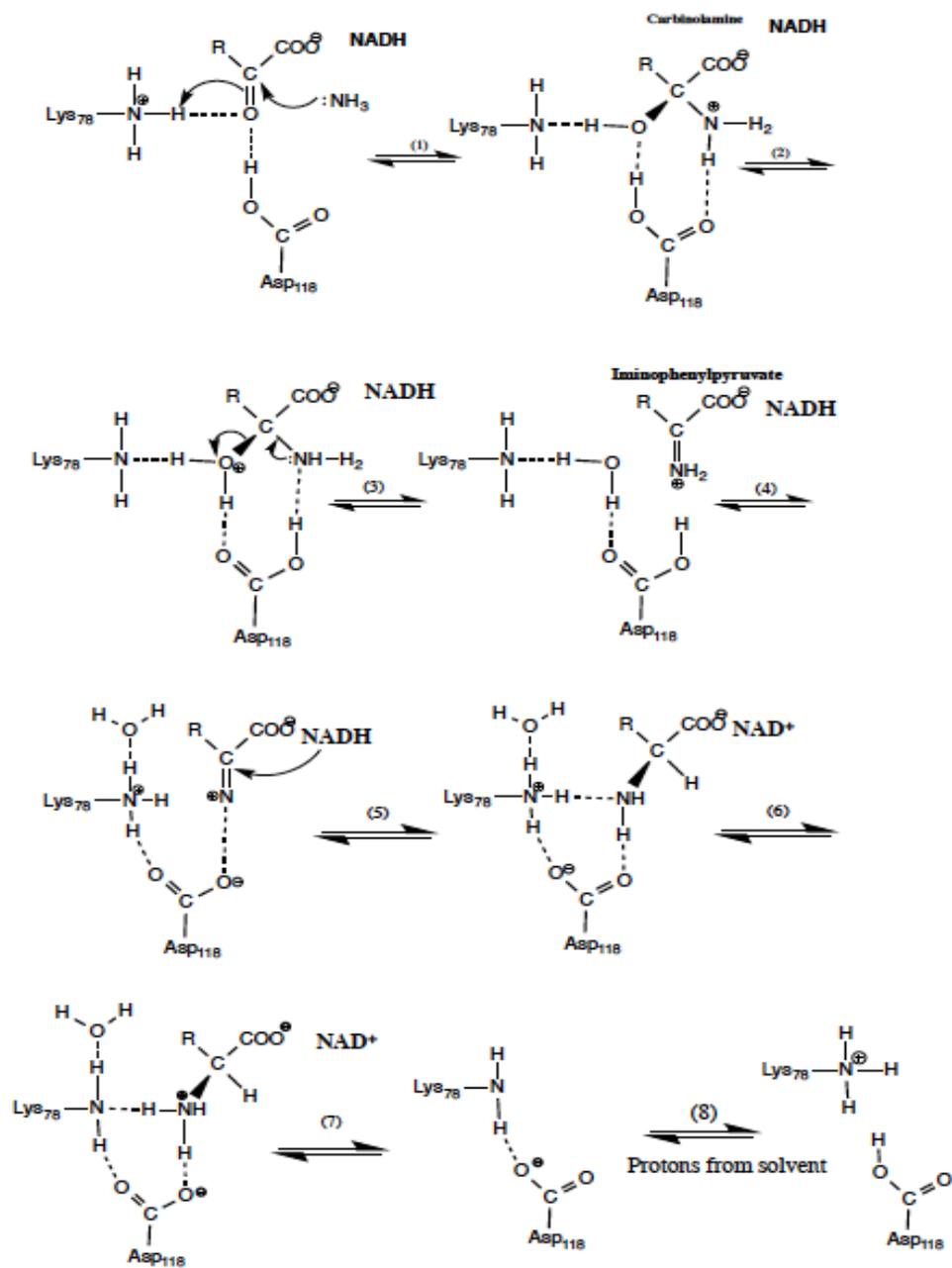


Figure 1.14. The modified chemical mechanism for phenylalanine dehydrogenase in the direction of reduction of phenylpyruvate to produce phenylalanine. Taken from (Vanhook et al, 1999).

a) Glutamate dehydrogenases

Accession number PDB	Species	Ligands	Resolution Å	Subunit	Reference
1HRD	<i>Clostridium symbiosum</i>		1.96	49,165	(Baker et al, 1992b)
1BGV	<i>Clostridium symbiosum</i>	Glutamate	1.90	49,165	(Stillman et al, 1993)
1B26	<i>Thermotoga maritima</i>		3.0	45,689	(Knapp et al, 1997)
1NR7	<i>Bos. taurus</i>		3.30	55,561	(Banerjee et al, 2003)
1NQT	<i>Bos. taurus</i>	ADP	3.50	55,561	(Banerjee et al, 2003)
1HWZ	<i>Bos. taurus</i>	Glu;GTP;NADPH	2.80	55,561	(Smith et al, 2001)
1HWY	<i>Bos. taurus</i>	2-Oxoglutaric acid, NAD	3.20	61,512	(Smith et al, 2001)
3MW9	<i>Bos. taurus</i>	Glu;GTP; NADH	2.50	61,512	(Peterson & Smith, 1999)
1L1F	Human		2.70	61,393	(Smith et al, 2002)
1BVU	<i>Thermococcus. litoralis</i>		2.50	46,608	(Britton et al, 1999)
1EUZ	<i>Thermococcus.p rofundus</i>		2.25	46,699	(Nakasako et al, 2001)
1GTM	<i>Pyrococcus. furiosus DSM 3638</i>		2.20	47,114	(Yip et al, 1995)
2BMA	<i>Plasmodium falciparum</i>		2.70	52,546	(Werner et al, 2005)
3R3J	<i>Plasmodium</i>		3.10	57,343	(Zocher et al,

	<i>falciparum 3D7</i>				2012)
1V9L	<i>Pyrobaculum islandicum</i>	NAD	2.80	46,906	(Bhuiya et al, 2005)
3MVO	<i>Bos. taurus</i>	Eu3+	3.23	61,512	(Bailey et al, 2011)
3MVQ	<i>Bos. taurus</i>	Zinc	2.94	61,512	(Bailey et al, 2011)
3QMU	<i>Bos. taurus</i>	Epicatechin-3-gallate	3.62	61,512	(Li et al, 2011)
3ETE	<i>Bos. taurus</i>	Hexachlorophene	3.0	61,512	(Li et al, 2009)
3ETD	<i>Bos. taurus</i>	Bithionol	2.50	61,512	(Li et al, 2009)
2YFQ	<i>Peptoniphilus. accharolyticus</i>		2.94	46,514	(Oliveira et al, 2012)
3AOG	<i>Thermus. Thermophiles</i>		2.10	46,113	
4BHT	<i>Escherichia. coli</i>		2.50	48,581	(Sharkey et al, 2013)

b) Phenylalanine dehydrogenase

Accession number PDB	Species	Ligands	Resolution Å	Subunit	Reference
1BGX	<i>Rhodococcus. sp</i>	NAD ⁺ + beta-phenylpropionate	2.30	36,609	(Vanhooke et al, 1999)
1BW9	<i>Rhodococcus. sp</i>	NAD ⁺ + Phenylpyruvate	1.50	36,609	(Vanhooke et al, 1999)
1C1X	<i>Rhodococcus. sp</i>	NAD ⁺ + L-3-	1.40	36,609	(Brunhuber et

		phenyllactate			al, 2000)
1C1D	<i>Rhodococcus. sp</i>	NADH + Phenylalanine	1.25	36,609	(Brunhuber et al, 2000)

c) Leucine dehydrogenase

Accession number PDB	Species	Ligands	Resolution Å	Subunit	Reference
1LEH	<i>Lysinibacillus. sphaericus</i>		2.20	39,043	(Baker et al, 1995)
3VPX	<i>Sporosarcina psychrophila</i>		2.55	39,345	

Table 1.1 shows structures of some amino acid dehydrogenases in the PDB

a) Glutamate dehydrogenases, **b)** Phenylalanine dehydrogenases and **c)** Leucine dehydrogenases.

Chapter Two: Materials and Methods

2.0 Overview

This chapter describes the molecular biology and X-ray crystallographic techniques used for the structure determination of target proteins in this thesis.

2.1 Gene amplification, cloning and overexpression of glutamate dehydrogenase from *M. smegmatis*

2.1.1 The design of primers

Gene amplification using the polymerase chain reaction (PCR) needs two oligonucleotide primers; a forward and a reverse primer. These primers were designed for the target gene (*msmeg5442*) using the genome sequence of *Mycobacterium Smegmatis* strain MC² 155. Pairs of primers should be exactly matching to the sequence at the 3' end of every strand of the DNA to be amplified. There are some conditions that must be available in the selected primers to amplify the target gene: the length of primers is usually between 18 to 30 bases, they should have %GC of 40 to 60, the melting point should be in the range of 52°C to 60°C, and the primers must not be self-complementary. The primers used for the amplification of *msmeg5442* were 20 and 21 bases in length and included 6 nucleotides sequence upstream of the restriction site for cutting the PCR products by restriction enzymes. The forward primer was designed to have an NdeI (CATATG) restriction site followed by 16 nucleotides of the *msmeg5442* gene. The reverse primer contained a HindIII (AAGCTT) restriction site followed by 17 nucleotides of genomic DNA from downstream of the gene.

The sequence of primers that used for amplification of target gene is described below:

Forward primer 5'-CATATGAGCGAACTACACCC-3'

Reverse primer 5'-AAGCTTTCAGATGAGCCCGAG-3'

2.1.2 Amplification of target gene using polymerase chain reaction (PCR)

Firstly genomic DNA of *Mycobacterium smegmatis* (*M. smegmatis*) strain MC²-

155 (a gift from Prof J. Green, Sheffield) was prepared. Bacterial genomic DNA was prepared by growing the bacteria in LB medium (1% tryptone, 0.5% yeast extract and 1% NaCl) at 37°C on a shaker at 250rpm for 72 hours. Extraction of the genomic DNA from the culture was carried out using Keyprep Bacterial DNA extraction kit. Successful extracting was confirmed by running a sample of the DNA on 1% gel agarose electrophoresis (AGE).

The gene of interest was amplified using the Polymerase Chain Reaction (PCR) with the genomic DNA as a template.

For the PCR reaction the following materials were added to the PCR tube:

Sterile H ₂ O	27.5µl
DNA Polymerase Buffer 5x	10µl
dNTPs (5mM)	2.5µl
Forward primer (5µM)	4.0µl
Reverse primer (5µM)	4.0 µl
Genomic DNA (<i>M. smegmatis</i>)	2.0µl
DNA polymerase	1.0 µl
Total volume	50 µl

The tube was the placed in a PCR machine and the reaction was set up under the following conditions:

Initial denaturation	98°C	2:0 min	
Denaturation	98°C	30 sec	X 25 Cycles
Annealing	60°C	30 sec	
Elongation	72°C	1:30 min	
Final Elongation	72°C	7:0 min	

2.1.2.1 Purification of PCR products from agarose gel

The resulting DNA fragments were run on a 1% TAE agarose gel at 100 V for one hour to separate the DNA. The gel was visualized over a UV trans illuminator and the bands that had the expected size were removed from the gel using a clean scalpel blade and transferred to an Eppendorf tube. The DNA fragments were purified from the agarose gel slices using a QiaQuick®Gel Extraction Kit (Qiagen) and the standard protocol. The tube that contained the agarose gel slices was weighed and 3 volumes of buffer QG (a proprietary buffer containing guanidine thiocyanate) was added to each volume of agarose (e.g. 300µl to 100mg agarose). The tube was then transferred to a heat block at 50°C for 10 min, with gently vortexing every two minutes. Once the agarose had dissolved, one gel volume of isopropanol was added to the microfuge tube and mixed before the contents were poured into a MinElute spin column and centrifuged at 12470 g for 1 minute. After centrifugation, the flowthrough was discarded and 500µl of buffer QG added before centrifuging at 112470 g for 1 minute. Again the flowthrough was discarded, before the addition of 750µl of buffer PE (a proprietary buffer containing ethanol) and centrifugation to remove salt and carbohydrates, the flowthrough was again disposed of, and the spin column was centrifuged at 12470 g for 1 minute to remove residual ethanol. The spin column was then transferred to a fresh microfuge tube and 10-50µl of EB buffer (10mM Tris-HCl, pH 8.0) was added to the centre of the silica matrix and incubated at room temperature for 1 minute before elution of the purified DNA by centrifugation at 12470 g for 1 minute. The purified DNA was then used in the ligation step.

2.1.3 Preparation of pET21a plasmid

A small-scale plasmid preparation was completed using the QIAprep Spin Plasmid Miniprep system. This system applies the principles of the alkaline lysis method of Birnboim and Doly (Birnboim & Doly, 1979) followed by plasmid purification from a silica resin. The kit was purchased from Qiagen and used as follows: A single white colony of DH5α *E. coli* containing the plasmid pET21a (figure 2.1) was picked from an LB agar plate and used to inoculate 3 ml of LB medium supplemented with 50µg/ml ampicillin. The culture was left

to grow overnight at 37°C on rotary shaker at 200 rpm. The overnight culture containing the required plasmid was pelleted by centrifugation at 4500 rpm for 10 minutes. The supernatant was discarded and the cells resuspended in 250µl of buffer P1 (50mM Tris-HCl, pH=8.0, 10mM EDTA and 100 µg/ml of RNAase A) and transferred to a 1.5 ml tube. 250µl of buffer P2 (200mM NaOH and 1% SDS) was then added to lyse the cells, and the tube contents were mixed via inversion. 350µl of buffer N3 (3M Potassium acetate pH=5.5) was added to neutralize the lysis reaction and the tube was gently inverted several times to mix the contents. The bacterial lysate was centrifuged at 12470 g for 10 minutes. The supernatant was recovered and applied to a QIAprep spin column, seated in a 2 ml microcentrifuge tube. The spin column was then centrifuged at 1300 rpm for 1 minute and the flowthrough discarded. The column was washed by adding 0.75 ml of buffer PE (60% Ethanol, 60mM Potassium acetate, 8.3mM Tris-HCl, 0.04mM EDTA) and centrifuged at 1247 g rpm for 1.0 minute. The flowthrough was removed and a second centrifugation at 12470 g was used to remove any residual buffer PE. The QIAprep column was then transferred to a fresh 1.5ml tube and the DNA was eluted by adding 50µl of EB buffer (10 mM Tris-HCl, pH 8.5) and incubated at room temperature for 2 minutes before centrifugation at 13000rpm for one minute to collect the pure plasmid.

2.1.4 Restriction digestion

DNA fragments and pET21a plasmid were cleaved with restriction endonucleases, NdeI (NEB) and HindIII (NEB) in order to provide appropriate ends for cloning experiments or to confirm the presence of particular gene fragments within a vector. Each was incubated with DNA (PCR products and pET21a plasmid) in NEBuffer (Buffer 4) to achieve optimum activity in the restriction digest. The reaction was carried out as following:

	pET21a vector	msmeg5442 gene
DNA	10µl	10µl
HindIII	1.0µl	1.0µl
NdeI	1.0µl	1.0µl
Buffer 4(NEB)	2.0µl	2.0µl

Sterile H ₂ O	6.0μl	6.0μl
Total volume	20μl	20μl

Reactions were incubated at 37°C for 3 hours. The resulting digested DNA (PCR products and pET21a plasmid) were visualized by agarose gel electrophoresis and the PCR products (msmeg5442 gene) and pET21a vector were purified using a QiaQuick®Gel Extraction Kit (Qiagen) and the protocol as described above (2.1.2.1).

2.1.5 Ligation of PCR product into plasmid

DNA fragments were ligated into the linearized vectors (pET21a and pET14b) expression vector using the reaction catalyzed by bacteriophage T₄ ligase (Sambrook & Russell, 2001) This enzyme forms phosphodiester bonds between adjacent 3'-OH and 5'-phosphate ends of DNA. T₄ ligase is capable of joining both cohesive and blunt ended DNA fragments. Ligations were performed in 20μl volumes using an approximate molar ratio of 1:1, 1:3 and 3:1 (plasmid: insert). Ligation buffer was added to a final concentration of 1X. The ligation reaction was set up as follows:

	Tube 1	tube2	tube3
T4 ligation buffer 5X	4μl	4μl	4μl
Vector DNA	3μl	5μl	9μl
Insert DNA	9μl	5μl	3μl
T4 ligase	1μl	1μl	1μl
Total volume	20μl	20μl	20μl

All tubes were mixed well and incubated at room temperature for 1.0 hour.

2.1.5.1 Transformation of ligation mixes into DH5α *E. coli*

100μl aliquots of competent DH5α *E. coli* were removed from -80°C storage and defrosted on ice for 5 minutes. When the cells had thawed, 5μl from the various ligation mixes were added to competent cells before being left to incubate on ice for 30 minutes. Cells were then shocked at 42°C for 90 seconds and returned to ice for 2 minutes. 400μl of LB medium was added and the bacteria were incubated at 37°C at 200 rpm for 1.0 hour. After centrifugation

for 1.0 minute at 12470 g, the transformed cells were resuspended in 100µl of LB before plating on LB agar supplied with 50µg/ml ampicillin and the plates were incubated at 37 °C overnight.

2.1.5.2 Screening of possible clones

A general procedure of screening transformed plasmids in *E. coli* was performed by plating the transformed cells on media containing suitable antibiotics. The choice of antibiotics is dependent upon the existence of selectable antibiotic marker in the cloning vector. The selectable marker that is present in pET21a is ampicillin, and therefore, the transformed cells were screened on LB agar containing 50µg/ml ampicillin (section 2.1.5.1). Some colonies from the transformed agar plate were picked to inoculate 5ml LB medium, supplemented with 50µg/ml ampicillin and grown overnight at 37 °C at 200rpm. Plasmids were extracted from the cells as described in section 2.1.3. Transformed cells were initially screened for the insertion by restriction digestion as described in section 2.1.4. The extracted plasmids were also sequenced using T7 forward and reverse primers.

2.1.6 Protein Expression

2.1.6.1 BL21 (DE3) transformation protocol

Transformation of the recombinant plasmids into *E. coli* cells, *BL21*(DE3) was carried out for overexpression of the target protein. 100µl aliquots of competent cells were removed from -80°C storage and defrosted on ice for 5 minutes. When the cells had thawed 2µl of the plasmid preparation were added before being left to incubate on ice for 30 minutes. The cells were then shocked at 42°C for 90 seconds and returned to ice for 2 minutes. 400µl of LB medium was added and the bacteria were incubated at 37°C at 200 rpm for 1.0 hour. 100µl from the culture was plated on LB agar supplied with 50µg/ml ampicillin and the plate was incubated at 37 °C overnight.

2.1.6.2 Small-scale overexpression trials

The cells containing the recombinant vector were grown overnight in 5 ml LB media containing 50µg/ml ampicillin. 1ml from this primary culture was added

to 50ml LB medium supplemented with the same antibiotic and grown at different temperatures (37°C, 30°C and 25°C) on a rotary shaker at 200 rpm. The cells were grown until an OD_{600nm} of 0.6 was reached. A 1ml sample of the culture was removed and cells pelleted on a micro-centrifuge at 13000 rpm for 1 minute. The supernatant was then discarded and the cell pellet stored at -20°C. After that, 1mM IPTG was added to the culture in order to induce expression of the protein of interest. The cells were then grown for a further 4 hours under the same conditions before centrifuging at 4499 g for 20 minutes at 4°C. The supernatant was discarded and the cell pellets stored at -20°C. The level of over-expression, protein solubility and optimum temperature were determined by analyzing the samples using SDS-PAGE.

2.1.6.2.1 Expression Level and Solubility Test

Although it is sometimes possible to refold active protein from the insoluble fraction, crystallization usually works better if the protein is expressed in the soluble form. For testing the whole protein content and its solubility, cells were lysed using the BugBuster® reagent (Novagen), a commercial reagent for rapid bacterial cell lysis, using the manufacturer's protocol. 2 ml BugBuster reagent plus 5µl Benzonase were added to the Falcon tube containing the cell pellets, with repeated pipetting until the cells were completely suspended. Benzonase is an endonuclease enzyme and was used to degrade all forms of DNA and RNA present in the sample. After this, the Falcon tube contents were incubated on ice for 30 minutes before centrifugation at 4499 g, at 4°C for 20 minutes, to pellet all the insoluble fraction, and leave soluble proteins in the supernatant. For analysis of the insoluble fraction, the remaining pellet was resuspended in 2ml 50mM Tris-HCl pH=8.0 and disrupted by three sonication bursts of 20sec at 16micron amplitude.

2.1.6.3 Large-scale over-expression

When a suitable condition had been determined for protein expression, the growth was scaled up using one litre of medium in 500 ml aliquots using 2 litre Erlenmeyer culture flasks. 10 ml samples from overnight cultures were added to two culture flasks containing 500ml LB with 50µg/ml ampicillin and grown at

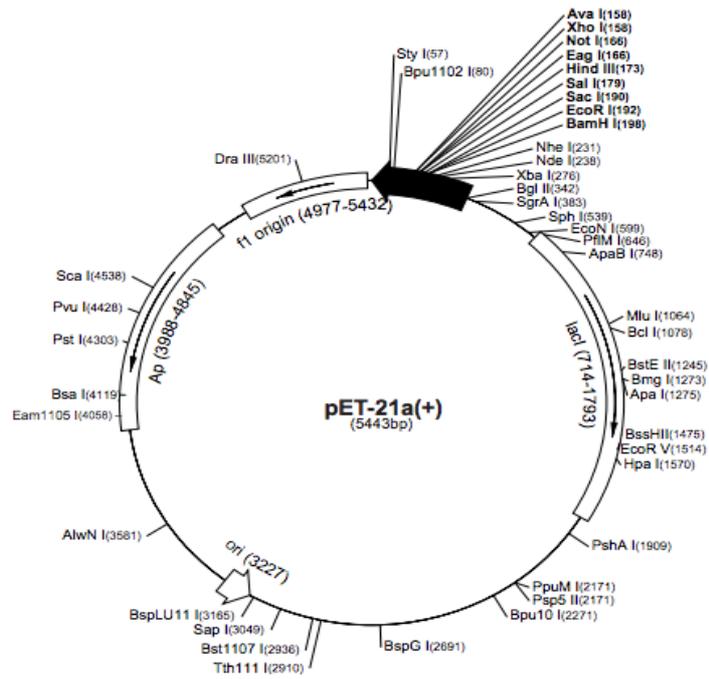


Figure 2.1. pET21a vector. The plasmid contains the multiple cloning sites (black arrow), T7 promoter (311-327), T7 terminator (26-72), lactose operator protein sequence (*lacI*) (714-1793) and ampicillin resistance (Ap) (3988-4845).

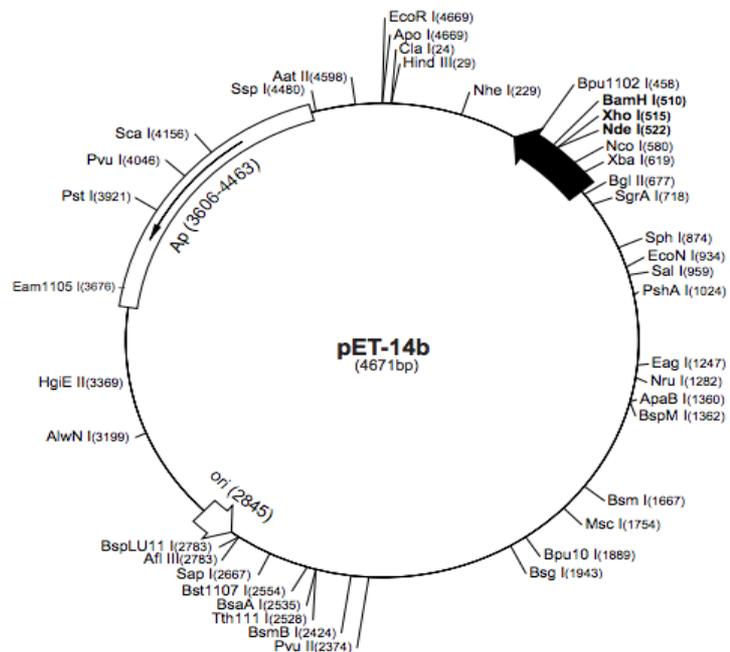


Figure 2.2. pET14b vector. The plasmid contains the multiple cloning sites (black arrow), T7 promoter (464-662), T7 terminator (404-450) and ampicillin resistance (Ap) (3606-4463).

37°C with shaking at 200 rpm until the optical density at 600nm reached 0.6. The cell cultures were then induced with 1 mM IPTG and were left to grow under the same conditions for a further 4 hours. Cells were then pelleted by centrifugation at 4499 g for 20 minutes at 4°C in a Beckman centrifuge before resuspension in about 100 ml of LB medium. The cell pastes were collected in a Falcon tube and harvested by centrifugation again under same conditions. The supernatant was discarded and the cell pellets stored at -80°C, until required.

2.1.6.3.1 Production of Seleno-L-Methionine incorporated protein

A starter culture growth in LB medium supplemented with 50µg/ml ampicillin was used to inoculate 1.0 liter of LB containing the same antibiotic split between two 2 litre Erlenmeyer culture flasks, yielding 2% inoculation. The cultures were left to grow at 37°C with shaking at 200 rpm until the A_{600} reached 0.6. At this point the cells were pelleted at 1844 g for 20 min prior to resuspension in about 100 ml minimal medium (table 2.1 describes the minimal media composition). The cells were centrifuged at 1844 g for 20 min before being resuspended in 50 ml seleno-L-methionine substituted minimal medium supplemented with 50µg/ml ampicillin. This culture was used to inoculate 2x2 litre conical flasks containing 500 ml substituted minimal medium with 40 mg/litre Seleno-L-methionine. The cells were left to grow at 37°C until with shaking at 200 rpm the A_{600} reached 0.6 before adding IPTG for induction. The cells were allowed to grow 8 hours, under the same condition. The cells were harvested by centrifugation and stored at -80°C, until required.

2.1.6.4 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Analysis of fractions obtained during preparation of cell free extracts or protein purification was carried out using SDS-PAGE gels.

i) Sample preparation

For analysis of the protein samples by SDS-PAGE gel, 15-20µl from these samples were mixed with 5µl loading buffer (250mM Tris-HCl, pH=8.0, 15%SDS, 50% Glycerol, 0.01% bromophenol blue and β-mercaptoethanol) in Eppendorf tubes, and boiled for 10 min prior to loading on the SDS gel.

ii) Gel preparation

Gels were constructed using a Biorad gel former. A 1 mm gel plate was used and

Selenomethionine-minimal medium		
To Autoclave	Chemicals	Concentration
	Potassium dihydrogen phosphate	4.5g/litre
	Dipotassium phosphate	10g/litre
	Tri-sodium citrate	0.5g/litre
	Ammonium sulfate	1.0g/litre
	Glycerol	5.0ml/litre
	Adenine	0.5g/litre
	Guanosine	0.5g/litre
	Thymine	0.5g/litre
	Uracil	0.5g/litre
	Deionized water	980 ml
After autoclaving	Chemicals	Concentration
	Lysine	100 mg/litre
	Phenylalanine	40 mg/litre
	Threonine	100 mg/litre
	Isoleucine	50 mg/litre
	Leucine	25 mg/litre
	Valine	50 mg/litre
	Magnesium sulfate	1.0 mg/litre
	Thiamine	4 mg/litre
	Selenomethionine	40 mg/litre

Table 2.1. Minimal Seleno-L-methionine (SeMet) medium composition.

gels were comprised from a 12% acrylamide separating gel with 6% stacking gel.

12% Resolving Gel

30% Acrylamide	2.5 ml
1M Tris-HCl pH8.8	2.35ml
Sterile water	0.95ml
10% SDS	62.5 μ l
10% APS	62.5 μ l
TEMED	6.25 μ l

6% Stacking Gel

30% Acrylamide	0.75ml
1M Tris-HCl pH6.8	0.47ml
Sterile water	2.46ml
10% SDS	37.5 μ l
10% APS	37.5 μ l
TEMED	3.75 μ l

Gels were run in a gel tank under SDS running buffer. The voltage was applied at 80V for 30 minutes and then was increased to 200V for 60 minutes.

10X SDS Running Buffer

Tris Buffer	30.3g
Glycine	144g
SDS	10g

Make 1 liter in distilled water to a final concentration 1X.

Staining and Destaining

Gels were removed from the gel plates and washed by distilled water and stained for 2 hr with Coomassie Blue Stain (0.1% Coomassie blue, 40% Methanol and 10% Acetic acid). Gels were removed from the Coomassie Blue Stain, washing by distilled water and destaining in a 20% Methanol and 10% Acetic acid solution until the bands were seen against the background.

2.2 Purification of all target proteins; *Mycobacterium smegmatis* GluDH, *Clostridium symbiosum* GluDH and phenylalanine dehydrogenase (PheDH) from *Bacillus sphaericus*.

In order to produce diffraction quality crystals the target protein must be purified to more than 95% purity. Protein purification is accomplished using several steps of liquid chromatography to isolate the desired protein from the complex mixture based on its unique physical, chemical and biological properties. Normally, particular purification protocols should be designed for every specific protein. Purification of proteins involve multiple steps, utilizing a combination of ion exchange, hydrophobic interaction, size exclusion and affinity chromatography. The purity of the protein of interest is most usually estimated by SDS-PAGE electrophoresis. Bradford or Lowry assays can be used most often to measure the concentration of the target protein. In this project, purification of all the protein targets utilized ion exchange, hydrophobic and gel filtration chromatography.

2.2.1 Cell Lysis

For all protein targets the first step in the purification was to lyse the cells that contained the desired protein. This was performed by sonication, which uses high frequency sonic pulses to disrupt the bacterial cell walls and membranes. Cells were defrosted, suspended in 50mM tris-HCl buffer, pH 8.0 and disrupted by sonication using two bursts of 20sec at 16micron amplitude. Debris was removed by centrifugation at 70000g for 10minutes.

2.2.2 Ion Exchange Chromatography

Separation of proteins by ion exchange chromatography depends on the differences in their net surface charges. Protein charges depend on two factors; the kind of amino acids that are located on a protein surface and the pH of its medium. Every protein has a particular pH, called the isoelectric point (pI), at this point the overall charge on the protein is zero. If the pH of the protein solution is lower than protein pI, the protein charge will be positive, but the overall charge on the protein will be negative when the pH is above the protein pI. There are two main types of ion exchange chromatography; anion exchange

chromatography utilises positively charge exchangers to separate negatively charged particles and cation exchange chromatography uses a negatively charged exchanger to bind positive particles. At low ionic strength, sample molecules in the mobile phase will bind to a column of beads (stationary phase) in accordance to their charges. The ionic strength of the mobile phase can then be increased gradually, by adding for example sodium chloride or another salt in a controlled manner. This gradual increase in ionic strength causes elution first of low net charge density proteins and gradually those with higher charge density.

2.2.3 Hydrophobic Chromatography

Separation using hydrophobic chromatography depends on the reversible adsorption of biomolecules according to their hydrophobicity. This method is normally used as a subsequent step in an isolation project when the protein samples have already undergone an ammonium sulfate precipitation step or after preliminary separation by ion exchange chromatography. In these cases, the sample contains a high salt concentration, which increases the interaction between the hydrophobic areas on the surface of a protein and hydrophobic groups attached to a matrix. The protein is thus generally applied to a column under high salt conditions, and eluted with decreasing concentrations of salt in a buffer. Phenyl Toyopearl is an example of a hydrophobic chromatography matrix, and consists of a matrix of cross-linked Toyopearl forming beads with phenyl groups exposed on its surface.

2.2.4 Gel Filtration

The gel filtration technique, which is sometimes called size exclusion chromatography, separates molecules according to differences in their size as they pass through a column filled with porous beads. Unlike ion exchange or affinity chromatography, molecules do not bind to the chromatography medium, which mean that buffer composition does not directly affect the resolution (the degree of separation between peaks). The column medium can be considered as a set of beads that contain pores of different sizes. Large molecules, that are unable to enter these pores, move around the outside of the beads, whereas molecules that are small enough to enter the holes in the beads have a longer

path before they exit the beads. Therefore, the large molecules will elute first from the column and the smallest molecules will elute last.

For the three proteins whose structures are described in this thesis, a combination of ion exchange, hydrophobic, and size exclusion chromatography was used to prepare samples for crystallization. The precise details of columns and buffer used, are described the relevant sections of chapter 3 for GluDHs from *C. symbiosum* and *M. smegmatis*, and chapter 5 for PheDH from *B. sphaericus*.

2.3 Protein crystallization

2.3.1 Robot screening

Crystallization conditions for the native proteins and their complexes were initially screened using commercial NeXtal suites (PACT, JCSG, PEG, pHClear, Classics and AmSO4) from Qiagen, with a sitting-drop vapour-diffusion approach using a Matrix Hydra II Plus one robot (Thermo Fisher Scientific, USA) by adding 0.2 μ l of the protein to the same volume of precipitant, for each well of the 96 conditions, and then equilibrating against a 40 μ l reservoir of the same precipitant. The plate was centrifuged at 2000rpm for two minutes to mix protein and precipitants before incubating at 17°C.

Full crystallization conditions are described in sections 3.2 and 3.7.2 of chapter 3 for *M. smegmatis* GluDH and *C. symbiosum* GluDH respectively, and section 5.1 of chapter 5 for *B. sphaericus* PheDH.

2.3.2 Optimization trials

For some of the structures determined, diffraction quality crystals were only obtained after optimization of the initial screen conditions. This was carried out using the hanging drop vapour diffusion method. Similar volumes from protein and precipitant were mixed in a ratio of 1:1 and placed on a pre-siliconised cover slip. The cover slip was placed upside down on the reservoir containing 1 ml of a precipitant solution and sealed with grease.

2.4 X-ray Crystallography

X-ray crystallography is a technique used in structural biology and materials science

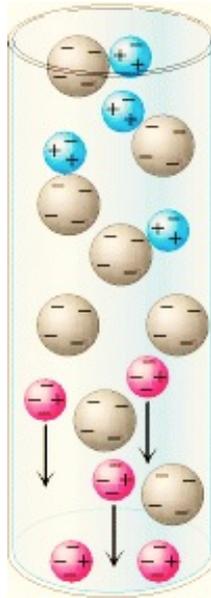


Figure 2.3. Ion-exchange chromatography. This method separates proteins depending on their net charges. Positively charged protein binds to the negatively charged bead. Negatively charged protein follows through. This figure has been adapted from Stryer, 1995.

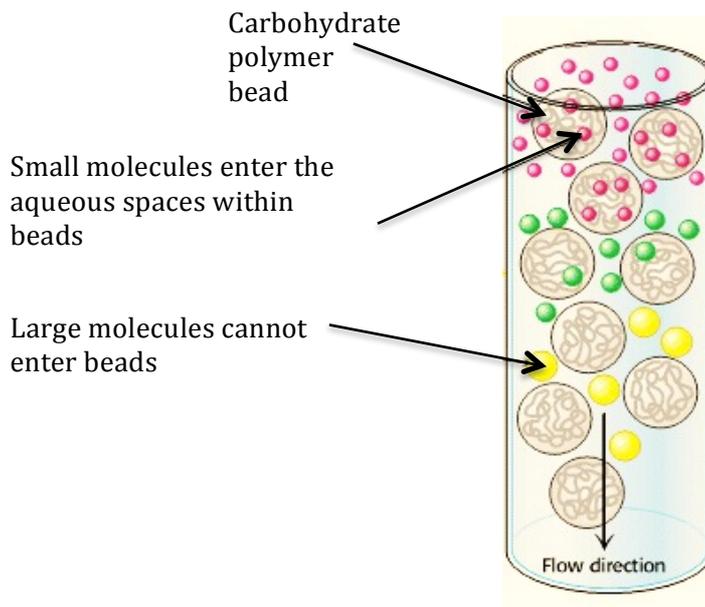


Figure 2.4. Gel filtration chromatography. A mixture of proteins in a small volume is applied to a column filled with porous beads. Because large proteins cannot enter the internal volume of the beads, they emerge sooner than do small ones. This figure has been adapted from Stryer, 1995.

for the determination of the atomic structure of a crystal. Solving a protein structure using X-ray methods commonly involves a number of separate steps, growing a crystal (crystallization), data collection, data processing, determining phases and structure refinement. X-ray diffraction theory is covered in a number of text books for example; *Crystallography made crystal clear*, by Rhodes (Rhodes, 2010) and “*Principles of protein x-ray crystallography*” by Drenth (Drenth, 2007). This chapter describes the X-ray crystallographic techniques used for determining the protein structures described in this project.

2.4.1 Protein crystallization

The formation of protein crystals is essential to the structure determination of protein by X-ray crystallography. However, due to variations in size, differing charge distribution, and stability of protein molecules, the conditions required for obtaining crystals of various proteins vary substantially and the crystallization process is complicated and time consuming. Protein crystallization is a phase transition phenomenon. Normally, a protein is soluble in an aqueous solution at a certain concentration, but if the protein concentration changes to more than its solubility limit, a protein might precipitate to form an ordered crystalline array (a crystal), or under some conditions, it may form an amorphous precipitate.

Protein crystallization is dependent on a number of factors including the purity and concentration of the protein sample, pH, temperature, and the concentration and kind of the precipitants used. For crystallization, the purity of the protein must usually be more than 95%. Not only should impurities be absent, but also the protein sample should be homogeneous, that is, all protein molecules should have the same protein conformation. A starting concentration of approximately 10mg/ml is often used, to aid supersaturation of the sample for crystallization. The temperature for most crystallization conditions is usually around 17°C, but this can be varied if necessary. Precipitants such as ionic compounds (ammonium sulphate) and organic compounds (polyethyleneglycol PEG) are the generally used agents for bringing the protein to supersaturation to induce crystal formation. The effect of these agents is essentially on the solvent rather than on the solute and they cause precipitation of protein out of solution.

Formation of a protein crystal can be broken down into two stages: the nucleation stage and crystal growth. As shown in the phase diagram (figure 2.4), formation of a nucleus and subsequent crystal growth happened within the supersaturation region, when suitable conditions have been determined. In the nucleation step, where the level of supersaturation is higher than that for growth, small aggregates of protein are formed, as the protein molecules come close together, producing the nuclei of crystals, from which the crystals will finally grow. Because the protein molecules coming out of solution to form the nuclei, the concentration of protein in solution decreases and the growth phase of the crystal is entered. In this phase, the protein concentration has dropped, thereby leading the system into the metastable zone, where no more nucleation happens, but a few crystals will grow by adding ordered layers of protein onto the nuclei. The supersaturated state of the protein solution in the nucleation phase is crucial for protein crystallization. This is because high saturation of the protein solution will cause a rapid growth of many crystal nuclei, leading to an excess of small crystals. Therefore, to get large protein crystals it is essential that the nucleation stage should be quite short, so that not many nuclei are formed, and then an increased crystal growth phase should arise, which means the crystals grow large enough for diffraction data to be collected.

There is no recognized technique for predicting precisely under what circumstances a particular protein will form crystals. Therefore, a number of commercial screens such as the Nextal Suites and the Hampton Research Crystal Screens 1 and 2, can be used for initial screens of protein crystallization. These screens use several common conditions of crystallization. Furthermore, by using a modern robot, a great number of crystallization conditions can be screened quickly. Once some crystals have appeared, or even crystalline precipitate, optimization of this condition should be done, by varying the parameters (precipitant concentration, pH, temperature and additives) in an attempt to obtain large protein crystals.

2.4.1.1 Vapour diffusion method

The conversion from a stable solution to a supersaturated solution can be accomplished by increasing the concentration of the protein and the precipitant in the protein-precipitant mixture. The physical method that causes this change in the

concentrations can be performed through dialysis or diffusion. Vapour diffusion is the most prevalent method in protein crystallization. In this method a crystallization solution is mixed with the same volume of the protein solution to form a drop, which is placed in an enclosed environment comprising a larger reservoir of the precipitant solution. Because the concentration of the solution in the sample is less than that of the reservoir, water vapour moves from the sample drop into the reservoir, until a supersaturation state is reached and the protein might crystallize, which can take from a few days to weeks, or months.

There are two techniques using the vapour diffusion method: hanging drop and sitting drop.

2.4.1.1.1 Hanging drop vapour diffusion method

In this method, a drop (1-10 μ l) of the crystallization solution is placed on a siliconized microscope cover slip and mixed with an equivalent amount of a protein solution. The cover slip is placed upside down on the reservoir containing between 500 μ l to 1 ml of a precipitant solution and it is sealed on top of the reservoir with grease.

2.4.1.1.2 Sitting drop vapour diffusion

The sitting drop vapour diffusion method depends on the same theory as the hanging drop vapour diffusion in using a closed system, but in this method, the protein/precipitant drop is usually placed on a micro bridge above the reservoir solution. Sealing tape can be used instead of a cover slip for an airtight closure on the system. The sitting drop method is used preferentially with proteins that have a low surface tension, which can make it difficult for a hanging drop to form. The other benefit of this method is that it is easily automated and therefore it is utilized by many commercially available crystallization robots.

2.4.2 Cryocooling

Protein crystals are very sensitive to X-rays, and can undergo considerable radiation damage when exposed to an X-ray beam at room temperature. This damage can seriously reduce the quality of the data that can be collected and thus the structural information that can be extracted from the crystals. Radiation

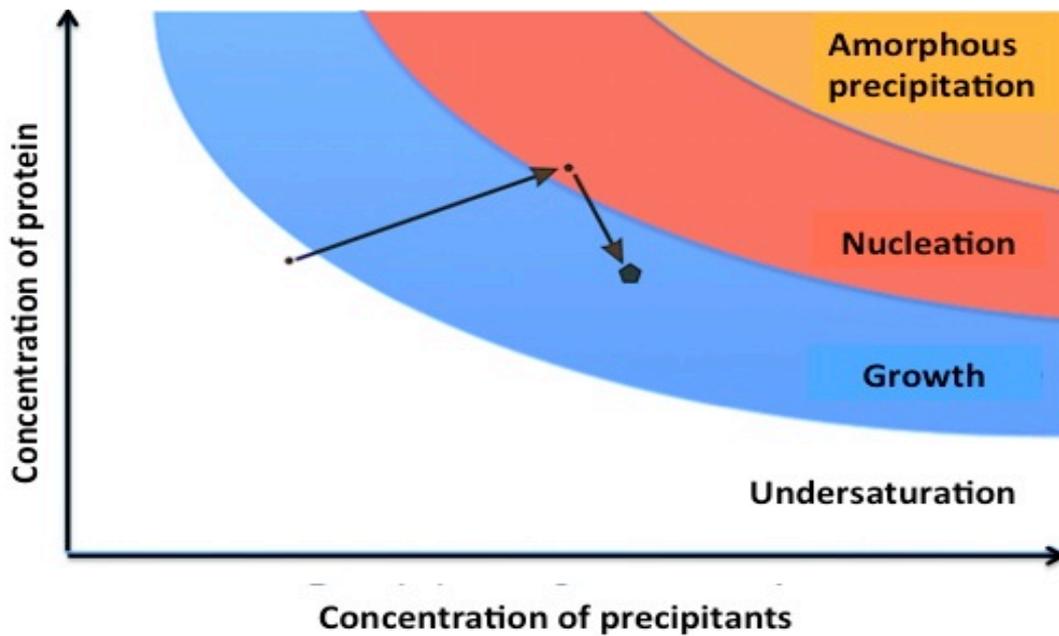


Figure 2.5. Phase diagram of protein crystallization. The scheme shows protein concentration against precipitants concentration during a crystallization vapour diffusion trail. The white zone illustrates an undersaturation region where crystallization will never occur. Because of the protein and precipitants drop equilibrates with the reservoir solution the concentration of protein and precipitants increase gradually into the nucleation region (red region in the plot). The protein then starts to come out of solution and so the concentration drops hopefully into the growth zone (blue zone in the plot), where crystal growth will occur. This scheme has been adapted from Rhodes (Rhodes, 2010).

damage is caused mainly by a primary interaction between the X-ray photons and the molecules in the protein crystals, which produces free radicals. The extent of the initial damage is X-ray dose dependent, but more damage can happen when the heat produced through thermal vibration of the molecules is large enough for the diffusion of the reactive radicals through the crystal. In order to reduce the extent of this secondary damage, a protein crystal can be kept at a cryo-temperature (around 100° K), which has the effect of freezing the reactive radicals in the crystal, and reducing the damage. Cryo temperatures are accomplished in the experiment by positioning the crystal in a stream of gaseous nitrogen at 100°K on the collection devices. Before cooling, the crystal should be transferred to a cryoprotectant solution, containing the precipitant solution used to grow the crystal plus an anti-freeze agent such as a low molecular weight polyethylene glycol (PEG) or glycerol. The presence of these agents prevents the formation of ice crystals that could interfere with the diffraction from the protein crystal. The crystal is placed in the cryoprotectant solution, and is then harvested in a thin nylon loop and is held by surface tension in the drop solution inside the loop and stored in the liquid nitrogen. The magnetic base of the loop is attached to the goniometer head, which allows the crystal to be rotated perpendicularly during data collection and for the precise alignment of the crystal into the path of the X-ray beam.

2.4.2 X-ray generation

X-rays are electromagnetic radiation that exists over the wavelength range of 0.1 Å (hard X-rays) to 100 Å (soft X-rays). The wavelengths of the X-rays that are used in crystallography usually lie between 2.0 Å and 0.5 Å, and they penetrate biological macromolecules sufficiently to be scattered from the crystal. In crystallography laboratories X-rays are produced by various sources, such as a rotating anode generator, but very high intensity X-rays can also be generated at synchrotrons.

2.4.2.1 Rotating anode generator

A rotating anode generator is the most popular X-ray generator used in the protein crystallography laboratories. In this device, electrons are liberated from the heated

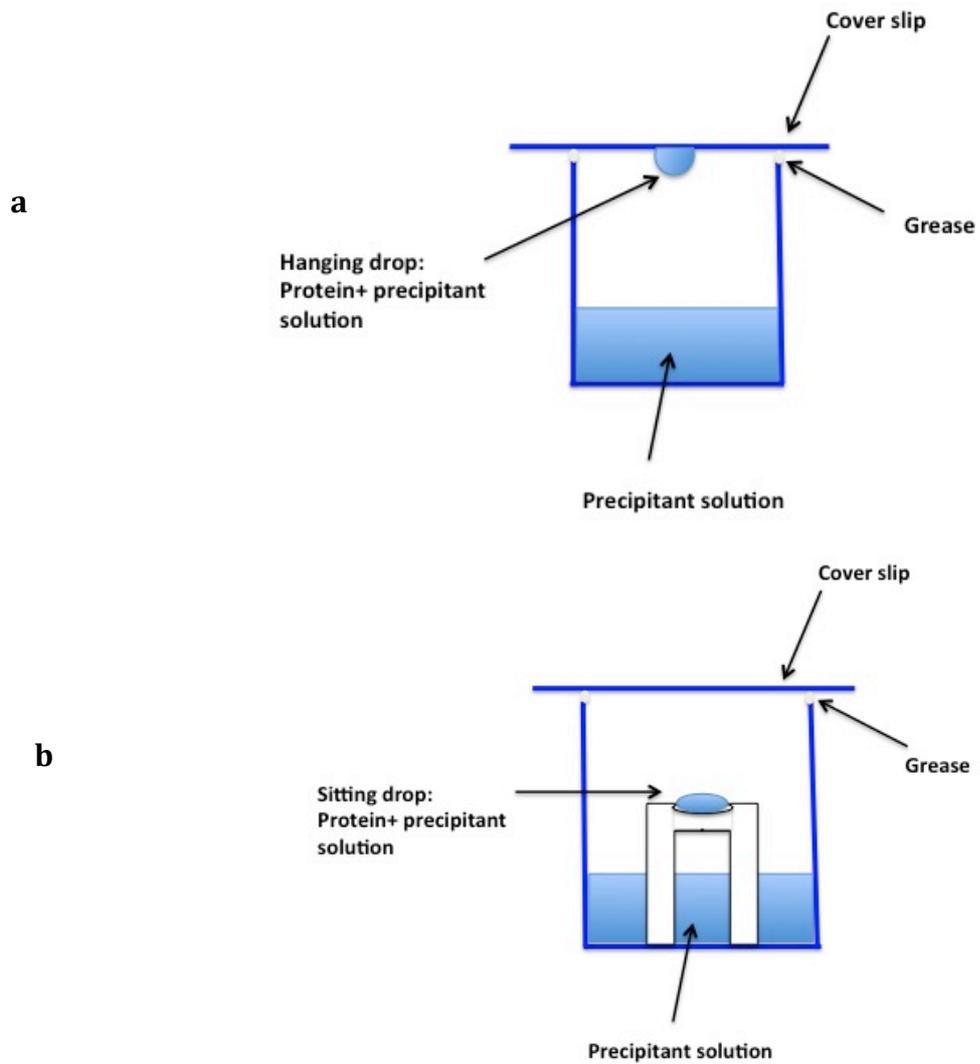


Figure 2.6. Illustrations of vapour diffusion methods (a) hanging drop and (b) sitting drop.

tungsten cathode and are accelerated towards the anode, through a high potential difference (typically 40kV) inside the evacuated lower. These accelerated electrons hit and interact with the anode material producing a beam of X-rays. In protein crystallography the anode is generally a copper cylinder rotating at high speed. A large proportion of the electron energy is converted to heat, which is removed by cooling the anode with circulating water. A copper anode emits X-rays at two wavelengths, $\text{CuK}\alpha$ ($\lambda=1.54\text{\AA}$) and $\text{CuK}\beta$ ($\lambda=1.392\text{\AA}$), which are the result of L to K shell electron transitions. For crystallography uses, X-rays of a single wavelength are needed, and therefore a graphite crystal monochromator or focusing mirrors can be used to select the stronger $\text{CuK}\alpha$ X-rays, and provide X-rays of a single wavelength at 1.542\AA .

2.4.2.2 Synchrotron radiation

Synchrotron radiation is electromagnetic radiation which is produced in a synchrotron, when charged particles are accelerated to very high speeds using an electric field and are forced to travel in a circular path using an electromagnet. Synchrotrons are constructed as polygons, with straight segments lined by bending magnets. These bending magnets are located at the corners of the polygon, which deflect the electrons to promote their passage around the ring. When the electrons change direction they change angular velocity, resulting in the release of X-rays at a tangent to their path. To produce even more intense X-rays, "insertion devices" could be inserted into the straight segments of the synchrotrons. These devices extend the spectrum to shorter wavelength and can be divided into two kinds: wigglers that produce a number of sharp extra bends in the electron trajectory, and undulators that are similar to wigglers but cause the emission of radiation at more specific wavelengths.

Synchrotron radiation sources have many advantages over the rotating anode X-ray source. The intensity of the beam is very much greater. This increase in intensity can aid collecting data on small or poorly diffracting crystals. Synchrotron radiation also allows the collection of data from crystals with large unit cell dimensions because of the lower divergence of the beam, making the diffraction spots sharper and thus avoiding overlap. Another advantage of synchrotron radiation is the ability to tune

the wavelength (to the absorption edge of heavy metals bound to the protein or chemically incorporated into the protein), allowing multiwavelength anomalous dispersion (MAD) experiments to be conducted.

2.5 Data processing

After collection of the diffraction pattern for the protein crystal the data must be indexed and integrated together to generate a complete dataset. The process of indexing determines the dimensions of the unit cell and the crystal space group by measuring the distance between diffraction spots and symmetry of the diffraction pattern. When the parameters of crystal and detector have been estimated to high accuracy, the integration of the data set can be achieved. The process of integration determines the intensities of each reflection. This process is accomplished automatically using MOSFLM (in home source) and Xia2 at Diamond Light Source with XDS and XSCALE (Winter, 2010).

The data set are scaled and merged by the program SCALA from ccp4 (Collaborative, 1994), which sums the multiple observations of reflections and merges these observations into average intensity. In this program the differences in reflection intensities produced by external effects (such as variation of the beam strength during an experiment) are excluded. The outputs of SCALA a *.mtz file includes the determined average scaled intensity for each reflection (I), the associated statistical errors (σI) and indices (h, k, l). The quality of scaled and merged data set is assessed using various methods of statistical analysis. The major indicators for assessment of the quality of data and resolution are the signal to noise ratio ($I / \sigma I$), R_{merge} and R_{pim} and completeness. The signal to noise ratio is the ratio of the intensity over the standard deviation for that intensity ($I / \sigma I$) and, in general, values lower than 2 indicate bad quality data. R_{merge} and R_{pim} calculate the agreement between multiple measurements of the same reflection. They differ in the application of a correction factor in R_{pim} that takes the multiplicity (the number of times of the same reflection are measured during the course of data set) of the data into account. Completeness is a measure of how much of the theoretically recordable data has actually been recorded. Values lower than 90% show poor quality data.

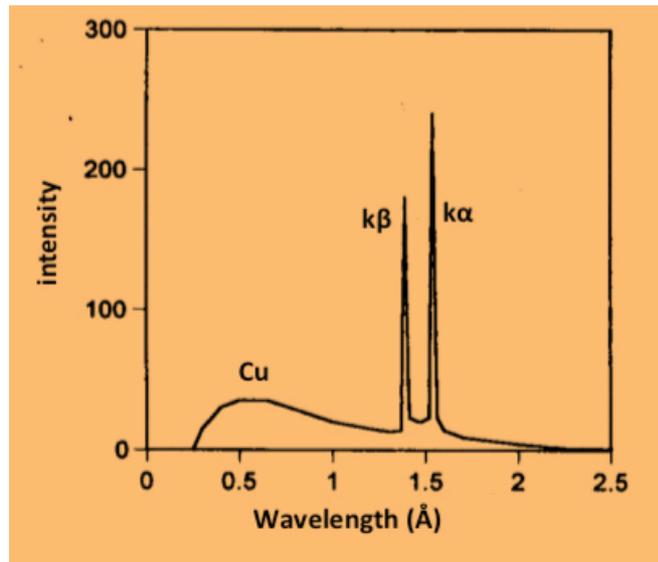


Figure 2.7. X-ray emission spectra generated using a copper target.

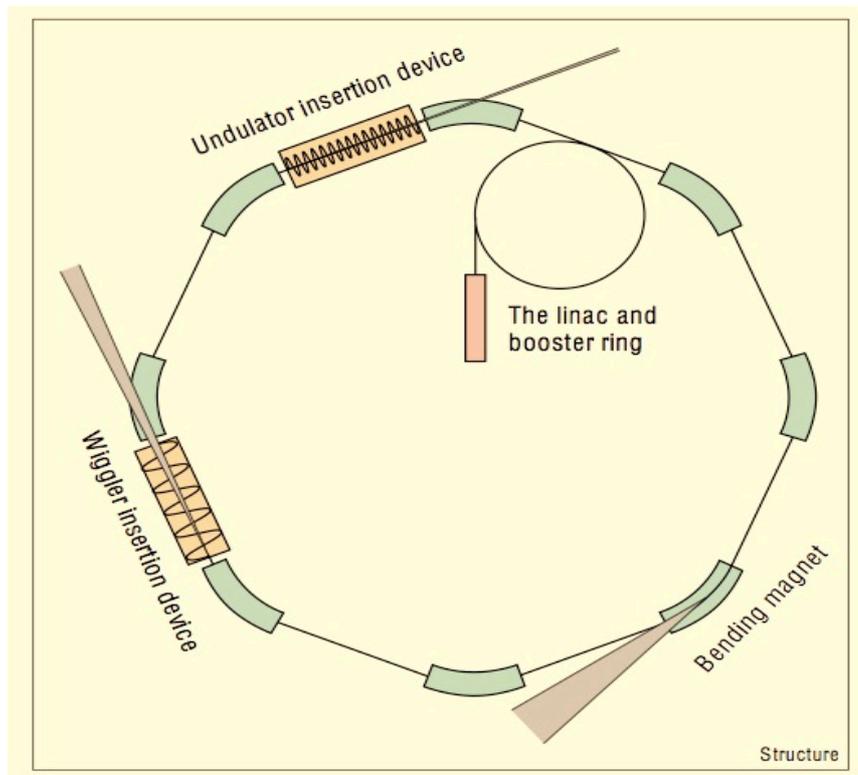


Figure 2.8. Schematic of a synchrotron radiation source illustrating the linac, booster and storage rings and magnetic devices that generate X-ray radiation. Green colour is the bending magnet; orange is the insertion devices and grey is the generated X-ray from each device. This figure has been adapted from (Mitchell et al, 1999).

The data were also run through the programme Matthews coef (Kantardjieff & Rupp, 2003) to predict the solvent content in the crystal and the number of molecules in the asymmetric unit based on the molecular weight of the protein, the unit cell volume and space group.

2.6 Determination of Phases by Molecular Replacement

Molecular replacement can be used to solve the phase problem in X-ray crystallography (Rossmann, 1990). MR depends on the presence of a previously solved protein structure, which is similar usually with more than 25% amino acid sequence identity to the unknown target structure from which the diffraction data are obtained and then could be used in determination of the target structure.

Molecular replacement utilizes the already known protein structure for structure determination of a related protein, such as a mutant protein, or a protein with close sequence similarity. If the model of the known structure is close to that of the target structure then an electron density map can be calculated using these calculated phases from the model, along with the structure factor amplitudes measured in the diffraction. For initial phases to be calculated, the model must be placed in the same orientation in the unit cell as the target model. Molecular replacement is therefore divided into two steps; the rotation function (orienting the model) and the translation function (moving the oriented model to the position occupied by the unknown).

2.6.1 Rotation Function

To estimate the orientation and position of the unknown structure the Patterson function can be used. A Patterson function can be considered as a vector map, i.e. it produces an electron density map of the vectors between scattering objects in the unit cell. The Patterson function $P(uvw)$ is determined independently of the phases by squaring the $\mathbf{F}(hkl)$ term for each reflection and it described as:

$$P(hkl) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \exp[-2\pi i(hu + kv + lw)]$$

The Patterson function for the rotated search model is compared to the Patterson function of the target molecule through all possible rotations. If both

are in the same orientation, the summation of the peaks will produce a large value, suggesting the rotation solution is correct.

2.6.2 Translation Function

When the right rotation function has been recognized, the position of the molecule in the unit cell (the three positional parameters) must be found. In the translation function a new Patterson is calculated on the conversely orientated (rotated) search model and the molecular replacement process searches through the three position parameters using inter-molecular vectors to find the position of the molecule in the unit cell, using the process developed by Crowther and Blow (Crowther & Blow, 1967).

Having placed the search model in the right orientation and position in the unit cell, an initial electron density map can be calculated by a Fourier transform, using the observed F and the calculated phases from the model. Phaser is a commonly used program for molecular replacement (McCoy et al, 2007).

2.6.3 Refinement and Re-building

Once the appropriate model and map have been determined using the molecular replacement, rounds of rebuilding and refinement should be done to improve the agreement between the observed and the calculated structure factor amplitudes.

Rebuilding of the unknown model in this thesis was accomplished using the program COOT (Emsley et al, 2010). Rebuilding includes changing the amino acids residues in the model to those in the target structure using the knowledge of the amino acid sequence of the target structure, to build an initial model of the protein structure. In addition, repositioning of side chains, adding water molecules, interpreting density features and modeling ligands are all part of the rebuilding process. After rebuilding, the model and observed amplitudes are input into Refmac5 (Murshudov et al, 2011) for refinement. During refinement the observations (structure factor amplitudes) are fixed and the parameters of the model (x , y , z co-ordinates, scales, and B-factor) are varied to give maximum agreement between observed and calculated structure factor amplitudes. The model is refined again and a new electron density map can then be calculated. Repeated cycles of refinement and model building provide better phases and

better electron density maps are obtained. Refmac5 gradually minimizes the differences between the observed structure factor and the calculated structure factors and a complete model can be built.

2.6.4 Rigid Body Refinement

Usually the first stage of refinement treats segments of the model as rigid bodies, which are refined as a unit (rather than separate atoms). This is called rigid body refinement. Rigid body refinement uses low-resolution data and allows more accurate placement of a model obtained by molecular replacement into an electron density map. Rigid body refinement is useful when refining models with multiple independent domains as it allows positional refinement of each domain individually. In this thesis the program suite Phenix was used for rigid body refinement (Adams et al, 2010).

2.6.5 Monitoring the Progress of Refinement

The progress of refinement is usually determined by the crystallographic R factor that is defined as the fractional differences between the observed structure factor amplitudes, from the experimental data, and calculated structure factor amplitudes produced from the model. The R factor can be described as:

$$R = \sum_{hkl} (|F_{obs}| - k|F_{calc}|) / \sum_{hkl} |F_{obs}|$$

where k is a scale factor. A low R factor means that the model agrees well with the diffraction data. Hypothetically the values of R factor can range from zero, for a model that entirely corresponds to the data, to more than 0.5, would indicate a low agreement between the two. The zero value of the R factor would be perfect, but because of the limits of resolution it is not possible to precisely determine the accurate position of all atoms in the unit cell and it is acceptable to obtain an R factor value roughly 10% of the data resolution. A free R factor, R_{free} is determined by selecting a small random subset of the diffraction data that is omitted from the refinement process, so as to prevent incorporation of bias. The value of the R free should be within 5% of the R factor, if the

differences between Rfactor and R_{free} are much higher, this suggests that the model has been over fitted to the data and is biased.

Chapter Three: Cloning, Expression, Purification, Crystallization, Data Collection and Structure Determination of *M. smegmatis* GluDH.

3.0 Overview

This chapter will explain the results for identification, amplification, cloning, overexpression, purification, crystallization, data collection and structure determination steps for the glutamate dehydrogenase protein from *M. smegmatis*. The purification, crystallization, data collection and structure determination of glutamate dehydrogenase from *Clostridium symbiosum* also has been described in this chapter.

3.1 Identification, amplification, cloning, overexpression, purification of *M. smegmatis* GluDH

3.1.1 Identification of Target Gene

The GluDH of *M. smegmatis* is encoded by the *msmeg_5442* gene and it was identified by searching the UniProtKB database to find both the DNA and protein sequence. A BLAST search at the NCBI was used to align the glutamate dehydrogenase protein sequence of *Mycobacterium smegmatis* with other glutamate dehydrogenases in the database.

3.1.2 PCR for amplification of *M. smegmatis* GluDH gene (*msmeg_5442*)

The Polymerase Chain Reaction (PCR) was used to amplify the *M. smegmatis* GDH gene (*msmeg5442*) from *M. smegmatis* genomic DNA (strain ATCC 700084 / mc (2)155). Primers were designed based on the nucleotide sequence and included restriction sites for NdeI and HindIII as described in section (2.1.1). The PCR products were run on a 1% agarose gel, and a 1.3 kb band was seen, as expected for the *msmeg5442* gene (1300bp). The PCR products were purified using QIAquick® gel extraction kit (Qiagen) and were stored at -20°C prior to use in cloning. Figure 3.1 shows the PCR products at the expected size.

3.1.3 Cloning of *M. smegmatis* GluDH gene (msmeg_5442)

The msmeg5442 gene PCR product was digested with the restriction enzymes NdeI and HindIII, before insertion into a pET21a plasmid, which was cleaved with the same enzymes (see section 2.1.4). After incubation with the digestion mixture for 2 hours at 37 °C, the samples were analyzed on a 1% agarose gel and the bands appeared at the expected sizes (figure 3.2). The digested insert and vector were purified using a QIAquick® gel extraction kit (Qiagen) prior to setting up the ligation reaction (see section 2.1.5). The ligation mixes were transformed into competent DH5α *E. coli* cells (Novagen) (see section 2.1.5.1). As plasmid pET21a carries an ampicillin resistance gene, the resulting transformants were spread on LB-agar plates supplemented with 50µg/ml ampicillin, and the plates were incubated overnight at 37°C. Some colonies could be seen on the plate and were screened for successful cloning.

3.1.4 Identification of possible clones

An initial screen of the successful clones was carried out by restriction digestion. A number of colonies were picked up from the plated clones and the possible constructs were extracted using a miniprep kit (Qiagen) (see section 2.1.3). The resultant miniprep samples were digested with the restriction enzymes NdeI and HindIII (see section 2.1.4) and again the samples were run on the 1% AGE to verify successful ligation. From the gel picture (figure3.2) it can be seen that all the clones gave bands at the expected size (1300 bp) showing the existence of the desired gene in these clones.

3.1.5 Expression of *M. smegmatis* GluDH

3.1.5.1 small-scale over-expression

The confirmed construct was then transformed into a competent *E. coli* strain, BL21(DE3) (Novagen), and plated onto an Amp agar plate. Firstly, a small-scale overexpression experiment was done (see section 2.1.6.2 for details) to test the expression of the cloned gene and the protein solubility. A single colony from the transformed BL21(DE3): pET21a-msmeg5442 plate was picked and used to prepare a 10ml starter culture using LB medium and grown overnight. One ml from this overnight culture was then used to inoculate 50ml of LB media

containing 50µg/mL Amp in a conical flask yielding 2% inoculation. Three separate cell cultures were carried out from the starter culture to determine the protein solubility at various temperatures; 37°C, 30°C and 25°C. The cells were left to grow at the desired temperature with shaking at 200rpm until the absorbance at 600nm reached about 0.6. After that, one ml from each of the cell cultures were transferred into Eppendorf tubes, and pelleted, before storage at -20°C. These were used on SDS-PAGE as the pre-induction control samples. Once the optical density of the cells had reached 0.6, IPTG was added to a final concentration of 1mM to induce GluDH protein expression by these cells. After 4 hours induction at 37°C, 30°C and 25°C on a rotary shaker at 200rpm, the cells were harvested by centrifugation at 1844 g at 4°C, for 20 minutes, and the cell pellets were stored at -80°C for subsequent SDS-PAGE analysis. Investigation of the expression was carried out on cell pastes.

The cell paste was defrosted, suspended in 50mM tris-HCl buffer, pH 8.0 and disrupted by sonication using two bursts of 20sec at 16micron amplitude. Debris was removed by centrifugation at 70K g for 10 minutes. The supernatant was transferred into the universal tube and pellets were suspended in the same buffer and disrupted by the same way. The un-induced sample was prepared using the following method: the cell pellets were resuspended in 100µl (1X) SDS loading buffer plus 7% β-mercaptoethanol by pipetting and vortexing. The β-mercaptoethanol reduces any disulfide bonds, allowing complete denaturation. The cell solution was then boiled at 100°C for 10 minutes prior to centrifugation at 13000 rpm for 1 minute in a bench top centrifuge. The sample is then ready for loading on SDS-PAGE gel. 15µl from soluble (supernatant) and insoluble fraction samples with 5µl loading buffer were boiled for 10 minutes before loading into 12% SDS-PAGE gel. The three samples, soluble, insoluble fraction and pre-induction, were then analysed by 12% SDS-PAGE gel electrophoresis for checking expression level and protein solubility. In each case the majority of the protein had been over-expressed in the insoluble fraction as seen by SDS-PAGE analysis (fig 3.3), but fortunately, there was some protein in a soluble form.

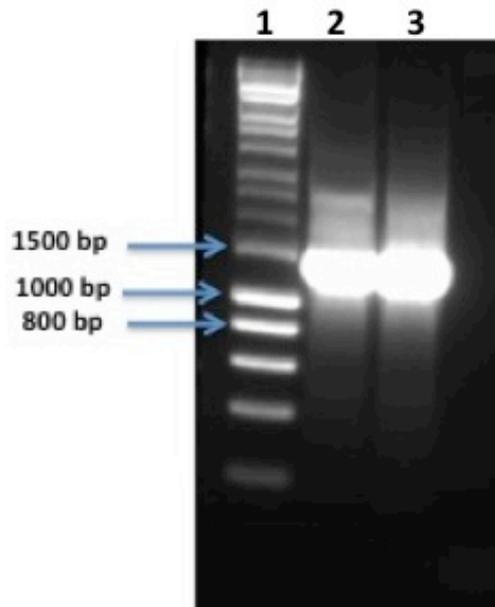


Figure 3.1 Electrophoretic pattern of agarose gel. PCR products of GluDH gene from *M. smegmatis*. The gel agarose picture shows that there are bands at the expected size (1.3kb), compared to the gene of interest. Lane 1 is a hyperladder I and Lanes 2 and 3 are PCR products for the desired gene.

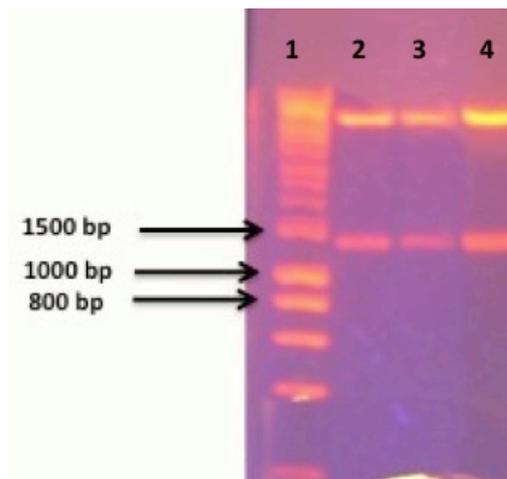


Figure 3.2 A picture of 1% agarose gel electrophoresis showing the screening results for the possible constructs of pET21a with *M. smegmatis* gene after digestion with NdeI and HindIII. Lane 1 is a hyperladder I and lanes 2, 3 and 4 are digested samples. All of these constructs represent the correct cloning.

3.1.5.2 Large-scale over-expression

Because there was no difference in protein solubility in the over-expression tests at the three temperatures, large-scale over-expression was under normal conditions (growth at 37°C with 1mM IPTG). Two 2L Erlenmeyer flasks containing 500 ml LB, supplemented with 50µg/mL Amp, and inoculated with the 10ml of starter culture were used. The flasks were left to grow at 37°C with shaking at 200rpm until the A_{600} reached 0.6, before adding IPTG for induction. The cells were allowed to over-express for four hours, under the same conditions. The cells were centrifuged at 1844 g for 20 min for harvesting and the cell pastes were stored at -80°C prior to use in purification experiments. 1ml samples were collected at the pre-induction and post-induction growth for verification at the expression and solubility levels. Figure 3.3 illustrates the large-scale protein expression growth.

3.1.6 Purification of GluDH from *M. smegmatis*

3.1.6.1 Ion exchange chromatography

About 2 g of the cell paste was resuspended in the lysis buffer 50mM TRIS pH=8.0. Insoluble materials were separated by centrifugation at 70,000 g for 15 minutes and the resultant cell free extract (the total protein about 330mg) was applied on a 10ml DEAE FF cartridge (two 5ml cartridges jointed together). Proteins were eluted by 100ml gradient of NaCl from 0 to 0.5M in buffer A (50mM TRIS pH=8.0). 4ml fractions were collected and GluDH activity was checked in the fractions according to the assay described below. The activity is represented on the chromatogram (figure 3.5 **b**) by the black line. Fractions were also analysed by SDS-PAGE (12.5% tris-glycine gel) (Laemmli, 1970) (figure 3.5 **a**).

3.1.6.2 GluDH assay

To assay the samples for GluDH activity samples were prepared in 1ml plastic cuvettes. The cuvette contained 20mM Na⁺ glutamate, 10mM NADP⁺ in 0.9ml of 50mM Na⁺ phosphate buffer pH 7.0. 0.1ml of the protein solution from the DEAE column was added to the cuvette, the solution was mixed and placed in a

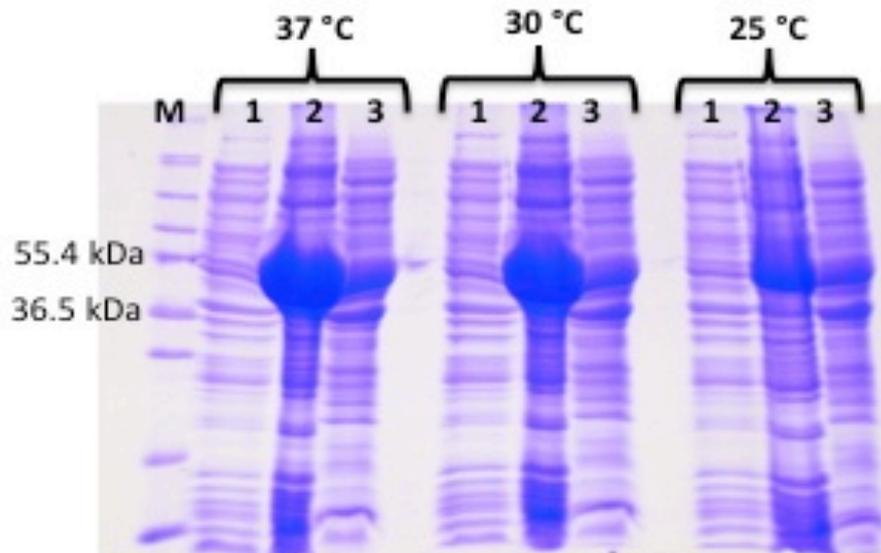


Figure 3.3 SDS gel showing the level of GluDH overexpression and solubility at different temperatures (37°C, 30°C and 25°C). M= Mark 12, 1= pre-induction, 2= insoluble fraction. And 3= soluble fraction. Mwt of GluDH = 48 kDa.

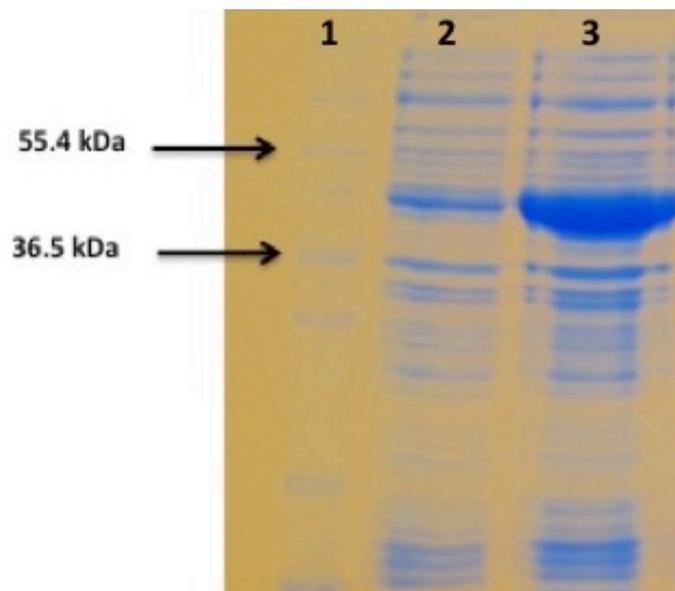


Figure 3.4: An SDS gel displaying the level of *M. smegmatis* GDH expression during large-scale growth at 37°C with 1mM IPTG. Lane 1 is a Mark12, lane 2 is a pre-induction sample and land 3 is a post-induction sample.

Helios spectrophotometer. The absorbance of the NADPH produced by the reaction was monitored at 340nm for 1min.

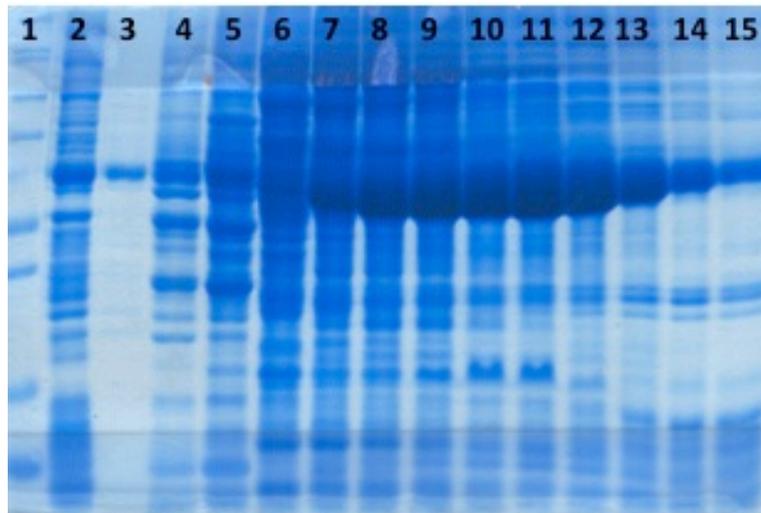
3.1.6.3 Hydrophobic and gel filtration chromatography

The most active fractions 23-26 (figure 3.5 **b**) were combined, (total protein was 64mg) and 4M ammonium sulphate solution was added to the sample to make the final ammonium sulphate concentration 1.5M. The sample was then applied to a 5ml Phe-HP cartridge. Elution was performed by 50ml ammonium sulphate concentration gradient from 1.5M to 0 in buffer A. 2.5ml fractions were collected and analyzed for GluDH activity and by SDS-PAGE (12.5% a Tris-glycine gel, Laemli) (figure 3.6, lanes 3-9). The most active fractions 8-10 were combined (10mg in total protein) and the volume of the sample was reduced to 2ml using a VivaSpin concentrator with molecular weight cut off of 30kDa. Then the sample was applied to a gel filtration column equilibrated with buffer A + 0.5M NaCl, and gel filtration was performed with the same buffer. 2ml fractions were collected after the void volume and were again analysed for GluDH activity and by SDS-PAGE (12.5% a Tris-glycine gel, Laemly) (figure 3.6, lanes 11-15). The fractions 7-11 were pooled to prepare GluDH sample for crystallization. The protein was concentrated using a VivaSpin concentrator and buffer was exchanged with 10mM Tris pH8.0 using a diafiltration cup. Finally the GluDH was concentrated to 12mg/ml, as estimated by method of (Bradford, 1976).

3.2 Crystallization of glutamate dehydrogenase from *Mycobacterium smegmatis*

M. smegmatis GluDH protein at concentration of 12 mg/ml was used directly after purification for crystallization trials. Two different samples, protein and protein in presence of 250mM L-glutamate (sodium salt, pH=7.5) plus 5mM NADPH, were automatically checked for crystallization using Matrix Hydra II Plus one robot (see section 2.3.1 for details). Crystals were grown in several conditions for both the native protein and the protein with substrate and cofactor. The best crystals for the native protein appeared in PEG screen well D7 containing 0.1M Tris-HCl pH=8.5 with 25%(w/v) PEG3000 and also

a



b

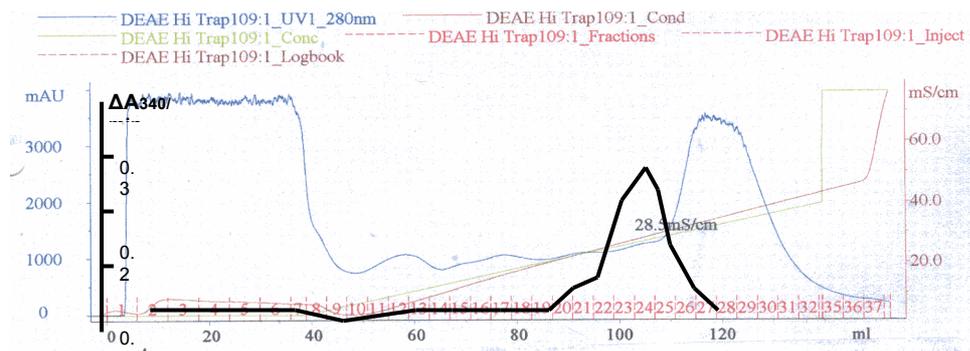


Figure 3.5. (a) SDS-PAGE gel showing the purification of GluDH using ion exchange chromatography. Lane order: 1= marker, 2= fraction 7, 3= fraction 10, 4= fraction 13, 5= fraction 18, 6= fraction 20 and 7-15= fractions 21-28. **(b)** Chromatogram represents the most active fractions after applying the sample on ion exchange.

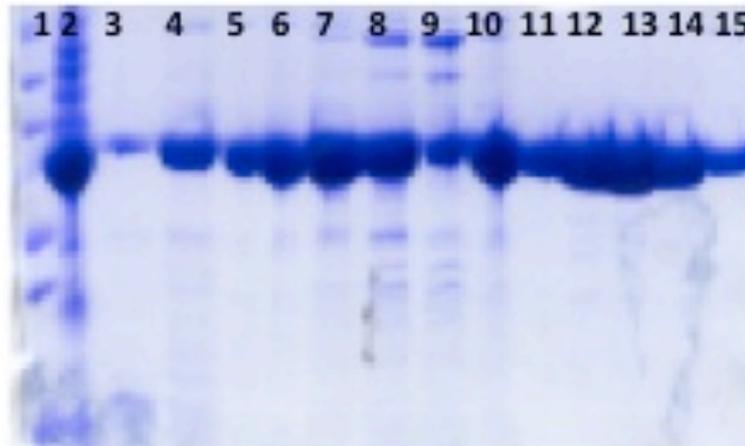
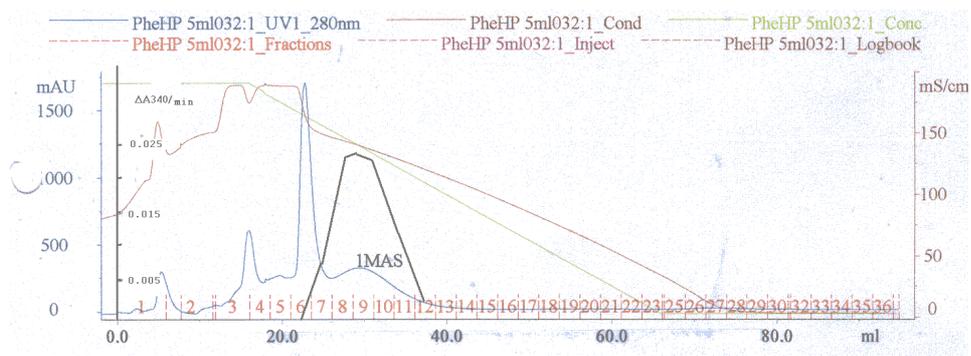


Figure 3.6. SDS-PAGE illustrates the purification of GluDH using Phe-HP and gel filtration columns. The molecular weight of GDH is 48.595 kDa. Lanes order as following: 1= marker, 2=sample applied on Phe-HP column, 3-9= fractions (6-12), 10=sample applied on a gel filtration column and lanes from 11-15= fractions (7-11).

a



b

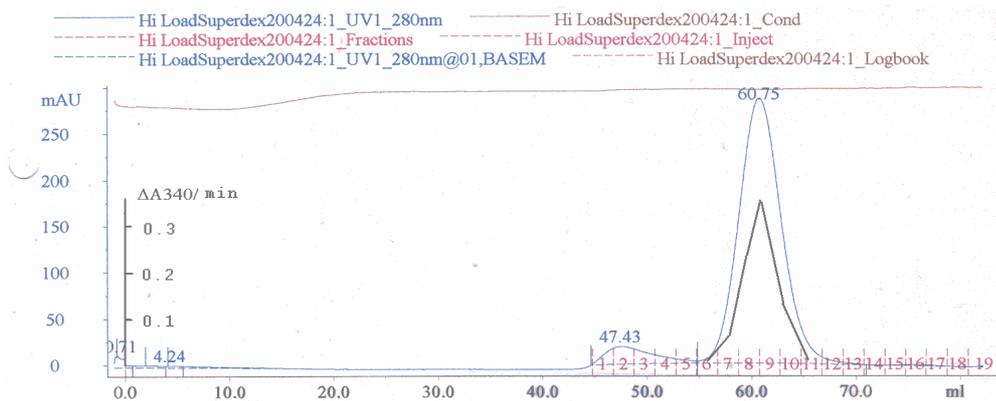


Figure 3.7. Chromatogram pictures for the purification of *M. smegmatis* GluDH. **a)** Hydrophobic chromatography (Phe-HP) purification stage showing the most active fractions 8-10. **b)** Gel filtration purification step showing elution with 2 ml fractions collected throughout.

PEG-D12, which comprises 0.1M Tris-HCl pH=8.5 plus 15% PEG20000 (figure3.10). For the protein with L-glutamate and NADPH the best crystals grew in the PACT screen well H1, containing 0.2 M sodium fluoride, 0.1 M Bis-Tris Propane pH=7.5 and 20% PEG3350 (figure3.10).

The *M. smegmatis* GluDH protein was also crystallized in the presence of 100 mM 2-oxo-glutarate and 10mM NADP⁺, again using commercial screens and the Hydra II Plus one robot (section 2.3.1). Crystals grew from condition D8 of the Classics screen, containing 0.1M Sodium HEPES pH 7.5 and 1.4 M tri-sodium citrate (figure 3.10). Crystals from this condition were optimized by varying the tri-sodium citrate concentrations from 1.0 to 1.8 M, a droplet containing 1 μ l protein plus 1 μ l well solution in a hanging drop vapour diffusion experiment.

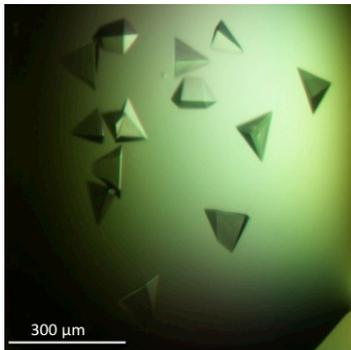
3.3 Data collection from crystals of *M. smegmatis* GluDH in apo state

Crystals were picked from those grown in the PEG suite, conditions D7 and D12 (figure3.8 a and b), according to their size and overall definition for initial diffraction analysis. The crystals were looped and soaked in cryo-protectant solution consisting of 30% ethylene glycol in the crystallization solution and immediately flash-cooled in liquid nitrogen before placing them onto the detector. Preliminary diffraction tests were performed using a Rigaku 007/Mar345 in-house detector system. Unfortunately, the diffraction from all the protein crystals of the enzyme glutamate dehydrogenase in the apo state was very poor with high mosaicity and overlapping diffraction pattern. One reason for this might be due to a movement in the domains of the hexameric GluDH, which causes disruption of the ordering of molecules inside the crystal, leading to poor diffraction. Therefore, further studies of this protein were carried out on the complex with its substrates and cofactor.

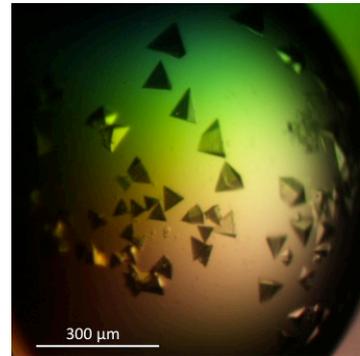
3.4 Data collection and processing of GluDH -NADPH-glutamate complex crystals

3.4.1 Data collection

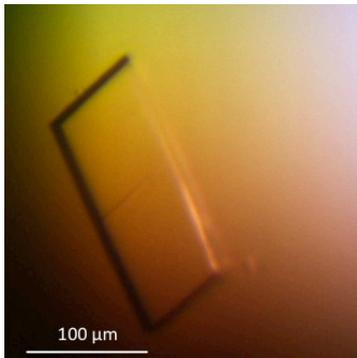
One crystal was harvested from the PACT suite, condition H1 robot screen



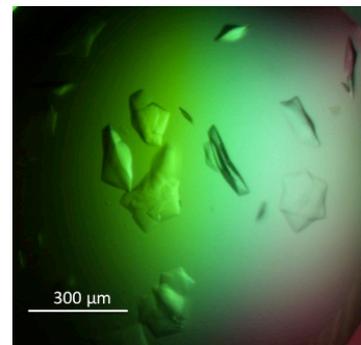
a) PEG D7-Robot Screen
 0.1M Tris-HCl pH=8.5
 25%(w/v) PEG3000



b) PEG D12-Robot Screen
 0.1M Tris-HCl pH=8.5
 15% PEG20000



c) PACT H1-Robot Screen
 0.2 M sodium fluoride
 0.1 M Bis-Tris Propane pH=7.5
 20% PEG3350



d) Classics D8-1 Robot Screen
 0.1M HEPES Sodium Salt pH=7.5
 1.4M Tri-Sodium Citrate

Figure3.8. Photographs of *M. smegmatis* GluDH crystals. **a** and **b** are the crystals in an apo state, **c** is the crystal in complex with NADPH and L-glutamate and **d** is the crystals of ternary complex of this protein with NADP⁺ and 2-oxoglutarate.

(figure3.8c) for data collection. The crystal was looped and transferred into a cryoprotectant containing the mother liquor and 25% ethylene glycol. This crystal diffracted to beyond 2.7Å resolution on the home source, so it was saved in the liquid nitrogen and taken to a beamline I04 of the Diamond Light Source Synchrotron (DLS), Oxford. 900 images were recorded at 100K from one crystal with a rotation angle of 0.2° and an exposure time of 0.4 s per image, using X-rays of wavelength 0.9795 Å (12658 eV) with the distance between the crystal and detector of 270mm. Diffraction data were collected and processed to a resolution of 1.78Å (Fig 3.9).

3.4.2 Data processing

The processing of the X-ray diffraction data involved several steps including, indexing (determination of the unit cell), measurement of the integrated intensities and merging and scaling the integrated data. Processing of the data from the crystal of GluDH-NADPH-glutamate was carried out at DLS using the xia2 system in the 3da mode (Winter, 2010). The data were indexed to space group C121 with the unit cell parameters $a=114\text{Å}$, $b=151.6\text{Å}$, $c=162.2\text{Å}$, $\alpha=\gamma=90^\circ$ and $\beta=101.4^\circ$. The asymmetric unit was suggested to comprise six subunits with a V_m of 2.34 according to the probabilities for the Matthews Coefficient determined by Mattprob (Kantardjieff & Rupp, 2003). Data-collection and processing statistics are shown in Table3.1.

3.5 Molecular replacement and model building of the *M. smegmatis* GluDH in complex with NADPH

As a number of GluDH structures are known, it was envisaged that the *M. smegmatis* GluDH structure could be determined using molecular replacement. The search model for the molecular replacement was obtained by comparing the amino acid sequence of *M. smegmatis* GluDH against the protein data bank (PDB), using the basic local alignment search tool (BLAST) (Altschul et al, 1997). The GluDH from *Clostridium symbiosum* was found to be the closest homologue of *M. smegmatis* GluDH with a sequence identity of 51%. Therefore, *C. symbiosum* GluDH (Accession number in PDB 1HRD) was used as a search model. As the predicted six chains in the asymmetric unit would be arranged as a complete hexamer, or two trimers with the crystallographic two-fold axis

Dataset	Overall	Low	High
High resolution limit Å	1.78	7.96	1.78
Low resolution limit Å	55.91	55.91	1.83
Completeness %	97.9	94.8	83.6
Multiplicity	3.6	3.6	2.8
I/sigma	12.1	29.4	2.0
Rmerge	0.060	0.031	0.473
Rmeas(I)	0.084	0.040	0.625
Rmeas (I+/-)	0.083	0.044	0.630
Rpim (I)	0.044	0.021	0.368
Rpim (I+/-)	0.056	0.031	0.413
Total observations	911386	10195	43840
Total unique	251813	2809	15805

Table 3.1. Data collection statistics for *M. smegmatis* GDH in complex with NADPH and L-glutamate.

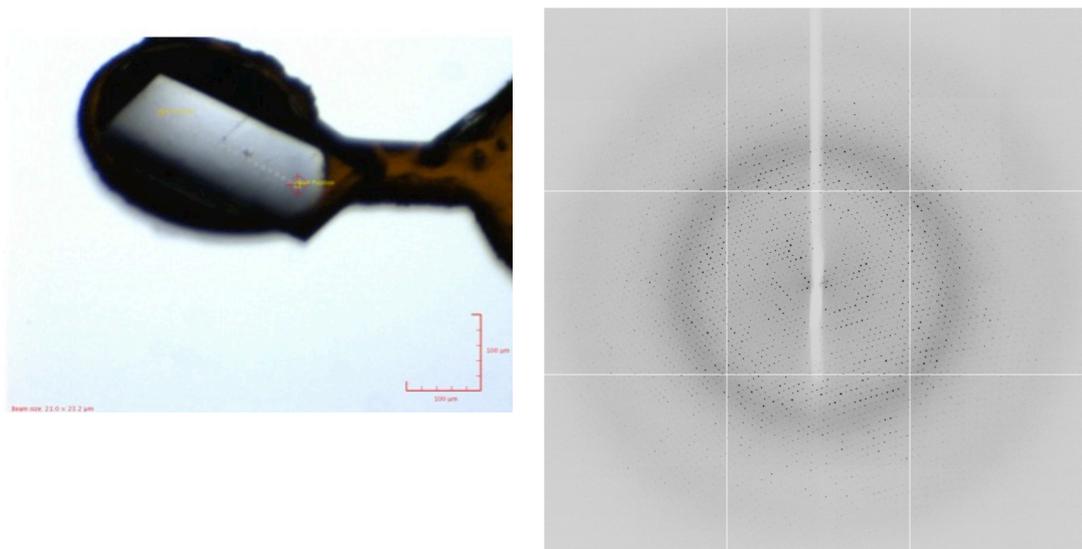


Figure 3.9. Diffraction picture for binary complex *M. smegmatis* GluDH with NADPH and L-glutamate.

building the hexameric molecule, a trimer of GluDH was used as the search model in program Phaser (McCoy et al, 2007).

The model of *C. symbiosum* GluDH (1HRD) was edited to remove water molecules and three subunits A, B and C were selected for an automated search in *Phaser* (McCoy et al, 2007). *Phaser* was run in the C121 space group, searching for two copies for the trimer in the asymmetric unit. *Phaser* found a solution with the Z score 10.3 for rotation and 10.4 for translation for the trimer one and Z-score of rotation function= 9.45 with translation function=21.7 for trimer two.

The initial model and electron density was reasonable for the N-terminal domain in each subunit but the electron density in the C-terminal domain (dinucleotide domain) was poor, possibly due to a different orientation of the domains compared to the search model. No interpretable density could be seen at the active site for the cofactor or the substrate for any of the six subunits. In an attempt to overcome this problem, rigid body refinement was carried out in Phenix with all 12 domains as separate entities. Although the density had improved, it was still very poor for some loops in the NADP⁺ domain. The Rfactor and Rfree for the resultant model from Phenix refinement were 0.41 and 0.47 respectively. This Phenix model was then subjected to xyz refinement in Refmac of the CCP4i suite (Collaborative, 1994), resulting in a structure with an Rfactor of 0.32 and Rfree of 0.36. Inspection of the electron density map over all the six subunits showed that the density was better for the D subunit in both domains, and that there was positive electron density in the difference map relating to the cofactor in all chains, but no density could be seen for the substrate. Despite presence of 250 mM glutamate being present in the crystallization conditions, the structure is thus believed to represent a binary complex with NADPH and is referred to as the *M. smegmatis* GluDH-NADPH binary complex for the rest of the thesis. Therefore, the decision was taken to use the chain D for rebuilding and thus the sequence of the search model (*C. symbiosum* GluDH) was substituted with the *M. smegmatis* GluDH sequence. Loops defined by poor electron density were deleted from the model and the first two residues at the N-terminal were removed as well due to the poor density around them. The NADPH cofactor was modeled in its electron density

and the rebuilt subunit D was superimposed onto the other subunits using the secondary structure-matching algorithm (SSM superpose) in COOT. Several cycles of model building and refinement were then conducted using COOT (Emsley et al, 2010) and REFMAC5 (Murshudov et al, 2011). In general, the features of the final electron density map were fine, with the exception of the regions at the residues S193-G205 and M385-E400 for all six chains A, B, C, D, E and F, where the density was missing. In addition, no density could be seen for residues G160-E161 in chains A and D and residue E161 in chain F. Therefore these loops were removed in the final structure. Although this structure is in the closed form, there is no electron density for L-glutamate and no substrate modeled in the resultant structure. The final model has an R-factor = 0.20 with R-free = 0.23 with 476 water molecules, and the NADPH bound in the catalytic pocket. The final model was restrained to standard bond lengths and angles and the root mean square deviations (r.m.s.d) of bond lengths=0.02Å and bond angles=2.0°. Validation of the final structure by the Molprobit server (Chen et al, 2010) and PROCHECK (Laskowski et al, 1993) showed that the quality of the final model was good and that most residues lie within the allowed regions of the Ramachandran plot, with the exception of 6 residues; Asp169 (in subunits B, C, D and F), Gly160 in chain E and Ser193 in subunit F, which were outliers, as shown in figure3.14.

3.6 Structure of 2-oxoglutarate-NADP-GluDH complex

3.6.1 Data collection and processing for low resolution data

The well-defined crystal of *M. smegmatis* GluDH in complex with NADP⁺ and 2-oxo-glutarate was harvested from the Classics suite condition D8 (figure 3.8d) for data collection. The crystal was looped and soaked in cryo-protectant solution containing the well solution (0.1M HEPES Sodium salt pH=7.5 and 1.4M Tri-Sodium Citrate) with 30% ethylene glycol before saving in liquid nitrogen. Data was collected at the Diamond Light Source, beamline I24. 1800 images were collected to a resolution 2.63Å with a rotation angle of 0.2° and an exposure time of 0.2 sec per image.

The processing of the X-ray diffraction data was carried out at DLS using the xia2 system in the 3da mode (Winter, 2010). Analysis of the data revealed that

Parameters	
Resolution (Å)	1.78Å
Number of reflections	239154
Protein molecule per asymmetric unit	6
Number of residues	449
Number of atoms	19682
Number of water molecules	479
Ramachandran favored (%)	97.26 %
Ramachandran outliers (%)	0.25 %
RMS bond length deviation (Å ²)	0.02 Å
RMS bond angle deviation (°)	2°
Average main chain B-factors (Å ²)	25 Å ²
Average side chain B-factors (Å ²)	31 Å ²
Average water B-factors (Å ²)	23 Å ²
R-factor (%)	0.1956
R _{free} (%)	0.2341
Molprobit score	1.54 92 nd percentile

Table 3.2. Shows the final refinement parameters for binary complex of *M. smegmatis* GluDH with NADPH.

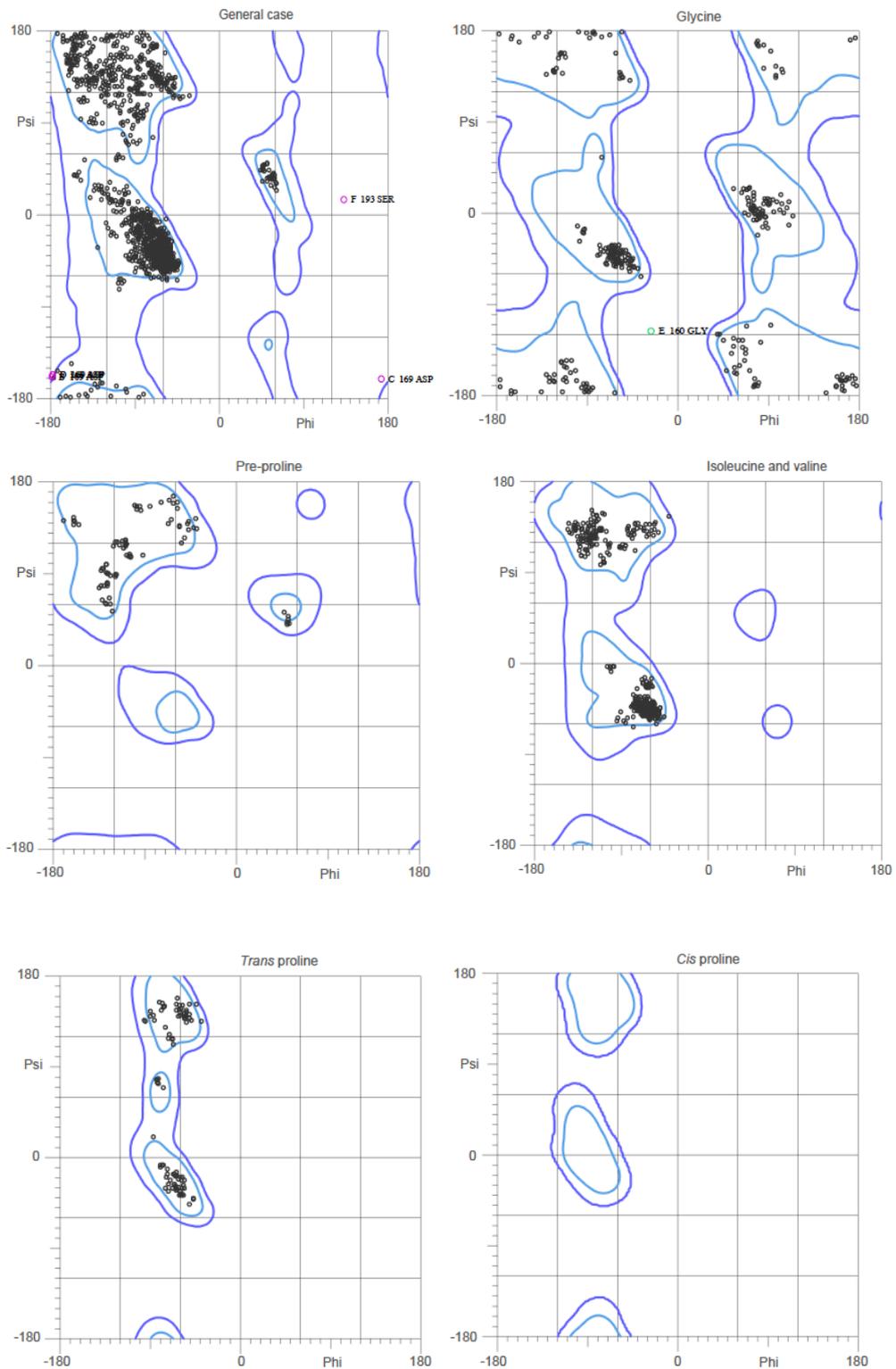
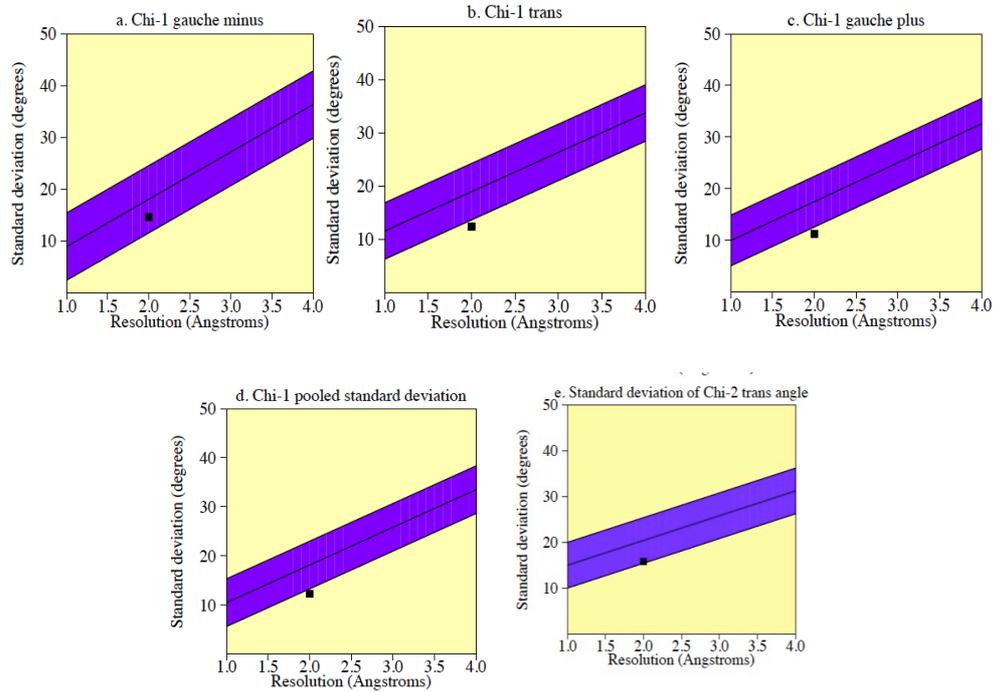


Figure 3.10. Ramachandran plot for the final structure of *M. smegmatis* GluDH + NADPH. The plot shows that 97.26% of the residues are within the favoured region and 0.25% are outliers. This figure was produced using Molprobit server (Chen et al, 2010).

Side chain parameters



Main chain parameters

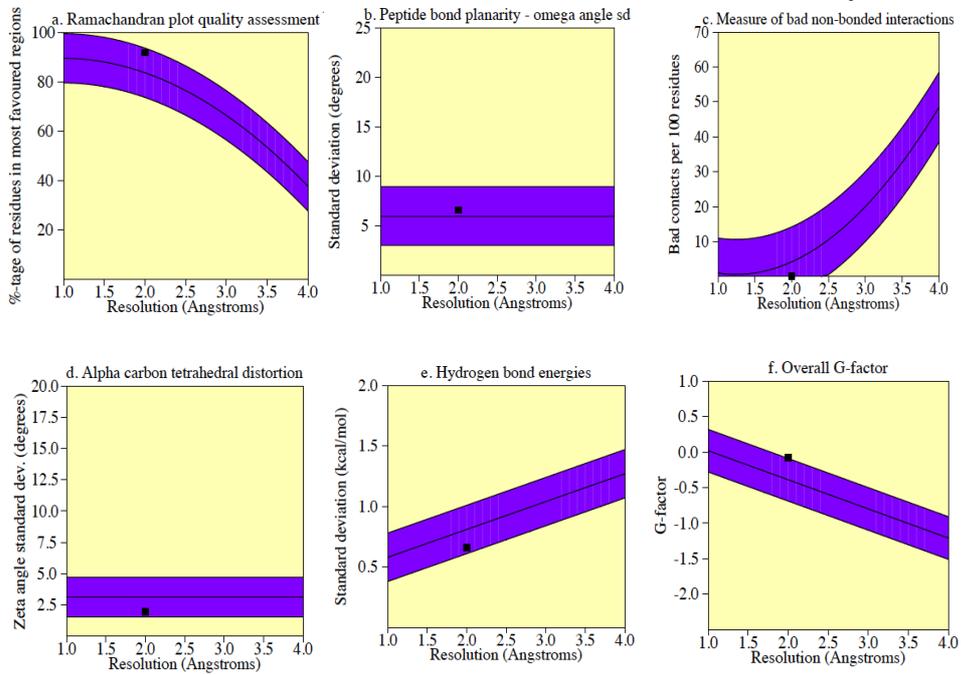


Figure 3.11. Main chain and side chain parameters for the final structure of *M. smegmatis* GDH + NADPH. This figure was produced using PROCHECK (Laskowski et al, 1993).

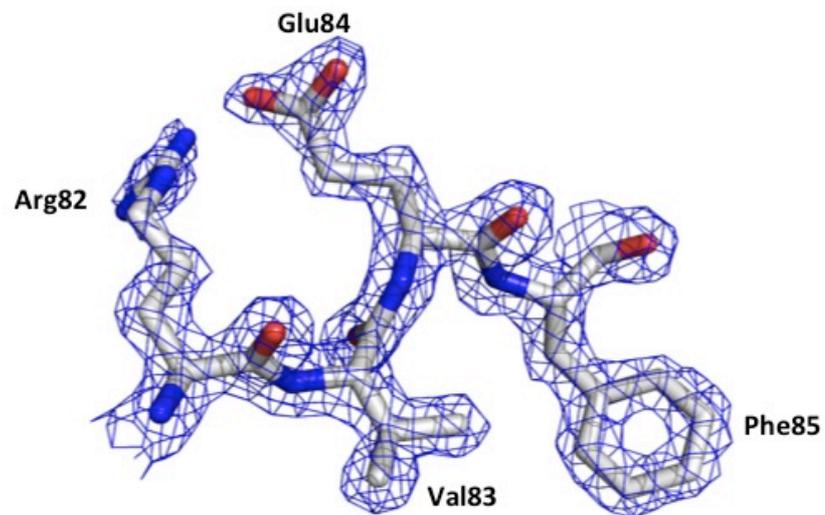
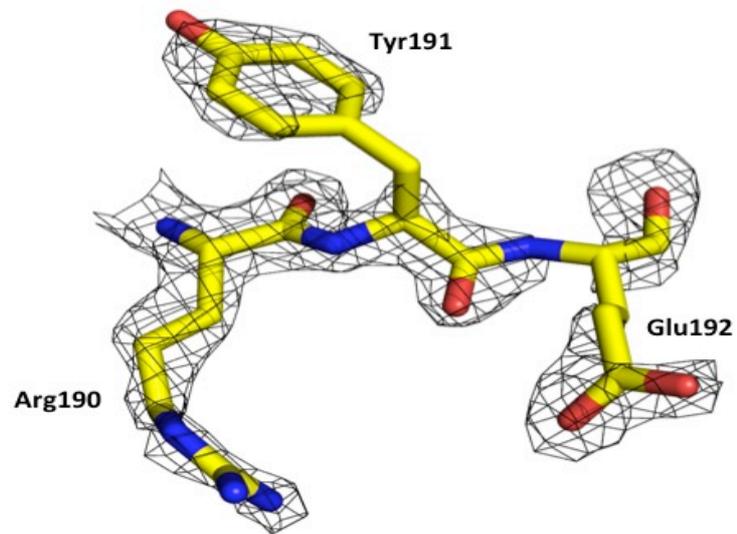
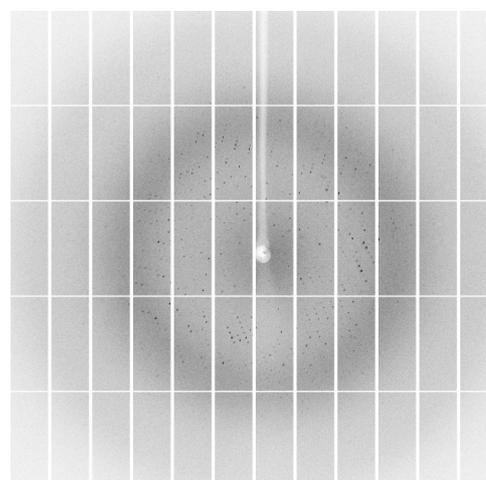
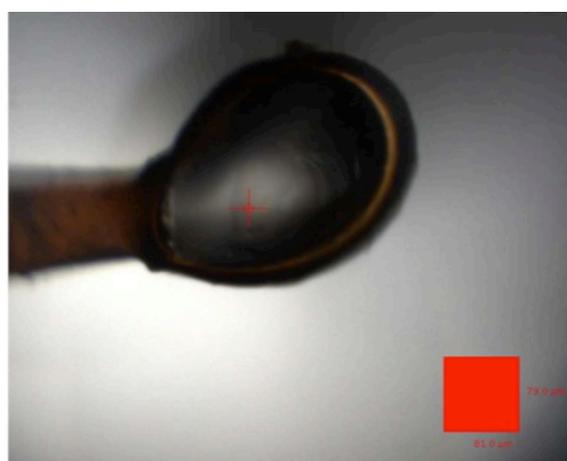


Figure 3.12. Electron density map at different regions in the final model of *M. smegmatis* GluDH + NADPH. Poor density can be seen at the beginning of the loop missing (yellow residues) and better density for the region of the N-terminal domain (light grey residues).

Dataset	
Space group	H32
Resolution range	11.77-2.63 Å
Unit cell dimension	
a=b	158.8Å
c	200.1Å
Completeness (%)	99.9
Multiplicity	20.0
I/sigma	13.7
Rmerge	0.216
Rmeas (I)	0.232
Rmeas (I+/-)	0.228
Rpim (I)	0.052
Rpim (I+/-)	0.071
Total observations	579884
Total unique	28935

Table 3.3. Data-collection and processing statistics for 2.63Å resolution data.



2.63Å dataset

Figure 3.13. Diffraction images for 2-oxoglutarate-NADP⁺ GluDH complex dataset at 2.63Å.

Dataset	
Space group	H32
Resolution range	4.92-1.64 Å
Unit cell dimension	
a=b	158.5Å
c	199.9Å
Completeness	99.9 %
Multiplicity	9.4
I/sigma	11.1
Rmerge	0.134
Rmeas (I)	0.153
Rmeas (I+/-)	0.151
Rpim (I)	0.049
Rpim (I+/-)	0.067
Total observations	1102694
Total unique	117275

Table 3.4. Data-collection and processing statistics for 1.6Å resolution data.

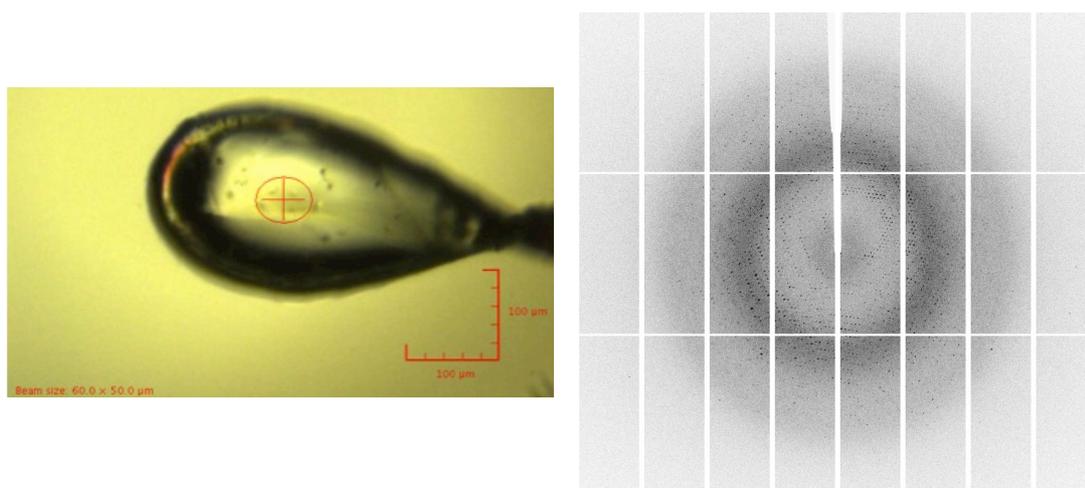


Figure 3.14. Diffraction images for 2-oxoglutarate-NADP GDH complex dataset at 1.64Å.

the crystal belonged to H32 space group with unit cell dimensions of $a = b = 158.8\text{\AA}$, $c = 200.1\text{\AA}$, $\alpha = \gamma = 90^\circ$ and $\beta = 120^\circ$. This crystal contains a dimer in the asymmetric unit with a V_m of 2.48, which is in the expected range of Matthews Coefficients (Kantardjieff & Rupp, 2003). Data-collection and processing statistics are shown in Table 4.2.

3.6.2 Molecular replacement and structure building for low resolution data

Chain A from the 1.78\AA model of the GluDH-glutamate-NADPH complex model was selected, using *Pdbset* program in CCP4i, as a suitable search model for molecular replacement. The X-ray diffraction dataset (2.63\AA) of GluDH-2-oxoglutarate-NADP⁺ complex was input, and Phaser was run to find two copies of the subunit in the asymmetric unit. Phaser solved the structure quickly and a subsequent round of refinement using Refmac5 (Murshudov et al, 2011) gave an Rfactor of 0.25. Inspection of the electron density map and the resulting model showed good density for the backbone structure with clear density of α -helices and with side chains visible. The map also contained interpretable density for the loops (residues 160 and 161, residues 193-209 and residues 385-400) that were missing in the previous structure. Unambiguous density for the cofactor (NADP⁺) and reasonable electron density for the substrate were also present in this map. Subsequently, the three missing loops were built in both subunits (A and B) using COOT (Emsley et al, 2010), and the cofactor and substrate (2-oxoglutarate) were modeled. After this a series of refining and rebuilding using REFMAC5 and COOT was completed. The final model of this ternary complex contains 894 residues and 65 water molecules with NADP⁺ and 2-oxo-glutarate bound in the catalytic pocket in each subunit. The R-factor was 0.15 with R-free=0.22 and the model was well defined in the electron density.

The final structure was validated by Molprobity server (Chen et al, 2010) and PROCHECK (Laskowski et al, 1993) which showed that the quality of the model was good and with most of the amino acid residues within the allowed regions of the Ramachandran plot, with the exception of Val208 in chain A and B. The

Parameters	
Resolution (Å)	2.63Å
Number of reflections	27462
Protein molecule per asymmetric unit	2
Number of residues	447
Number of atoms	6980
Number of water molecules	65
Ramachandran favoured (%)	97.19%
Ramachandran outliers (%)	0.22%
RMS bond length deviation (Å)	0.014Å
RMS bond angle deviation (°)	1.7°
Average main chain B-factors (Å ²)	33Å ²
Average side chain B-factors (Å ²)	37Å ²
Average water B-factors (Å ²)	26Å ²
R-factor (%)	0.15
R _{free} (%)	0.22
Molprobit score	1.69 99 th percentile

Table 3.5. Shows the final refinement parameters for ternary complex of *M. smegmatis* GluDH with NADP⁺ and 2-oxo-glutarate, 2.63 Å resolution.

root mean square deviations (r.m.s.d) of the bond lengths=0.014Å and bond angles=1.7°.

3.6.3 Data collection and processing for higher resolution data

As the data for the crystal obtained from the robot trial were only to medium resolution (2.63Å) attempts were made to improve the crystals, and hence the resolution of the data, by optimization of the crystallization conditions. Therefore, the Classics suite D8 condition (0.1M HEPES Sodium Salt pH=7.5 and 1.4M Tri-Sodium Citrate) was optimized by changing the concentration of Tri-Sodium Citrate from 1.0M to 1.8M in steps of 0.2M.

1 µl protein sample and 1 µl crystallization solution were used in a hanging drop trial and crystals appeared after two weeks of incubation at 17 °C, in the original condition (1.4M Tri-Sodium Citrate and 0.1M HEPES Sodium Salt pH=7.5). Crystals were selected for data collection according to their size and overall definition and were looped and soaked in a cryo-protectant solution consisting of the crystallization solution with 30% ethylene glycol. Data were collected at beamline I04-1 at Diamond Light Source. 1200 images were collected to a resolution 1.64Å with a rotation angle of 0.15° and an exposure time of 0.1 sec per image, using X-rays of wavelength 0.9200 Å (13476.5 eV).

Processing of the X-ray diffraction data was carried out using the Xia2 system by the 2da mode (Winter, 2010). Analysis of these data revealed that the crystal belonged to the same space group as before (H32 with unit cell dimensions of $a = b = 158.5\text{Å}$, $c = 199.9\text{Å}$, $\alpha = \gamma = 90^\circ$ and $\beta = 120^\circ$). The crystal contained a dimer in the asymmetric unit with a V_m of 2.47.

3.6.4 molecular replacement structure building for high-resolution data

The previous 2.63Å model of this ternary complex (GluDH-2-oxoglutarate-NADP⁺) was refined against the higher resolution data set (1.64Å) using Refmac5, resulting in an initial structure with R-factor and R_{free} , 0.22 and 0.24 respectively. Rebuilding this model to fit the electron density was performed in COOT, and multiple cycles of structure refinement and rebuilding were conducted through Refmac5. An examination of the electron density map over the two subunits of the dimer showed that the polypeptide chain was well defined in both chains. The electron density of the substrate (2-oxoglutarate)

was much better than that of the 2.63Å structure (figure3.15). Several repeated cycles were carried out for structure building, refinement and evaluation using COOT and Refmac5. The final structure of this ternary complex comprises 894 residues, 943 water molecules and four molecules of ethylene glycol. The R-factor was 0.14 with R-free=0.17 and the model well defined in the electron density.

The final model has been restrained to standard bond lengths and angles and the root mean square deviations (r.m.s.d) of bond lengths=0.012Å and bond angles=1.52°. Validation of the final structure by Molprobit server (Chen et al, 2010) and PROCHECK (Laskowski et al, 1993) revealed that the final model was of good quality and most residues lie within the allowed regions of the Ramachandran plot with the exception of Val208 and Asp169 residues in both subunits that were in outlier regions, though these residues were fitted well in the density map.

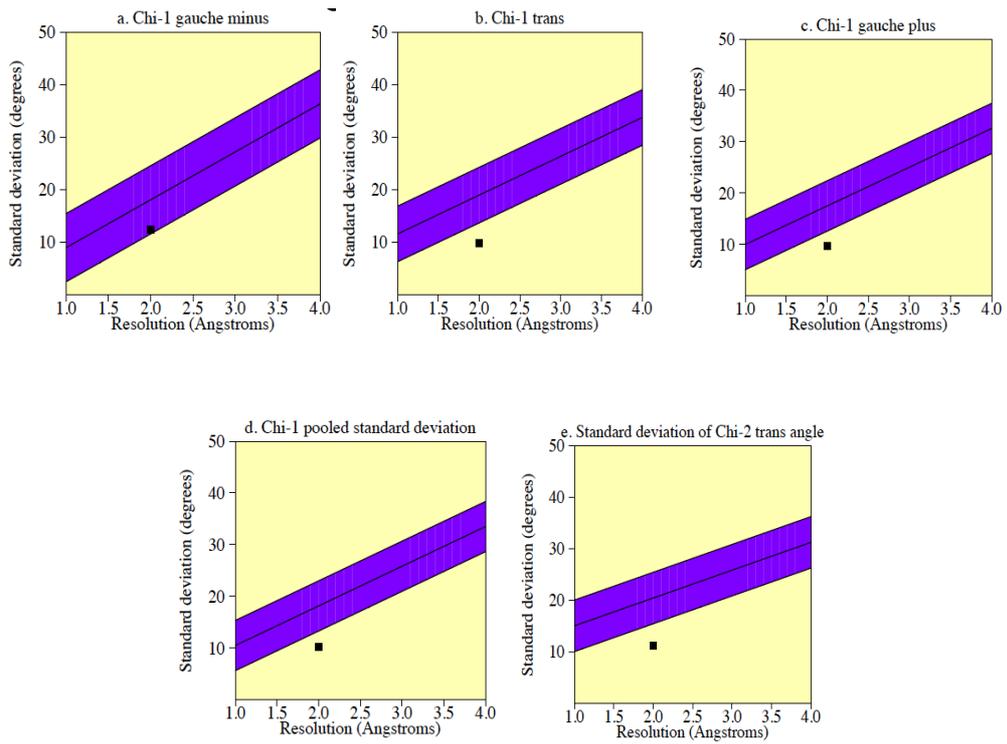
3.6.4.1 Refinement of the substrate

Inspection of the electron density at the active site of the two subunits showed clear density for the substrate and cofactor. By examining the substrate density in chain A and chain B it was found that there were two density blobs at the C α position instead of one density peak for the carbonyl group. After modeling the 2-oxoglutarate in its density and refining the model using refmac5, a positive density could be seen near to the C α . This might be due to the presence of the 2-oxoglutarate in two conformations. Moving the carbonyl group into the positive density in COOT and refining the structure again using Refmac5 resulted in a positive density blob in the other position. Two molecules of 2-oxoglutarate were modeled with occupancy 0.5 in the electron density in two subunits and the model was refined using refmac5. The resultant structure has a positive density around the C2. Additionally, analysis of the B-factors for the C1, C2 and C3 with O5 for the two 2-oxoglutarates revealed that the average of the B-factors of the carbon atoms (C1, C2 and C3) are 18.5 Å², which is higher than that for the O5 (10 Å²). From the refinement of these models above, it can be concluded that the substrate is not one molecule of 2-oxoglutarate and does not look likely to be 2 molecules of 2-oxoglutarate. In addition, the electron

Parameters	
Resolution	1.64Å
Number of reflections	111037
Protein molecule per asymmetric unit	2
Number of residues	447
Number of atoms	7805
Number of water molecules	943
Ramachandran favoured (%)	98.45%
Ramachandran outliers (%)	0.45%
RMS bond length deviation (Å)	0.012 Å
RMS bond angle deviation (°)	1.52°
Average main chain B-factors (Å ²)	11Å ²
Average side chain B-factors (Å ²)	19Å ²
Average water B-factors (Å ²)	26 Å ²
R-factor (%)	0.14
R _{free} (%)	0.17
Molprobit score	1.15 99 th percentile

Table 3.6. Shows the final refinement parameters for ternary complex of *M. smegmatis* GluDH with NADP⁺ and 2-oxoglutarate, 1.64Å resolution.

Side chain parameters



Main chain parameters

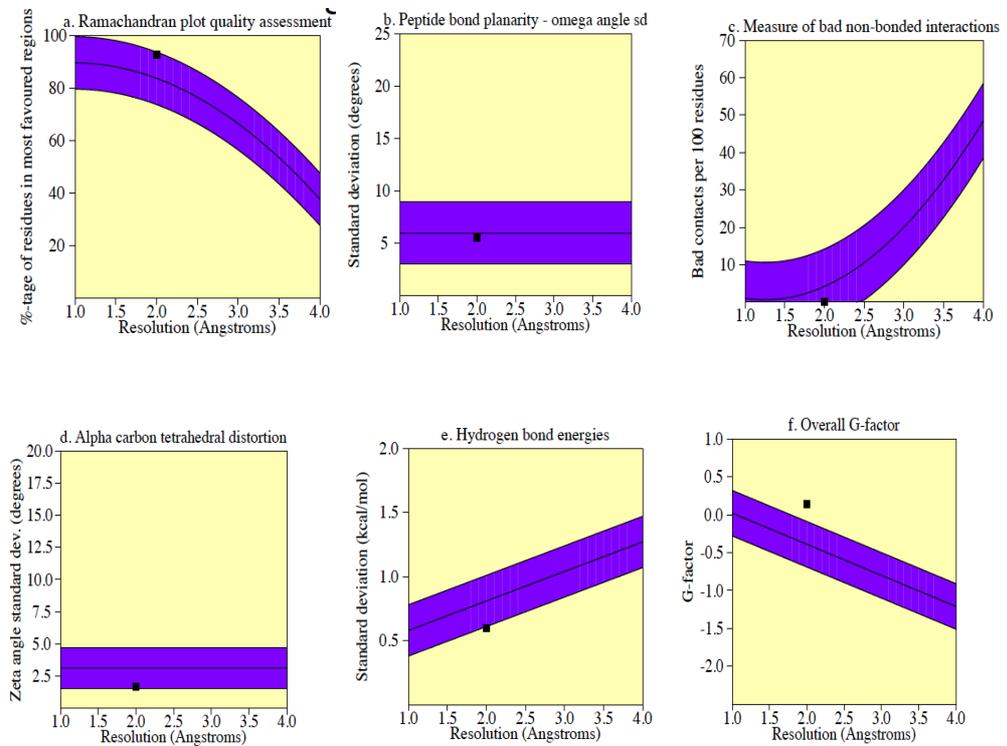


Figure 3.15 Side-chain and main-chain parameters for the final model of GluDH-NADP-2-oxoglutarate from *M. smegmatis*. The figure produced by PROCHECK (Laskowski et al, 1993).

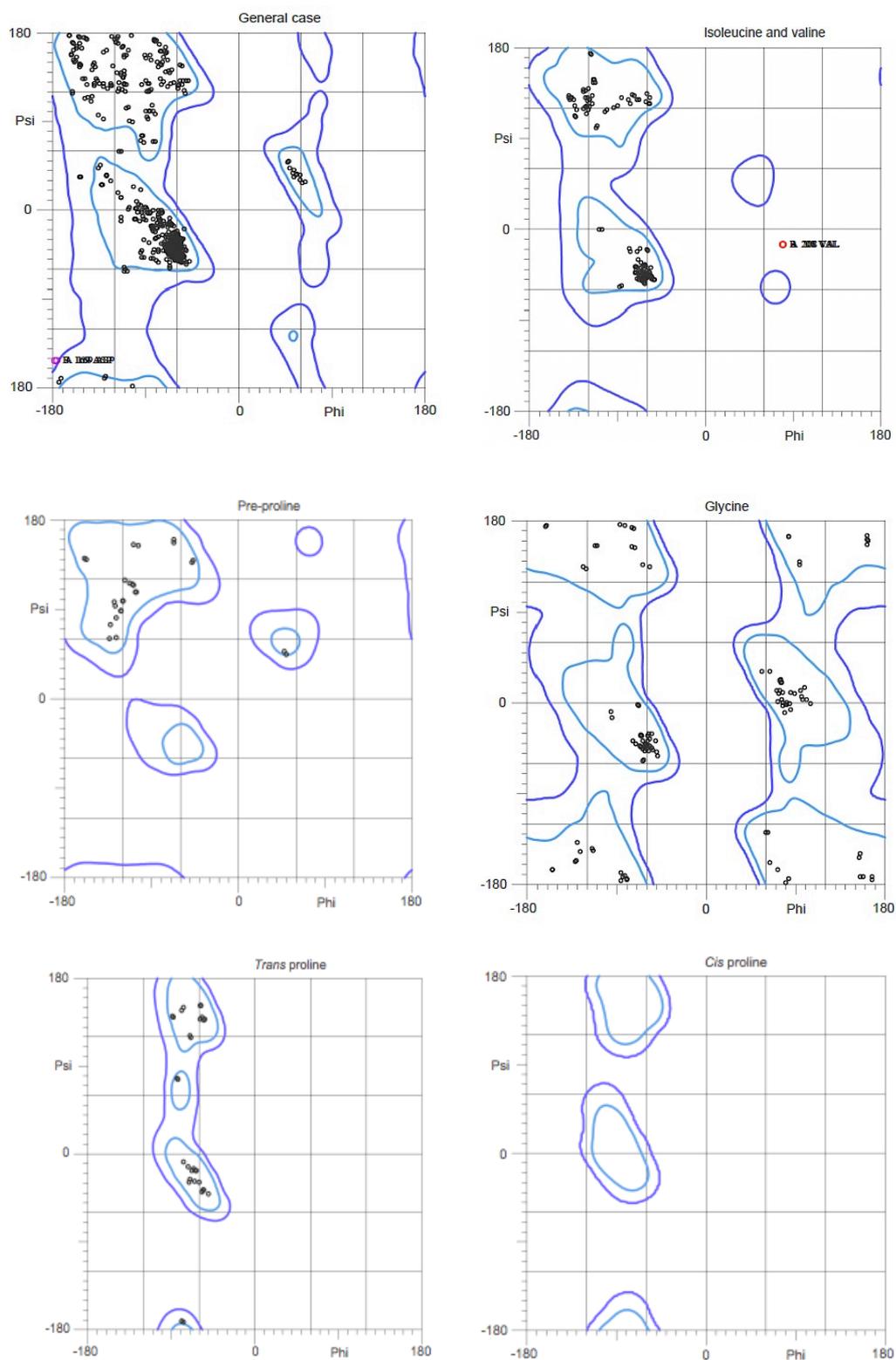


Figure 3.16. Ramachandran plot for the final model of ternary complex model of *M. smegmatis* GluDH + NADP⁺ + 2-oxoglutarate. The plot displays 98.45% of the residues are within the favoured region and 0.45 % are outliers. This figure was produced using Molprobit server (Chen et al, 2010).

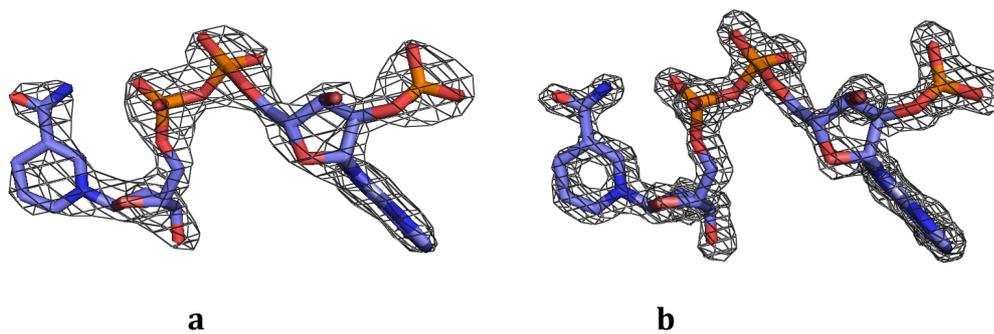


Figure 3.17 shows the electron density maps around the cofactor NADP⁺ at 2.63Å resolution (**a**) and at 1.64Å resolution (**b**). Maps are shown at 2-sigma counter level.

density around the α -carbon (C2) looks sp^3 (tetrahedral) rather than sp^2 (trigonal). 2-oxoglutarate is a keto acid, and thus there exists a possibility that this compound is present in the crystal as a gem-diol, rather than the keto acid. The ketone-gem diol equilibrium in aqueous solution is modulated by groups attached to the carbon and can range from $K=10^{-3}$ for acetone, to $K=1.2 \times 10^6$ for hexafluoro acetone (Clayden et al). As the $C\alpha$ in 2-oxoglutarate is attached to an electron withdrawing carboxylate group, an appreciable fraction of the compound would most likely be in the gem-diol form in aqueous solution. Therefore, the substrate was modelled as a gem-diol 2-oxoglutarate within the electron density and refining the structure using Refmac5 showed no difference peaks about the $C\alpha$, which means that the gem diol form of 2-oxoglutarate fits the density perfectly. In addition, analysis of the B factors for O5 and O6 with C1, C2 and C3 for the diol revealed that the B factors for the oxygen atoms (O5 and O6) are same as the average of B factors for carbon atoms (C1, C2 and C3).

3.7 Glutamate dehydrogenase from *Clostridium symbiosum*

Obtaining crystal structures of the complexes of substrates, cofactors and products of a particular enzyme is fraught with difficulty, as binding of the various ligands often seems to preclude crystallization. In addition, crystal contacts or the crystallization buffer often keep the enzyme in a conformation that precludes substrate or cofactor binding. It was thus decided early on in the project to also study glutamate dehydrogenases from other bacterial species, to optimize the chance of obtaining relevant complex structures. To this end, the GluDH from *Clostridium symbiosum* was used as another source of the enzyme. Structures of the apo enzyme (1.96Å) and the GluDH- glutamate binary complex (1.9Å) have been reported (Baker et al, 1992a; Stillman et al, 1993) and deposited in the protein data bank, accession numbers 1HRD and 1BGV for the apo enzyme and the binary complex respectively. In addition a GluDH-NAD⁺ binary complex has been described to 2.5Å (Baker et al, 1992a; Stillman et al, 1993). As the crystals of this enzyme diffracted well on first generation synchrotron sources, using X-ray film to collect the data, it was envisaged that higher resolution structures could now be obtained. In addition, the wealth of

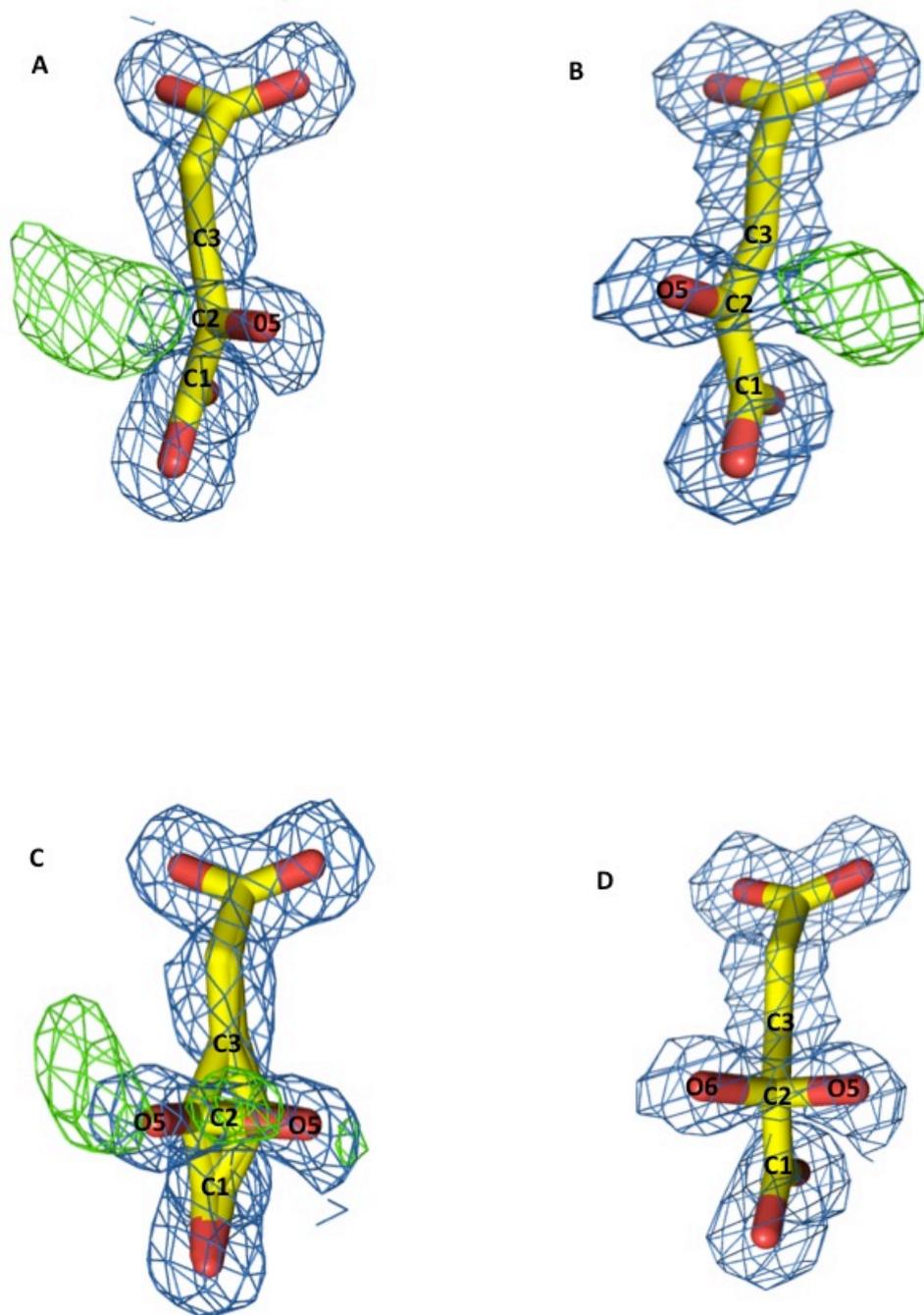


Figure 3.18. Electron density for the substrate of the ternary complex of *M. smegmatis* GDH + NADP⁺ + 2-oxoglutarate. a and b the different density of the two conformations of carbonyl group, c the electron density for the two molecules of 2-oxoglutarate and d the electron density for the gem-diol. In each case the map (counteracted at 2-sigma) is shown in blue, with the positive difference map (counteracted at 3-sigma) in green.

B factor with occupancy=0.5 for the 2-oxoglutarate		
	Molecule 1	Molecule 2
O4	8	8
C5	12	11
O3	9	9
C4	17	16
C3	16	19
C2	22	17
O5	9	10
C1	18	17
O1	13	12
O2	12	12

(a)

	B factor with occupancy=1.0 for the gem-diol
O4	10
C5	12
O3	10
C4	16
C3	19
C2	26
O5	22
O6	21
C1	18
O1	13
O2	13

(b)

Table 3.7. Shows B factor values for the 2-oxoglutarate (table a) and gem-diol (table b).

crystallization screens now available would perhaps allow ternary complexes of the enzyme to be crystallized.

3.7.1 Purification of glutamate dehydrogenase of *Clostridium symbiosum*

Glutamate dehydrogenase protein was expressed by Fiona Rodgers in Sheffield according to the previous method (Teller et al, 1992). About 3 g of the cell paste was resuspended in the lysis buffer, 50mM TRIS pH=8.0. Insoluble material was removed by centrifugation at 70,000 g for 15 minutes and the cell free extract (total protein 125mg) was loaded to 5ml DEAE FF cartridge and eluted with 50ml NaCl gradient from 0 to 0.5M. 2ml fractions were collected and analyzed by SDS-PAGE (4-12% Bis-Tris NOVAX gel). Fractions 16-21 were combined; a total protein of 54mg. 4M-ammonium sulphate solution was added to the sample to give a final concentration of 1.5M and the sample was applied on a 5ml Phe-HP cartridge. The protein was eluted by 50ml of reverse gradient of ammonium sulphate concentration from 1.5M to 0M in buffer A. 2ml fractions were collected and analysed by SDS-PAGE. Fractions 22-24 were combined (total protein 16mg) and the volume of the sample was reduced to 2ml using a VivaSpin concentrator with a molecular weight cutoff of 30kD and applied on a gel filtration column equilibrated with bufferA+0.5M NaCl. Gel filtration was performed at flow rate 1.5ml/min with the same buffer. The GluDH eluted at peak of 60ml, which corresponds to an apparent MW of 290kDa (a hexamer). 2ml fractions were collected after void volume and analysed on SDS-PAGE. The progress of purification was analyzed by SDS-PAGE (4-12% Bis-Tris NOVAX gel). For crystallization the buffer in the sample was exchanged with 10mM Na⁺ Phosphate pH 8.0 using a VivaSpin concentrator with the diafiltration cup and the protein was concentrated to 11 mg/ml.

3.7.2 Crystallization of glutamate dehydrogenase from *C. symbiosum*

Co-crystallization of GluDH protein was carried out using the hanging-drop vapour-diffusion technique (see section 2.3.2 for details). In this experiment 1µl of the protein sample (11mg/ml) with L-glutamate (250mM) was mixed with an equal volume of 25%-50% ammonium sulphate in 100mM sodium phosphate buffer pH=7.0, and then equilibrated against the same precipitant at

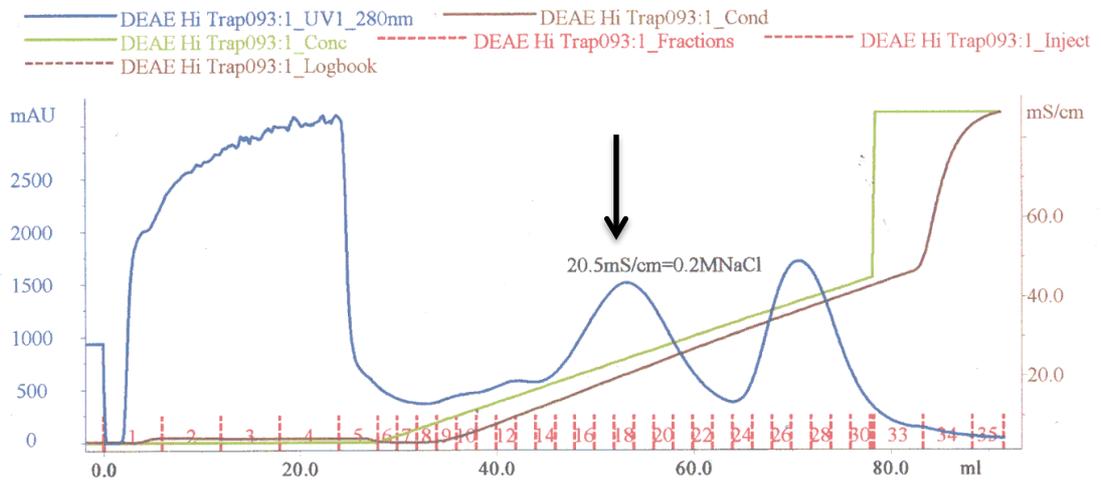
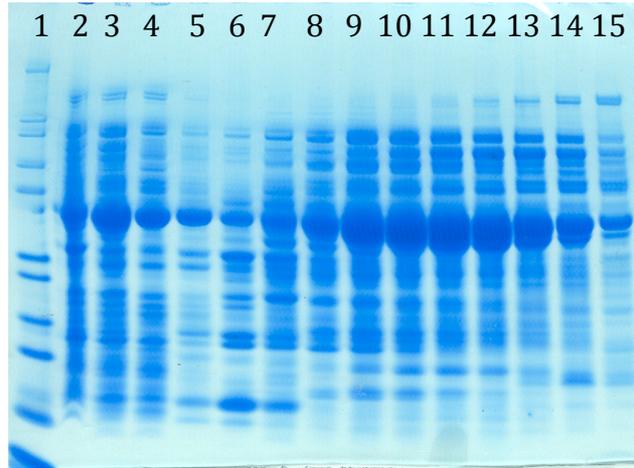


Figure 3.19. SDS-PAGE analysis and chromatogram trace of *C. symbiosum* GluDH (MW=49.296 kDa) DEAE purification. The SDS gel shows fractions from the DEAE purification stage of *C. symbiosum* GluDH. Lanes are ordered as following: 1= marker 12, 2= cell debris, 3= cell free extract, 4= fraction4, 5= fraction10, 6=fraction12, 7=fraction14 and 8-15=fractions 16-24. The trace shows loading and elution with 2 ml fractions collected starting at the beginning of the gradient at 50 ml. Arrow indicates peak containing GluDH protein.

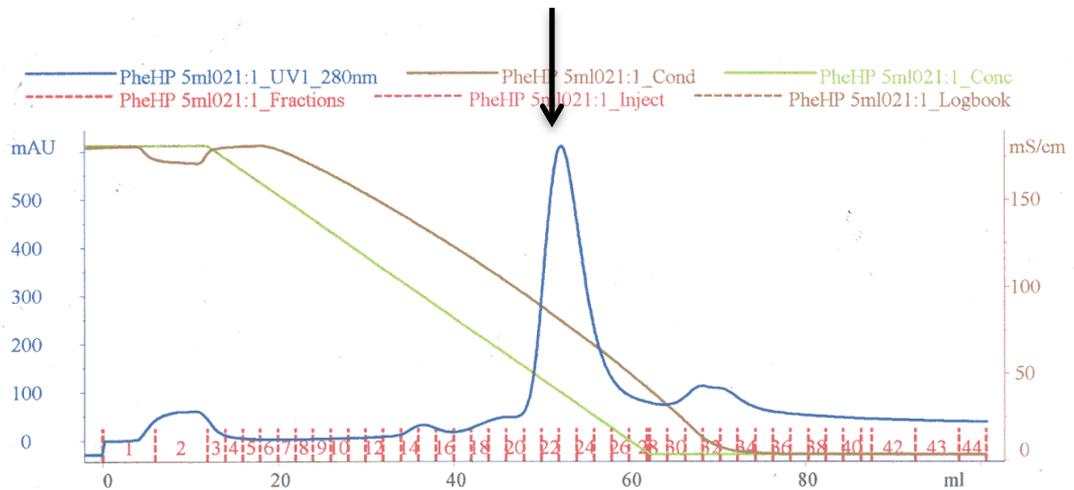
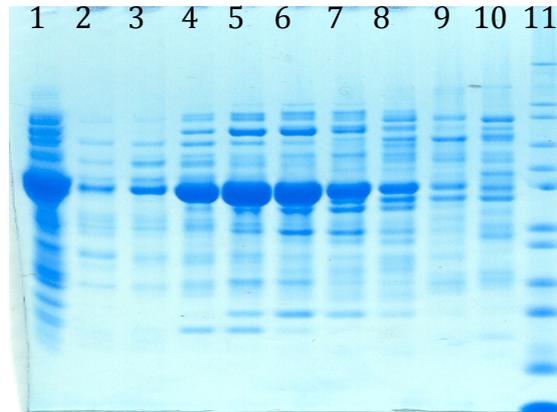


Figure 3.20. SDS-PAGE analysis and chromatogram trace of *C. symbiosum* GluDH (MW=49.296 kDa) Phe-HP purification. The SDS gel shows fractions from the Phe-HP purification stage of *C. symbiosum* GDH. Lanes are ordered as following: 1= sample applied on Phe-HP column, 2-8=fractions 20-26, 9=fraction28, 10=fraction31 and 11= marker 12. The trace shows loading and elution with 2 ml fractions collected starting at the beginning of the gradient at 50 ml. Arrow indicates peak containing GDH protein.

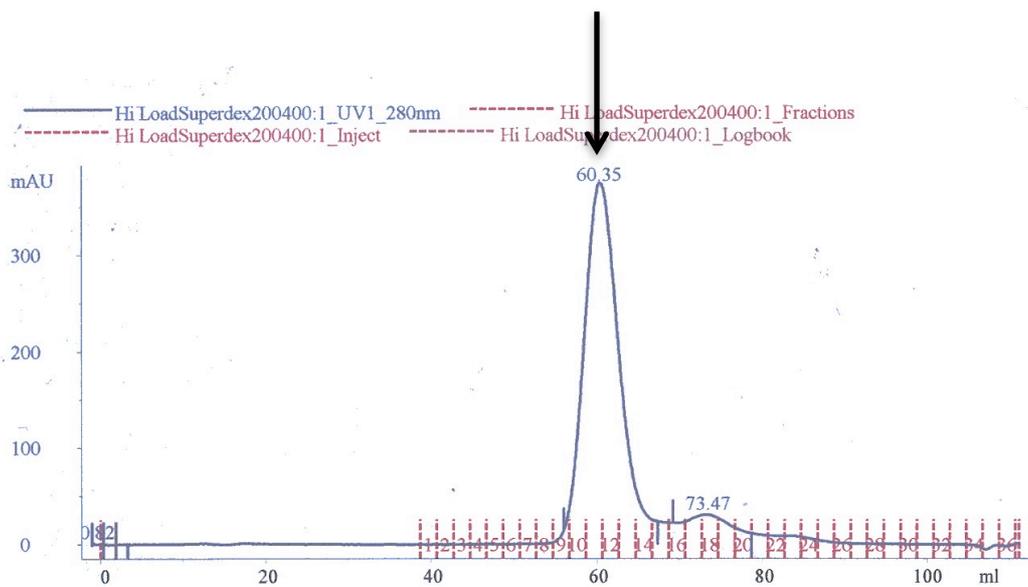
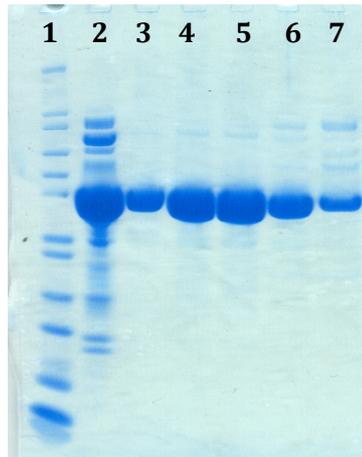


Figure 3.21. SDS-PAGE analysis and chromatogram trace of *C. symbiosum* GluDH (MW=49.296 kDa) gel filtration purification. The SDS gel shows fractions from the gel filtration purification stage of *C. symbiosum* GDH. Lanes are ordered as following: 1= marker 12, 2= sample applied on gel filtration column and 3-7 fractions 10-14. The trace showing elution with 2 ml fractions collected throughout. Arrow indicates peak containing GDH protein.

17°C. The best crystals grew in 45% ammonium sulphate (in 100mM sodium phosphate) after two weeks. The same procedure was used to grow apo crystals, and then these were used to soak NAD⁺ into the crystals to determine the mode of binding of the cofactor to *C. symbiosum* GluDH.

3.7.3 Data collection and processing for crystals of *C. symbiosum* GluDH with L-glutamate

Crystals were selected for data collection according to their size and overall definition. Diffraction data from GluDH-glutamate binary complex crystal of the enzyme was collected at the Diamond Light Source, beamline I02. 1000 images were collected with a rotation angle of 0.1° and an exposure time of 0.1 sec per image using X-rays of wavelength 0.97949 Å. Processing the X-ray diffraction datasets was performed by Xia2 system using the 3da mode at DLS. The data were processed to a resolution of 1.24 Å in space group H32, with unit cell parameters $a=b=158.7\text{Å}$, $c=101.4\text{Å}$, $\alpha=\beta=90^\circ$ and $\gamma=120^\circ$. This crystal has a monomer in the asymmetric unit with a V_m of 2.5, which is in the range of Matthews Coefficients (Kantardjieff & Rupp, 2003). Data-collection and processing statistics are shown in Table 3.5.

3.7.4 Molecular replacement and model building for the binary complex of *C. symbiosum* GluDH with glutamate

The structure of *C. symbiosum* GluDH with L-glutamate (accession number 1BGV) (Stillman et al, 1993) was used as a search model for the molecular replacement. Diffraction data (1.24 Å) were input, and Phaser (McCoy et al, 2007) was run to search for the one copy subunit in the asymmetric unit. Phaser found a solution with Z-score of 24.87 and 56.23 for the rotation and translation respectively, and Rfactor=0.27 with Rfree=0.25 following rigid body refinement in REFMAC5 (Murshudov et al, 2011). Multiple cycles of model building and refinement were carried out using COOT (Emsley et al, 2010) and REFMAC5. Clear density was present in the active site that could be attributed to L-glutamate. The final model contains 524 water molecules with L-glutamate bound in the active site with an Rfactor and the Rfree of 0.13 and 0.15, respectively. The root mean square deviations (r.m.s.d) of bond lengths=0.0068Å and bond angles=1.26°. Validation of the final structure using

	Overall	Low	High
High resolution limit	1.24 Å	5.55	1.24
Low resolution limit	31.09 Å	31.09	1.27
Completeness	99.9 %	99.5	100.0
Multiplicity	5.3	5.8	5.3
I/sigma	10.3	25.8	2.5
Rmerge	0.068	0.046	0.540
Rmeas (I)	0.085	0.054	0.677
Rmeas (I+/-)	0.084	0.055	0.679
Rpim (I)	0.036	0.022	0.293
Rpim(I+/-)	0.048	0.031	0.406
dF/F	0.067		
dI/s(dI)	0.789		
Total observations	728472	9418	53544
Total unique	137055	1623	10109

Table 3.8. Data-collection and processing statistics for *C. symbiosum* GluDH in complex with L-glutamate.

Parameters	
Resolution (Å)	1.24Å
Number of reflections	130130
Protein molecule per asymmetric unit	1
Number of residues	449
Number of atoms	3477
Number of water molecules	524
Ramachandran favoured (%)	97.12%
Ramachandran outliers (%)	0.22%
RMS bond length deviation (Å)	0.007Å
RMS bond angle deviation (°)	1.26°
Average main chain B-factors (Å ²)	15Å ²
Average side chain B-factors (Å ²)	19Å ²
Average water B-factors (Å ²)	34 Å ²
R-factor (%)	0.13
R _{free} (%)	0.15
Molprobrity score	0.96 99 th per centile

Table 3.9. Shows the final refinement parameters for binary complex of *C. symbiosum* GluDH with L-glutamate.

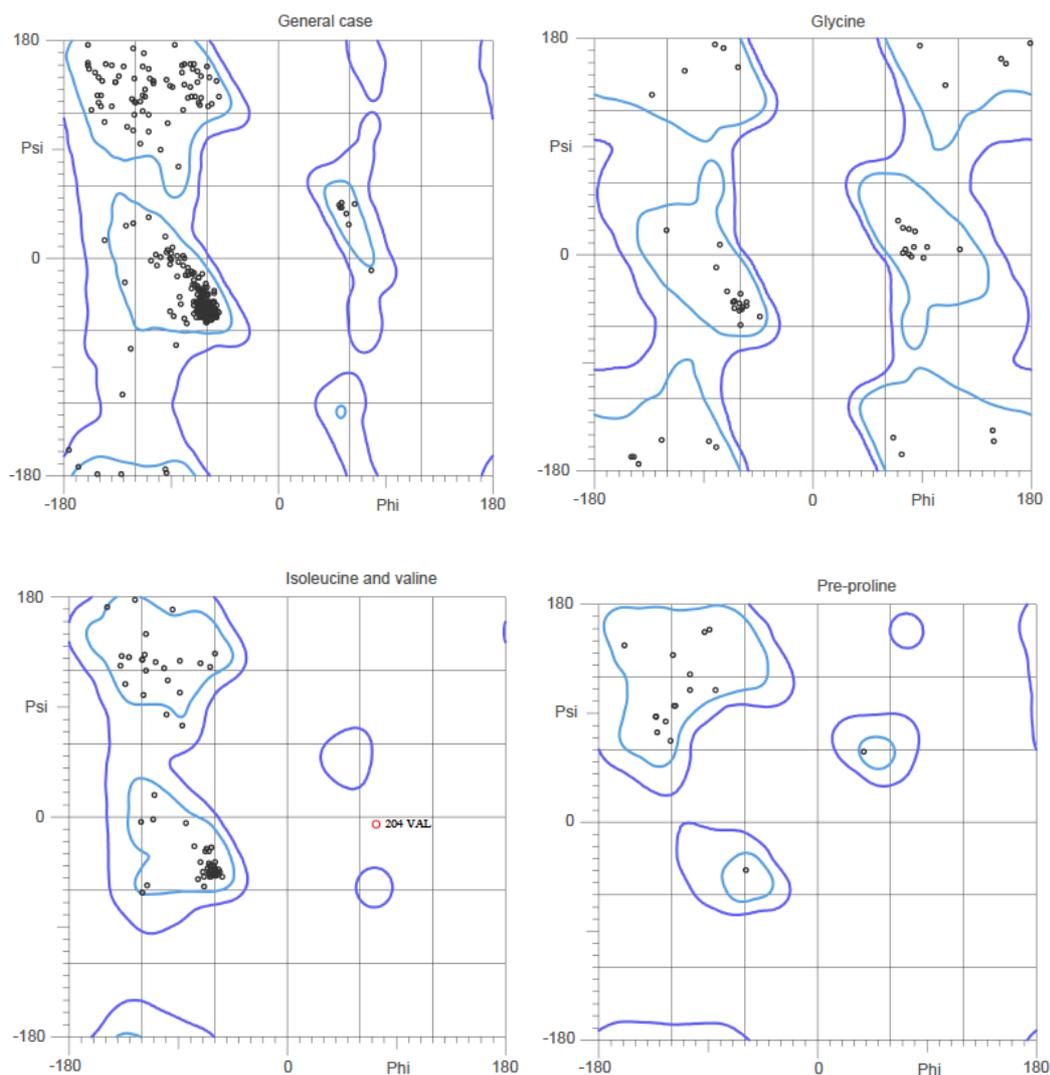


Figure 3.22. Ramachandran plot for *C. symbiosum* GluDH in complex with L-glutamate. The plot shows 97.12% of the residues are within the favoured region and only Val204 is Ramachandran outlier. This revised Ramachandran plot was generated using Molprobit server (Chen et al, 2010).

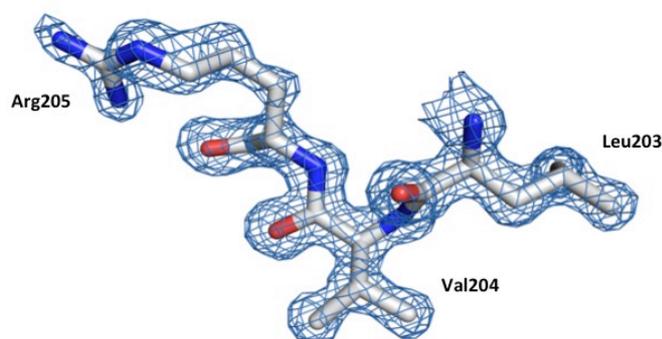


Figure 3.23. The electron density map around Val204. Map is shown at 2-sigma counter level.

the Molprobit server (Chen et al, 2010) and PROCHECK (Laskowski et al, 1993) showed that this model is of a good quality with most of the amino acid residues lying within the allowed regions of the Ramachandran plot with the exception of Val204 (figure3.22), even though this residue has a good electron density (figure3.23).

3.7.5 Data collection and processing for *C. symbiosum* GluDH soaked in NAD⁺

Data collection and processing for *C. symbiosum* GluDH with NAD⁺ was conducted using the same method as above, at the Diamond Light Source, beamline I04-1. Data from this crystal was processed to a resolution 1.80 Å and was indexed to a space group C2 with the unit cell parameters a=136.2 Å, b=150.2Å, c=91.9Å, $\alpha = \gamma = 90^\circ$ and $\beta = 129.640^\circ$. The crystal contains a trimer in the asymmetric unit with a V_m of 2.46. Data-collection and processing statistics are shown in Table (3.6).

3.7.6 Molecular replacement and model building for binary complex of *C. symbiosum* GluDH with NAD⁺

The space group C2 for these crystals was the same as that for the apo structure of *C. symbiosum* GluDH (1HRD), with similar cell dimensions and thus this model was used as the search in molecular replacement using Phaser (McCoy et al, 2007) in CCP4i. The three subunits A, B and C were selected for an automated search in Phaser. The Rfactor=0.29 with R_{free}=0.31 following refinement in REFMAC5 (Murshudov et al, 2011). The fit of the primary model to the electron density was generally good for subunits A and B, but domain II in subunit C was not in the density map. Rebuilding this structure was done using COOT (Emsley et al, 2010) and the whole domain II of chain C was moved to fit the electron density. Several cycles of model building and refinement were then conducted using COOT (Emsley et al, 2010) and REFMAC5 (Murshudov et al, 2011). Inspection of the electron density within the active site for all subunits showed that there was clear positive difference density linked with the nucleotide-binding site of the 3 subunits. NAD⁺ was modeled into this electron density in the active site and a number of refinement cycles were carried out using REFMAC5 (Murshudov et al, 2011). The final model has

R-factor = 0.17 with $R_{\text{free}} = 0.22$ and 500 water molecules, in addition to NAD^+ bound in the catalytic pocket. The root mean square deviations (r.m.s.d) of bond lengths=0.017Å and bond angles=1.8°. The final structure was validated by Molprobit server (Chen et al, 2010) and PROCHECK (Laskowski et al, 1993) which display the quality of the model was good with most of the amino acid residues lying within the allowed regions of Ramachandran plot, with the exception of Val204 in A, B and C subunits and Asp165 in B and C subunits.

	Overall	Low	High
High resolution limit	1.80 Å	8.05	1.80
Low resolution limit	52.45 Å	52.45	1.85
Completeness	98.4%	99.4	98.6
Multiplicity	3.9	4.1	3.9
I/sigma	10.3	39.1	1.9
Rmerge	0.084	0.024	0.594
Rmeas (I)	0.115	0.031	0.817
Rmeas(I+/-)	0.112	0.032	0.796
Rpim(I)	0.058	0.015	0.408
Rpim(I+/-)	0.074	0.021	0.523
Total observations	499004	6228	37509
Total unique	129036	1502	9563

Table 3.10. Data-collection and processing statistics for *C. symbiosum* GluDH with NAD⁺.

Parameters	
Resolution (Å)	1.8Å
Number of reflections	122544
Protein molecule per asymmetric unit	3
Number of residues	449
Number of atoms	11004
Number of water molecules	500
Ramachandran favoured (%)	97.32%
Ramachandran outliers (%)	0.37%
RMS bond length deviation (Å)	0.017 Å
RMS bond angle deviation (°)	1.8°
Average main chain B-factors (Å ²)	19Å ²
Average side chain B-factors (Å ²)	26Å ²
Average water B-factors (Å ²)	22Å ²
R-factor (%)	0.17
R _{free} (%)	0.22
Molprobity score	1.52 93 rd percentile

Table 3.11. Shows the final refinement parameters for binary complex of *C. symbiosum* GluDH with NAD⁺.

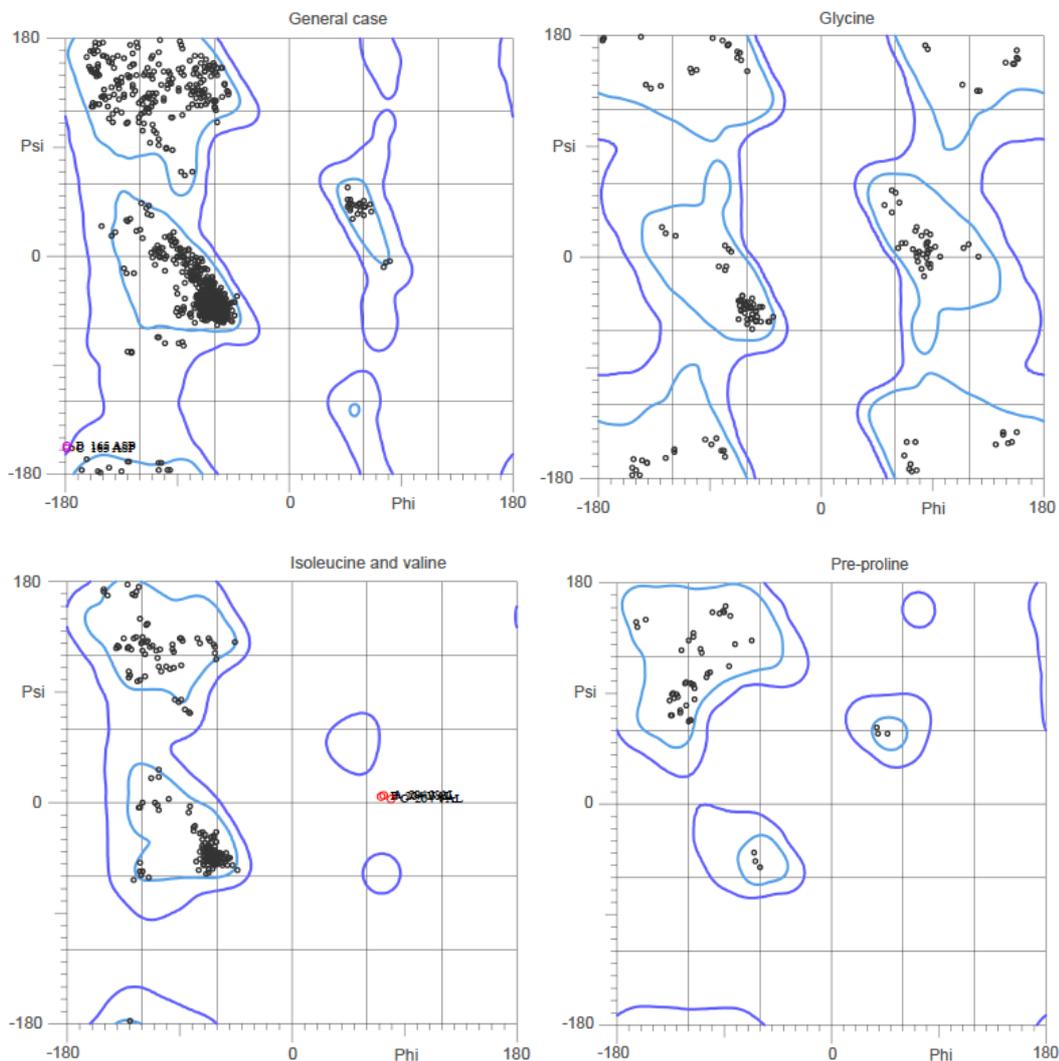


Figure 3.24. Ramachandran plot for binary complex of *C. symbiosum* GluDH with NAD⁺. The plot shows 97.32% of the residues are within the favoured region. Ramachandran outliers are Val204 in A, B and C subunits and Asp165 in B and C subunits. This revised plot was produced using Molprobit server (Chen et al, 2010).

Chapter four: Structural analysis of glutamate dehydrogenase from *M. smegmatis* and *C. symbiosum*

4.0 Introduction

This chapter describes the final model of the ternary complex of *M. smegmatis* GluDH with NADP⁺ and 2-oxoglutarate as well as the structure of *M. smegmatis* GluDH with NADPH. Comparison of the conformational changes between the two structures is discussed, together with a comparison to *C. symbiosum* GluDH and other GluDHs. The structural basis for cofactor specificity in the glutamate dehydrogenases is also discussed.

4.1 Subunit structure of *M. smegmatis* GluDH

The molecular structure of GluDH from *M. smegmatis* is analogous to that seen in other bacterial GluDHs such as *C. symbiosum* (Baker et al, 1992a). The enzyme consists of six identical subunits, each one folded into two domains separated by a deep groove, which comprises the catalytic pocket. Domain I contains the N-terminal segment of the polypeptide chain and involves amino acid residues 1-207 and 377-449. In domain I, the first amino acid residues 1-56 are organized into a cluster of four α -helices followed directly by folding the residues 59-99 into three antiparallel strands. The remaining residues of domain I then fold into a string of successive α/β secondary structure segments. In this domain, the mixed β -sheet (four parallel/antiparallel strands) is located in the domain core and flanked by α -helices on each side.

Domain II (C-terminal domain) is smaller than domain I and it is built from the adjacent stretch of residues 208-376. This domain comprises a seven stranded β sheet in the domain centre, which is again surrounded by α -helices on both sides. The folding pattern of the C-terminal domain is similar to a classical dinucleotide-binding fold, but one of the β strands (β_g) lies in the reverse direction. The whole subunit is built from 13 α -helices with 241 residues and 11 β strands with 65 residues, which represents 68 % of the polypeptide chain.

4.2 Subunit assembly

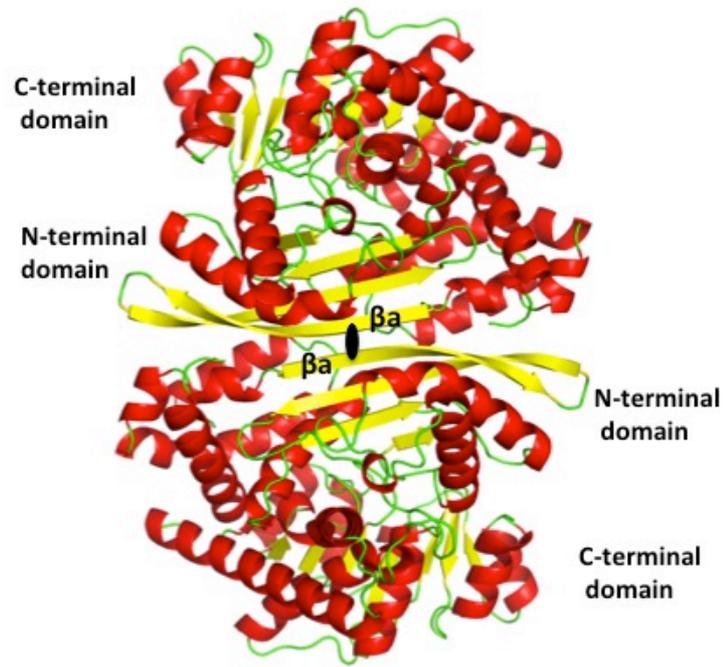
The subunits of GluDH assemble into a hexamer with D₃-symmetry with three 2-fold axis perpendicular to the three-fold axis. This hexamer is like a squat cylinder with the dimensions 108 Å in the direction of the molecular 3-fold axis of symmetry and with a radius of 49 Å through the largest part of the molecule. Domain I and domain II meet in the upper part around the three-fold axis, where the N-terminal domain is close to the 3₂ symmetry point and is responsible for all intersubunit interactions. The dimer interface is centred on the binding between the extended antiparallel β_a in domain I and its 2-fold symmetry related strand β_a (figure 5.1a), from the adjacent subunit.

There are three separate interfaces that should be considered when the hexamer is formed. Dimer interfaces, trimer interfaces and other interactions that are formed through the assembly of the hexamer. The accessible surface areas for monomer, dimer, trimer and hexamer were calculated using the Areaimol program in the CCP4 suite. Within the asymmetric unit, the buried surface area for the dimer interface between the two subunits is 1400 Å², which amounts to 7.8% of the total solvent accessible surface area of the monomer, and involves 40 residues. The subunits around the 3-fold axis are contacted over the area of 1900 Å² (10.5%) and the additional contacts on assembly of the hexamer include buried surface area of 370 Å², which amounts to 2.0% of the total accessible surface area of the monomer. As these buried surface areas show that the maximum contact is linked with the trimers, the GluDH molecule might be described as a dimer of trimers, rather than a trimer of dimers. These results are in agreement with the findings observed in GluDH from *C. symbiosum* (Baker et al, 1992a).

4.3 Alternative conformations of Serine residues in a ternary complex model of *M. smegmatis* GluDH

Within the electron density for the GluDH/NADP⁺/2oxo-glutarate complex there were clear incidences of alternative conformations for some residues. These included Ser16 and Ser339 in both subunits. Inspection of the difference density map showed clear positive density features, and thus the side chains of

a



b



Figure 4.1. Schematic illustrating the assembly of *M. smegmatis* GluDH into the active hexamer with 3₂-symmetry. a) Two subunits come together to form the dimer around the two-fold axis, mediated via $\beta\alpha$ of N-terminal domain (here the two-fold axis comes out of the paper toward you). b) Subunits assemble into the hexamer by the three-fold axis.

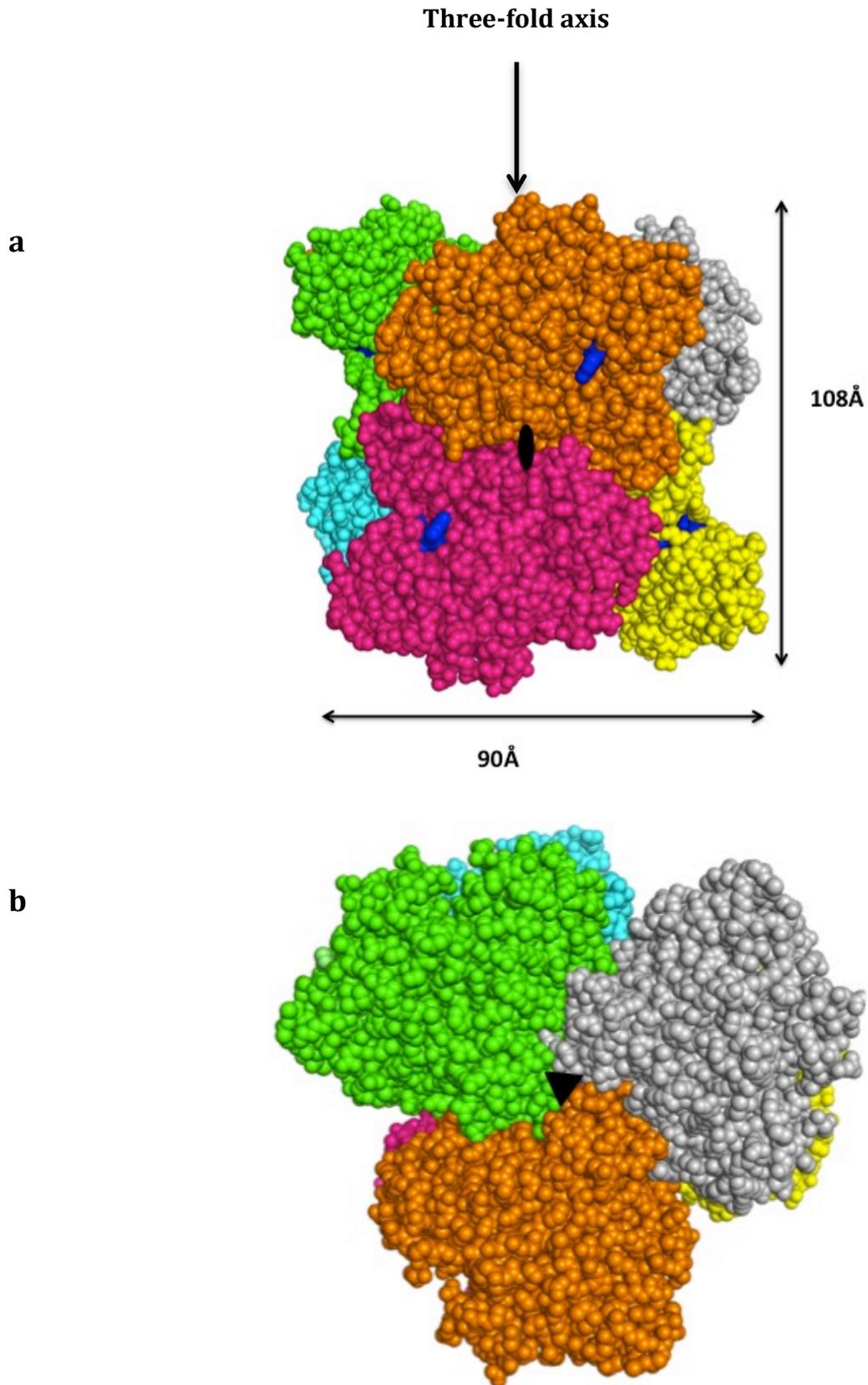


Figure 4.2 The overall organization of *M. smegmatis* GluDH hexamer with each subunit individually coloured. **a)** view down a molecule two-fold axis used to form the dimer, with the three-fold axis vertical. **b)** view down the three-fold axis. The NADP⁺ is shown in a blue color.

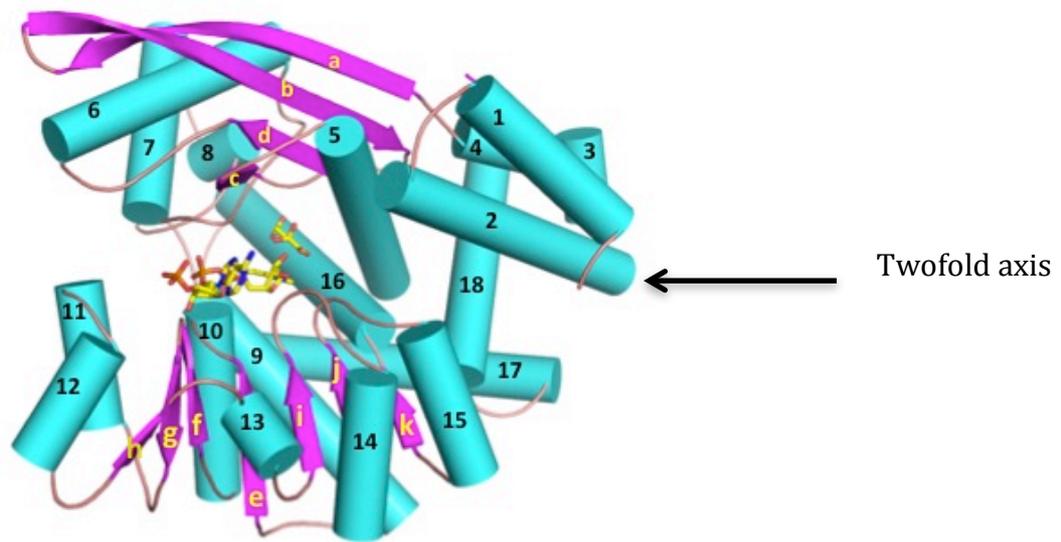


Figure 4.3. Schematic diagram illustrates the overall organization of a single subunit of *M. smegmatis* GluDH in complex with NAD⁺ and 2-oxoglutarate viewed in the twofold axis; α -helices appear as cylinders (numbered from 1-18) and β -strands as arrows (labeled from a-k).

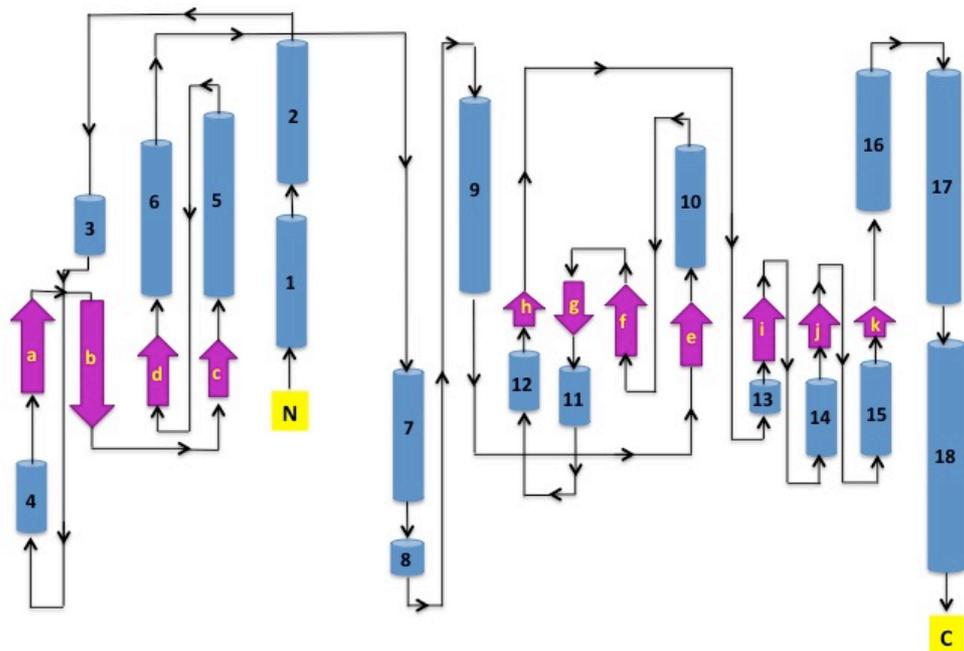


Figure 4.4. A schematic diagram showing the connection of the secondary structure elements. α -helices appear as blue cylinders and β -strands as pink arrows.

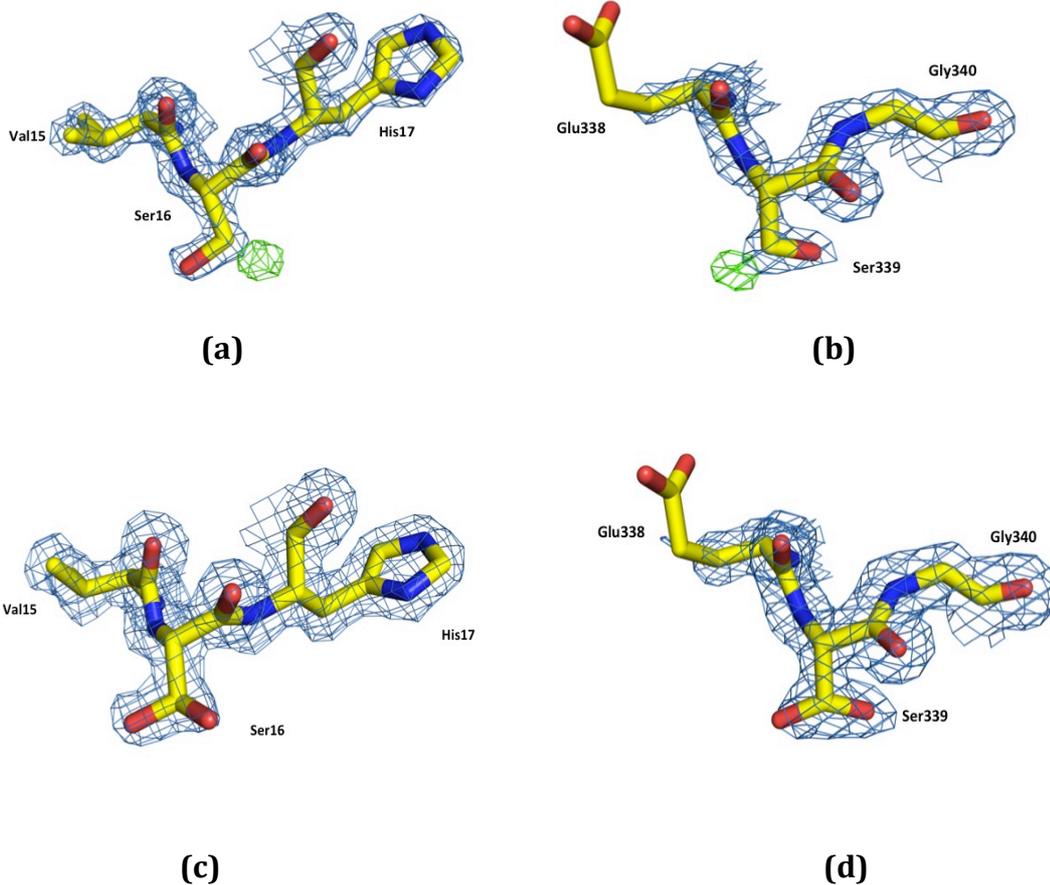


Figure 4.5 The two conformations of the side chain of Ser16 on the left and Ser339 on the right figure in the ternary complex of *M. smegmatis* GluDH. The electron density map is contoured to 1.0 sigma.

these residues were modelled in two conformations, each one with 0.5 occupancies. Figure 5.4 shows both residues fitting in the electron density.

4.4 Binding of the dinucleotide with the enzyme

4.4.1 Comparisons between dinucleotide binding dehydrogenases

Structural comparisons among the dehydrogenase enzymes investigating the nature of the dinucleotide-binding site have revealed that the secondary structure features of the cofactor-binding domain are similar between these enzymes. The dinucleotide binding domain (domain II) has a Rossmann fold and it is composed from six (or seven in GluDHs) stranded-parallel (one strand is in reverse direction in GluDHs) β -sheet surrounded by α -helices on both sides (Birktoft & Banaszak, 1984; Hol et al, 1978). This domain contains the conserved secondary structure unit ($\beta\alpha\beta$), which is responsible for the binding of the dinucleotide. In this conserved unit, the α -helix which connects the two β strands (β_a to β_b) is named the dinucleotide binding helix (Birktoft & Banaszak, 1984; Hol et al, 1978). The $\beta\alpha\beta$ motif (also called the fingerprint region) possesses conserved residues containing a glycine rich loop with the sequence GXGXXG or GXGXXA (where X is any amino acid), which forms a linkage between the end of the β_a and the start of the dinucleotide binding helix. The two β -strands of the $\beta\alpha\beta$ unit are formed from hydrophobic amino acid residues and are protected against the solvent on one side by the hydrophobic face of the dinucleotide binding helix. The other significant feature of this fingerprint region is that an acidic residue frequently occurs at the C-terminal end of β -strand (β_b) which makes hydrogen bonds with the 2' and 3' hydroxyl groups of the adenine ribose, in those enzymes that bind NAD⁺ or FAD⁺ (Baker et al, 1992c).

4.4.2 Recognition of the adenine ribose 3'OH

Analysis of the three-dimensional structure and primary sequence of the fingerprint region in a number of dehydrogenase enzymes has shown that two different hydrogen-bonding patterns (type I and type II) can be made between the glycine-rich turn and the dinucleotide (Baker et al, 1992c). This pattern appears to rely on the nature of the last amino acid in the fingerprint sequence

(GXGXXG/A). These six residues are termed P1 to P6, and an acidic residue on the second β strand (β_b) is termed P7. In the same study, Baker and co-workers showed that, if the residue at P6 is glycine, a type I hydrogen bonding configuration can be formed. In this case, the oxygen atom of the P1 carbonyl group interacts with the α -carbon of P6 by van der Waals' forces, permitting the main chain N-H of P2 to direct to the side chain of the residue at P7. The arrangement of this system allows the formation of hydrogen bonding between the N-H of P2 and the side chain of the residue at P7, which makes two hydrogen bonds with 2' and 3' ribose hydroxyls. However, if the residue at a P6 is alanine, a type II hydrogen bonding pattern can be seen. In this configuration, the peptide between P1 and P2 is forced into another arrangement due to existence of the β carbon at P6, which permits the P2 to form hydrogen bonds with the adenine ribose (Baker et al, 1992c). In the *M. smegmatis* GluDH structure residue P6 is an alanine (Ala246), as is found in *C. symbiosum* GluDH (Ala241), despite the fact that *M. smegmatis* GluDH binds NADP⁺ preferentially (Sarada et al, 1980), compared to the NAD⁺ binding of *C. symbiosum* GluDH. In both these enzymes a direct H-bond is made between the backbone NH of P2, and the 3'OH of the adenine ribose (figure 4.6 a and b). In the *M. smegmatis* GluDH structure a second hydrogen bond is found between the side chain hydroxyl of Ser242 (P2) and the 3'OH of the ribose in the *M. smegmatis* GluDH. In the *C. symbiosum* GluDH the P2 residue is a phenylalanine, so no hydrogen bond can be made between the side chain of phenylalanine and the 3'OH of adenine ribose. Figure 4.6 shows the interactions between the 3'OH adenine ribose and the glycine-rich turn in the *M. smegmatis* GluDH ternary complex and the *C. symbiosum* GluDH binary complex.

4.4.3 Binding the nicotinamide ring and associated ribose moiety and pyrophosphate

The conformation that nicotinamide nucleotides adopt is similar among the family of dehydrogenase enzymes, with the exception of the glycosidic bond of the nicotinamide ribose. Based on the configuration of the nicotinamide moiety around the glycosidic bond, these enzymes can be classified into two types: A type dehydrogenases, which eliminate the hydrogen from the A side (*re*-face)

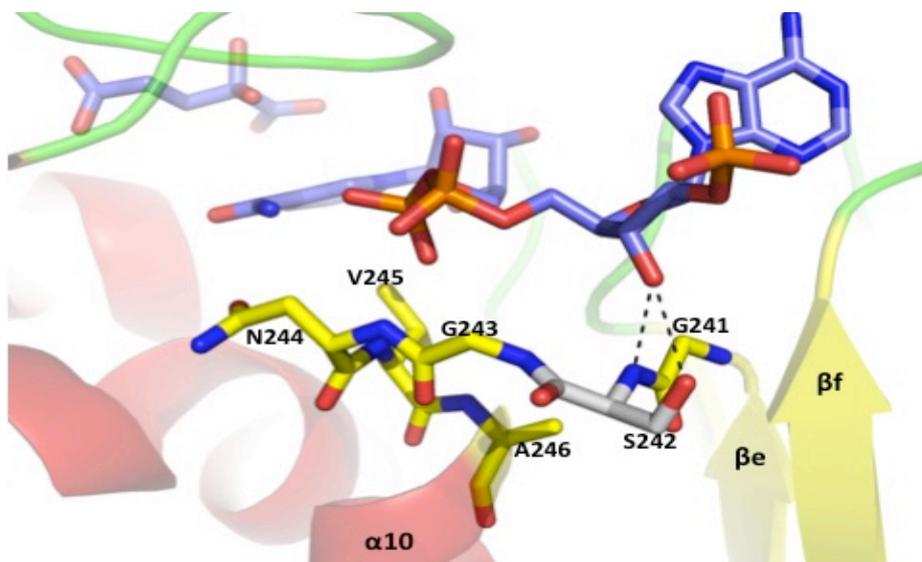
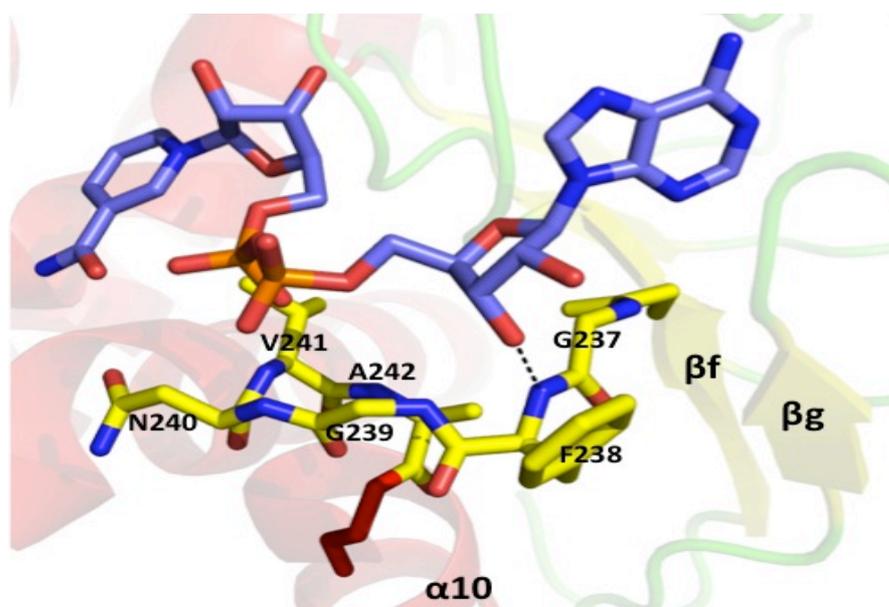
a**b**

Figure 4.6. A schematic diagram illustrating the interaction of the $\beta\alpha\beta$ unit with the cofactor for a) Ternary complex of the *M. smegmatis* GluDH with 2-oxo-glutarate and NADP^+ . In this complex the $\beta\alpha\beta$ motif is shown as $\beta_e \alpha_{10} \beta_f$ and the hydrogen bonds between 3'OH adenine-ribose and Ser242 (P2) can be seen. P1-P6 residues that connect the β_e and the α_{10} are shown in yellow colour; G241-S242-G243-N244-V245-A246 (GXGXXA). b) binary complex of *C. symbiosum* GluDH with NAD^+ . Only one hydrogen bond is formed between the NH backbone of F238 (P2) and 3'OH of adenine ribose. The glycine-rich turn is shown as $\beta_f \alpha_{10} \beta_g$ and P1-P6 residues in this binary are G237-F238-G239-N240-V241-A242.

of the nicotinamide where the NAD(P⁺) adopts the *anti* form, for example; s-MDH, LDH and ADH, and the B type dehydrogenases, that transfer the hydrogen from the B side (*si*-face) where the nicotinamide nucleotides take the *syn* conformation, such as GluDHs. Examination of the electron density map for the cofactor in both subunits of the *M. smegmatis* GluDH ternary complex, shows that this enzyme follows the pattern, with the conformation of the nicotinamide moiety and its linked ribose ring analogous to that observed in the GluDH from *C. symbiosum*, adopting the *syn* conformation (figure4.7a). As a result of this conformation, the carboxamide of the nicotinamide ring makes four hydrogen bonds with adjacent groups. The oxygen atom of the CO group acts as a hydrogen bond acceptor, and forms two bonds with the side chains γ -OH of Thr213 and H-N of Arg209 residues. Furthermore, the carboxamide NH₂ makes two hydrogen bonds, one with the side chain CO of Asn244, and one with O2PN of the NADP⁺ pyrophosphate moiety. The γ -methyl group of the Thr213 side chain is packed against the A face of the nicotinamide, and acts to protect it from the solvent, whereas the B face points towards the active site. The nicotinamide ribose lies against the residues Ala348 and Asn349 allowing the 3' OH and 2'OH to make hydrogen bonds with N-H backbone and NH side chain of Asn349 respectively. Additionally, hydrogen bonds are made between the 2'OH and the side chains of Arg97 and Asp169 (figure5.3). The residues Ala323 and Thr324 are stacked against one face of the ribose ring, close to the C3 and C4 carbon atoms of the ribose.

The pyrophosphate moiety of NADP⁺ of *M. smegmatis* GluDH ternary complex is accommodated at the N-terminal end of the dinucleotide binding helix (α_{10}), with the helix dipole stabilizing the pyrophosphate. The four oxygen atoms of the pyrophosphate interact with neighboring groups. An oxygen atom of the phosphate group (O1N) binds to the main chain N-H of Val245 by hydrogen bonding, whereas the O2N hydrogen bonds to H-N of the nicotinamide moiety. Two hydrogen bonds are made between oxygens O1A and O2A to the N-H backbone of Gly171 and Gly244, respectively.

The binding of the NAD⁺ dinucleotide to the *C. symbiosum* GluDH is in general very similar to that seen for NADP⁺ in *M. smegmatis* GluDH. The residues that

make interactions with the dinucleotide are conserved between these two enzymes. However some differences in the interactions are seen in the nicotinamide moiety due to the difference in the domain closure between the two structures. For instance, in the binary complex of *C. symbiosum* GluDH with NAD⁺ that is in the open form, no hydrogen bond can be seen between Arg205 and the CO of carboxamide. In addition the interactions between the 2'OH of nicotinamide ribose with Asp165 and Arg93 are not made and O2A of the pyrophosphate does not form a hydrogen bond to the main chain of Gly167. All of these residues are conserved in *M. smegmatis* GluDH (Arg209, Asp169, Arg97 and Gly171, respectively), but do interact with the coenzyme due to the more closed nature of the enzyme. Table 4.1 shows the hydrogen bonds made between the cofactor and GluDH in both these complex structures; ternary complex *M. smegmatis* GluDH and *C. symbiosum* GluDH bound to NAD⁺.

4.4.4 Comparison between the binding of dinucleotide in the NADPH binary complex and NADP⁺ and 2-oxoglutarate ternary complex of *M. smegmatis* GluDH

The cofactor is bound in the two structures (GluDH-NADPH and GluDH-NADP-2xoglutarate) in a very similar way, with the exception of the binding of the nicotinamide moiety. In the ternary complex model, the carboxamide of the nicotinamide ring forms four hydrogen bonds with the adjacent groups; O2N and the side chains of Thr213, Arg209 and Asn244. In the binary complex structure, however, Arg209 and Asn244 residues adopt different conformations compared with the ternary complex, and thus do not interact with the carboxamide moiety. Moreover, superimposing the two structures shows that the orientation of R209 is different because the loop that this residue fall on, is also in a different conformation, which is related to the missing residues in this loop in the electron density map (193-204) (figure 4.11).

4.4.5 Coenzyme specificity

As structures of both the NADP⁺ complex GluDH from *M. smegmatis* and the NAD⁺ complex of GluDH from *C. symbiosum* have been determined, the nature of coenzyme specificity in GluDHs can be addressed. NADP⁺ binding to *E. coli*

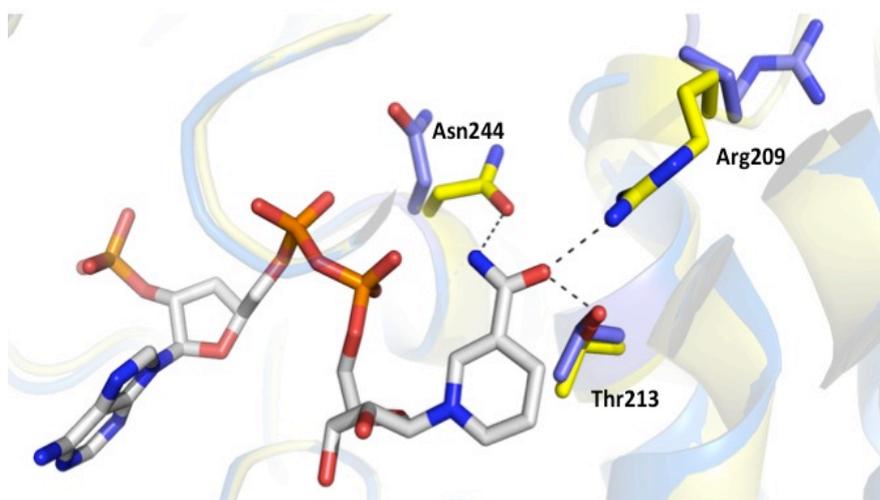


Figure 4.7. A schematic of the interaction of the nicotinamide ring in the binary complex structure (GluDH-NADPH) (blue colour) and ternary complex model (GluDH-NADP⁺-2oxo-glutarate) (yellow color) for *M. smegmatis* GluDH. In the binary complex Arg209 and Asn244 are directed away from the carboxamide and are not involved in nicotinamide binding.

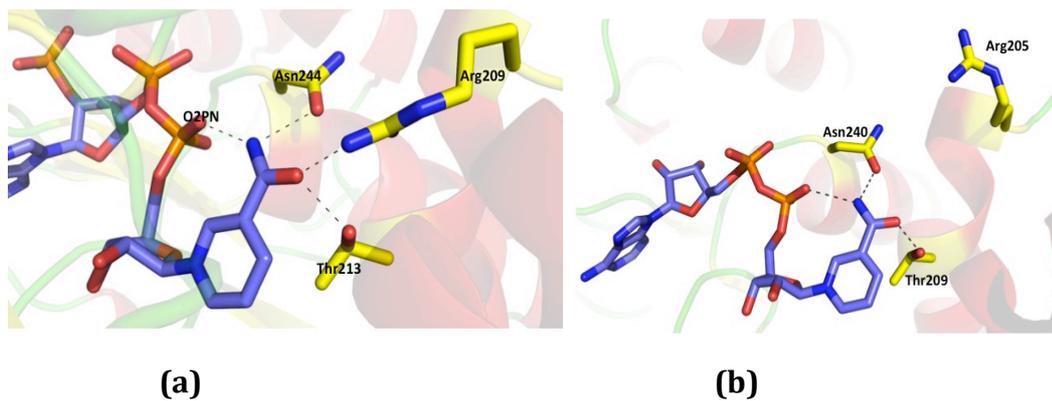


Figure 4.8. Schematic diagram of the interactions made by the carboxamide of the nicotinamide ring with neighbouring groups in **a)** *M. smegmatis* GluDH and **b)** *C. symbiosum* GluDH.

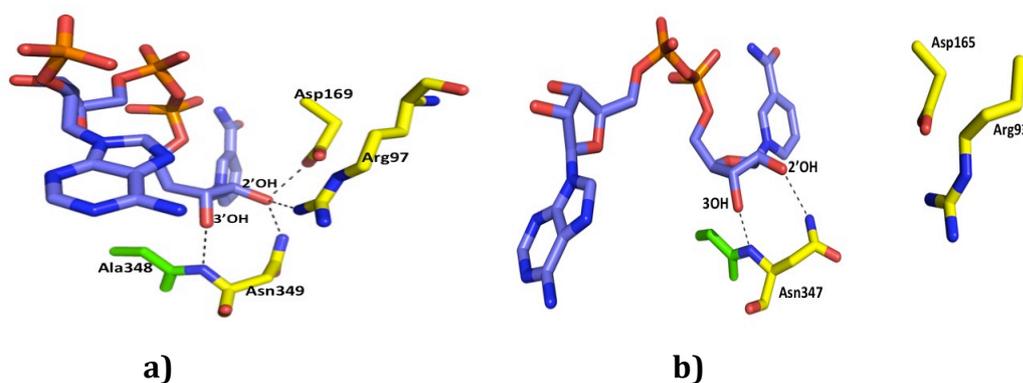


Figure 4.9. Schematic diagram of the 2' and 3'OH of the nicotinamide ribose forming hydrogen bonding interactions with adjacent residues (shown in yellow colour), for a) *M. smegmatis* GluDH/NADP⁺/gem-diol and b) *C. symbiosum* GluDH/NAD⁺.

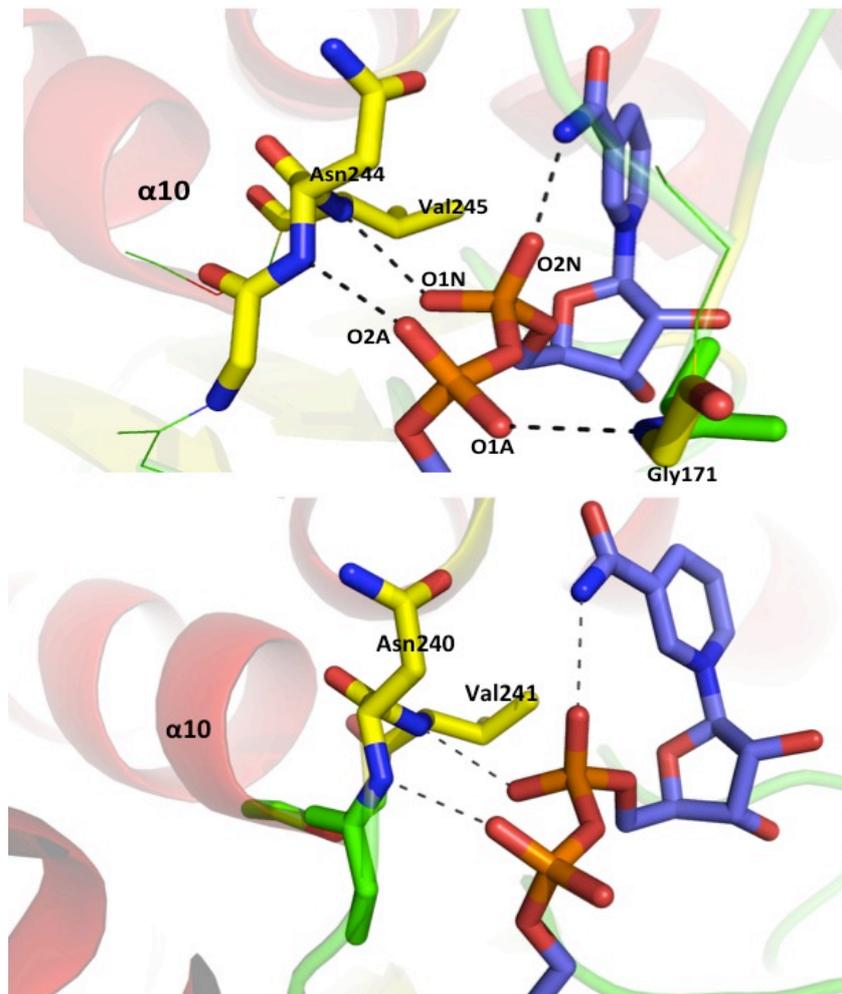


Figure 4.10. Schematic diagram illustrating the pyrophosphate binding, which lies at the end of $\alpha 10$ and forms hydrogen bonds with the main chains of Asn244, Val245 (P4 and P5) and Gly171. In addition, a hydrogen bond between the H-N of the nicotinamide carboxylate and O2PN of the pyrophosphate is shown.

NAD ⁺ /P ⁺	Hydrogen bonds of NADP ⁺ in <i>M. smegmatis</i> GluDH ternary complex	Hydrogen bonds of NAD ⁺ in <i>C. symbiosum</i> GluDH binary complex
Nicotinamide ring		
Carbonyl group	i-Side-chain of Thr213 2.9 Å ii- side-chain Arg209 2.86 Å	i- side-chain of Thr209 2.53 Å ii- No hydrogen bond
Amino group	i- O2PN of pyrophosphate 2.78 Å ii- CO of side-chain of Asn244 3.33 Å	i- O2PN of pyrophosphate 2.94 Å ii- CO of side-chain of Asn240 3.13 Å
Nicotinamide ribose		
2'OH	i- NH2 side-chain of Asn349 2.85 Å ii- side-chain of Asp169 2.66 Å iii- side-chain of Arg97 2.93 Å	i-no hydrogen bond (Asn347) ii-no hydrogen bond (Asp165) iii-no hydrogen bond (Arg93)
3'OH	i- main chain NH2 of Asn349 3.12Å	i- main chain NH2 of Asn347 3.19 Å
Pyrophosphate		
O1N	i-main chain NH2 of Val245 2.93 Å ii- main chain NH2 of Asn244 3.13 Å	i- main chain NH2 of Val241 2.97 Å ii- main chain NH2 of Asn244 3.27 Å
O2N	i- NH2 of carboxamide 2.78 Å	i- NH2 of carboxamide 2.94 Å
O1A	i- main chain NH2 of Gly171 2.97 Å	i-no hydrogen bond (Gly167) ii-side-chain of Arg285 3.32 Å
O2A	i- main chain NH2 of Asn244 2.92 Å	i- main chain NH2 of Asn244 3.07Å
3'OH adenine ribose	i-NH2 of main chain of Ser242 3.02 Å ii- side chain of Ser242 3.12 Å	i- NH2 of main chain of Phe238 2.94 Å

Table 4.1. Hydrogen bonds between NAD(P)⁺ and GluDH enzymes.

GluDH has been modelled recently by Sharkey and co-workers (Sharkey et al, 2013) and the residues that lie at P7 and P8 (at the C-terminal end of the β strand following the GXGXXG/A motif in the Rossmann fold) have been suggested as the determinant of NADP⁺/NAD⁺ specificity. The 2' phosphate of NADP⁺ in the *M. smegmatis* GluDH/NADP⁺/2oxoglutarate complex makes interactions with the side chains of R291 and K285 (R292 and K286 in *E. coli* GluDH) as had been predicted by Sharkey *et al* (Sharkey *et al*, 2013). The role of the residue at P7 has remained unclear. It had previously been suggested that if an acidic residue was present at this position, the enzyme would select for NAD⁺, as the carboxylate could interact with both the 2' and 3' hydroxyls of the adenine ribose, as seen in, for example, lactate dehydrogenase (Adams et al, 1973). However, the NAD⁺-dependent *C. symbiosum* GluDH has a glycine at P7, and the NADP⁺-dependent *E. coli* and *M. smegmatis* GluDHs have an aspartate at this position. Various theories for altered pKs of this residue, and whether it makes water mediated hydrogen bonds have been put forward (Baker et al, 1992c; Dawson), but, in the NADP⁺ complex of *M. smegmatis* GluDH, this residue points away from the adenine ribose, towards the N-terminal helix α 12 of the α 11-loop- α 12 motif, and could be involved in the stabilization of the helix dipole of this helix. This orientation allows the P7 residue to hydrogen bond to the NH backbone of Ile292. This is in agreement with the position of the equivalent residue (Val293) in the *E. coli* apo structure (Sharkey et al, 2013). The 2' phosphate also makes a hydrogen bond to the side chain of Ser266 (position P8), which again is in agreement with the model proposed by Sharkey *et al* for *E. coli* GluDH (Ser264 is the equivalent residue). However, the structure of the *M. smegmatis* GluDH/NADP⁺/2oxoglutarate ternary complex, which is in a very closed conformation (see section 4.5.3), reveals that an additional interaction is formed between the 2' phosphate and the amine moiety of the side chain of the Lys137, which lies in domain I. This interaction can only occur when the cleft is closed and perhaps enables the enzyme to sense when both substrates are bound.

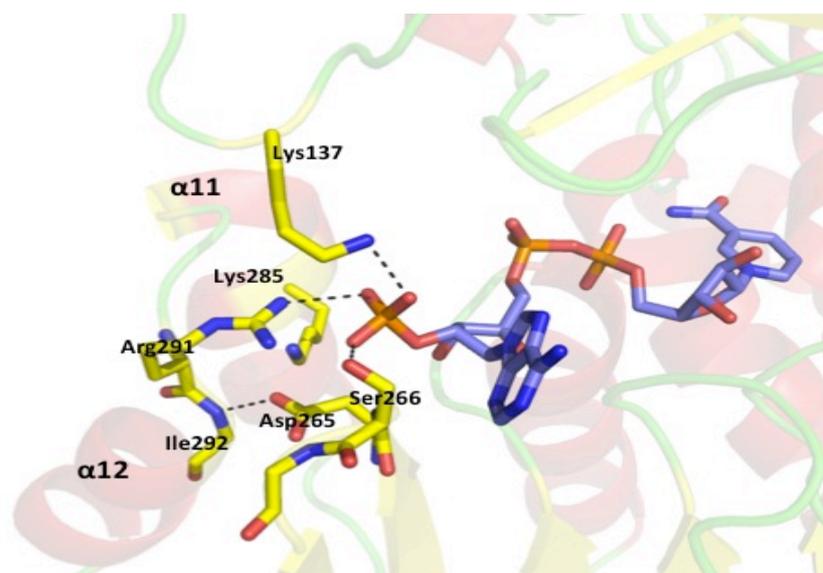


Figure 4.12. Shows the three positively charged residues (Lys137, Lys285 and Arg291) and the Ser266, which are adjacent to 2'phosphate group in the GluDH-NADP⁺- 2-oxoglutarate complex from *M. smegmatis*.

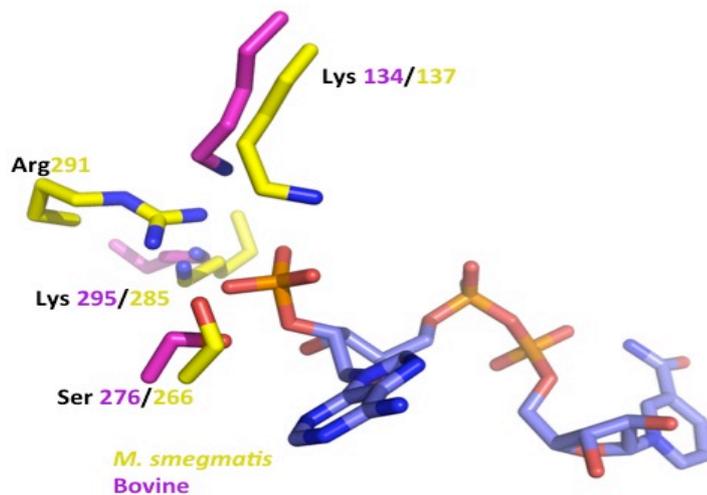
The character of the NADP⁺-binding site in *M. smegmatis* GluDH shows a close similarity with bovine-GluDH, a dual-dependent enzyme, especially in the vicinity of residues that interact with the 2'phosphate group of the adenine ribose. The two basic residues Lys137 and Lys285 with Ser266 in *M. smegmatis* GluDH are conserved in bovine-GluDH as Lys134, Lys295 and Ser276 respectively. However, the orientation of the Lys137 in *M. smegmatis* GluDH is different to Lys134 in bovine-GluDH (both residues are conserved) and the reason for that might be due to the presence of the Arg291 in *M. smegmatis* GluDH (which is equivalent to Thr257 in bovine-GluDH) adjacent to the 2'phosphate group (figure 4.14). In both enzymes these residues might permit charge compensation of the 2'phosphate group. As mentioned above the side chain of Asp265 (P7) makes a hydrogen bond with the main chain of N-H of Ile292. This residue (Asp265) is equivalent to Glu275 in bovine-GluDH, which forms two hydrogen bonds by its side-chain and main chain with the 2'phosphate group (Smith et al, 2001), and thus points in a different direction.

4.5 Conformational change of GluDH from *C. symbiosum* and *M. smegmatis*

4.5.1 Open and closed conformations in *C. symbiosum* GluDH

Several enzymes undergo molecular rearrangement as a result of the binding of substrates and cofactors (Lesk & Chothia, 1984). In such enzymes the catalytic pocket is located in a cleft between two domains and the structural rearrangements close the domains often through association with a substrate and a cofactor. Domain closure is involved in the catalytic activity, where it aids in orientation of the catalytic residues about the substrate, differentiates against unsuitable substrates, and eliminates water from the catalytic pocket (Lesk & Chothia, 1984). Structural studies on the glutamate dehydrogenase family have shown that the two domains of the enzyme subunit can exist in a number of different conformations, including an open form, required for substrate binding, and a closed form essential for catalysis (Baker et al, 1992a; Smith et al, 2001; Stillman et al, 1993). In the *C. symbiosum* GluDH for example; the apo enzyme (1HRD) is

a



b

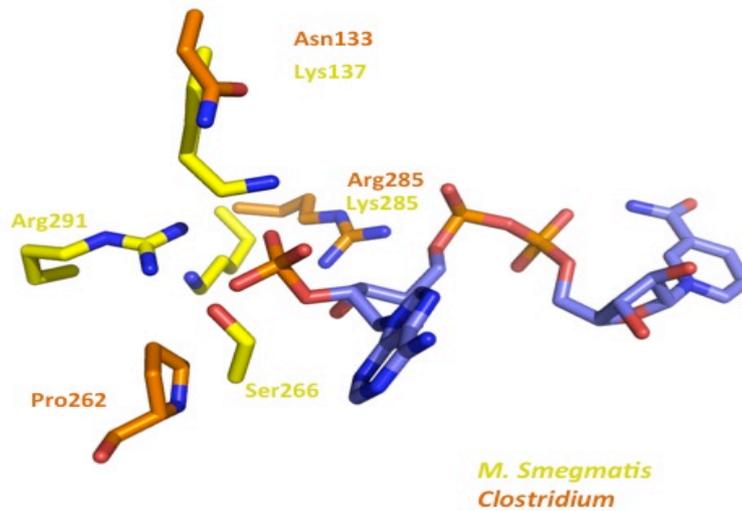


Figure 4.14. The binding site of the NADP⁺ of *M. smegmatis* GluDH with the residues implicated in interacting with the 2'phosphate group displayed with yellow carbon colour. **a)** The pink color residues are a superposition of the corresponding residues of the bovine GluDH/Glutamate/NADPH complex (3MVQ). **b)** The orange color residues are a superposition of the corresponding residues extant in the *C. symbiosum* GluDH (1HRD).

in the open conformation, and the glutamate bound binary complex is in a closed conformation (Baker et al, 1992a; Stillman et al, 1993). The structural rearrangement between the two forms can be described as a rigid body movement of roughly 15° about an axis lying between the two domains (Stillman et al, 1993).

4.5.2 K89L mutant of apo-enzyme of *C. symbiosum* GluDH present in a closed form

Stillman and co-workers have also solved the three-dimensional crystal structure of the K89L mutant of GluDH from *C. symbiosum*. Comparison of this structure with the previous structures of wild-type *C. symbiosum* GluDH, have shown that the structure of this mutant is similar to that of the GluDH/L-glutamate binary complex. This means that the K89L mutant enzyme adopts a closed conformation, even in the absence of substrate (Stillman et al, 1999). It is unclear whether the K89L mutation facilitates the domain closure, by mimicking the substrate binding by removing the positive charge of the lysine residue, which lies at the interface between domain I and domain II (Stillman et al, 1999).

The structure of the wild-type apo form of the *E. coli* GluDH also adopts a closed conformation (4BHT, (Sharkey et al, 2013), showing that the domain movement in GluDH is not dependent on substrate or cofactor binding. To determine the level of domain closure in the structures presented in this thesis, the N-terminal domains I of a number of GluDH structures were superimposed on domain I of the A subunit of the *C. symbiosum* apo enzyme (1HRD), which is the reference structure. These overlapped structures were, in turn, superimposed on the C-terminal domain II of 1HRD. Using the superposition matrices describing this second superposition, the domain movement between the structures was calculated as screw axis rotation, using the method of Gerstein (Gerstein et al, 1993). The results of these analyses are shown in table 4.2. In each case the rotation that describes domain closure lies between the domains, and approximately parallel to $\alpha 17$. These superpositions show the range of the domain movements possible in the GluDH family. Compared to the *C.*

Protein	Subunit	Rotation	Distance
<i>C. symbiosum</i> GluDH (1HRD)	A	0°	Cα A95-Cα G239 22.4 Å
<i>M. smegmatis</i> GluDH ternary complex	A	25.6°	Cα H99-Cα G243 13.6Å
	B	25.6°	13.6Å
<i>C. symbiosum</i> GluDH binary complex (GluDH-Glutamate) 1.2 Å	A	17.7°	Cα A95-Cα G239 17.5 Å
<i>C. symbiosum</i> GluDH binary complex (GluDH-NAD ⁺)	A	3.3°	Cα A95-Cα G239 21.7 Å
	B	3.7°	20.8 Å
	C	16.4°	17.9 Å
<i>C. symbiosum</i> GluDH binary complex (GluDH-Glutamate) (1BGV)	A	15.4°	Cα A95-Cα G239 17.8 Å
K89L mutant of <i>C. symbiosum</i> GluDH (1K89)	A	15.4°	Cα A95-Cα G239 17.7 Å
K89L/S380V mutants of <i>C. symbiosum</i> GluDH (1AUP)	A	23.5°	Cα A95-Cα G239 14.9 Å
Bovine ternary complex of GluDH (GluDH/NADPH/Glutamate) (3MVQ)	A	26.5°	Cα S96-Cα G253 14.1 Å
	A	25.9°	13.7 Å
Bovine ternary complex of GluDH (GluDH/NAD ⁺ /2oxoglutarate) (1HWY)	A		
Apo <i>E. coli</i> GluDH (4BHT)	A	14.9°	Cα H98-Cα G241 18.0 Å

Table 4.2. Shows the domain motion in GluDHs. The relative of the domain movement is described by the rotation and the distance between the two domains, which is measured at the equivalent residues.

symbiosum apo structure (1HRD) the *C. symbiosum* L-glutamate binary complex (1BGV) has a domain closure of $\approx 15^\circ$. The 1.2Å *C. symbiosum* L-glutamate binary complex structure determined here has a slightly more closed structure of 18° , whereas the *M. smegmatis* ternary complex is in a fully closed form, with a domain closure of $\approx 26^\circ$. Interestingly, the ternary complexes of bovine GluDH (NADPH/Glu or NAD⁺/oxoglu) have a similar fully closed structure to the *M. smegmatis* ternary complex, also with a domain closure of $\approx 26^\circ$. In contrast, the *E. coli* apo structure adopts a similar closed structure to the *C. symbiosum* L-glutamate binary complex with a domain closure of $\approx 15^\circ$.

The relative domain closure can also be illustrated by measuring the distance between equivalent residues in the two domains across the different structures. As there are some variations in the fold of the structures around helices $\alpha 11$ and $\alpha 12$ in domain II, which lie at the jaws of the cleft, and would thus give ambiguous results, the distance compared here is that between the α carbons of the second glycine of the GXGXXG/A turn and the residue at the C-terminal end of βc of domain I (His99 in *M. smegmatis* GluDH). The distance between these two α carbons varies from 22Å in the *C. symbiosum* apo open structure (1HRD) to 13.5 Å in the *M. smegmatis* GluDH fully closed ternary complex structure (table4.2). Analysis of these structures appears to show that the movement between the open and closed (15°) structures is not dependent on substrate binding, as closed structure can be observed without substrate. However, the fully closed structure (26°) is only observed when both substrate and cofactor are present. On substrate binding there are a number of changes in the torsion angles of the side chains of the residues that interact with the glutamate or 2-oxoglutarate substrate. These residues (S380, K89, R205, Q114, D165, K125 and K113) in *C. symbiosum* GluDH (table4.3) are conserved across the family and so the mode of binding of substrate, and the reaction mechanism are likely to be very similar.

One aspect of the closed structure is that, when the domains close, additional interactions are made between domain I and the cofactor which cannot occur in the open form; these extra interactions are:

- i) The side chain of Lys137 makes a hydrogen bond to the 2'phosphate of NADP⁺
- ii) The backbone NH of Gly171 forms a hydrogen bond to O1A of the pyrophosphate (NAD⁺ or NADP⁺)
- iii) The side chain of R205 interacts with the carbonyl of the nicotinamide carboxamide.

4.6 Catalytic pocket

4.6.1 Substrate binding in GluDH

The previous structure of the *C. symbiosum* GluDH/L-glutamate binary complex (1BGV) has shown how the L-glutamate substrate is accommodated in the active site of the enzyme (Stillman et al, 1993). The substrate binds at the base of the active site cleft between domains I and II, accommodated in a pocket on the enzyme surface. Recognition of the substrate side chain occurs by K89 and S380 residues, which interact by their side chains with the γ -carboxyl group of L-glutamate. The side chains of V377 and A163 form hydrophobic interactions with the β and δ -carbon atoms of the amino acid substrate. Further interactions are made between α -amino group of the glutamate to the side chain of D165 and the main chain carbonyl group of G164, with the α -carboxyl group interacting with the side chains of K113 and Q110. The active site also comprises five glycine residues (G122, 123, 90, 91 and G376) that play a role in the shape of the catalytic pocket and K125, which lies close to the amino group of the substrate. This structure also enabled a reaction mechanism to be proposed (Stillman et al, 1993). All of these amino acid residues are conserved in the GluDH family (Britton et al, 1993b).

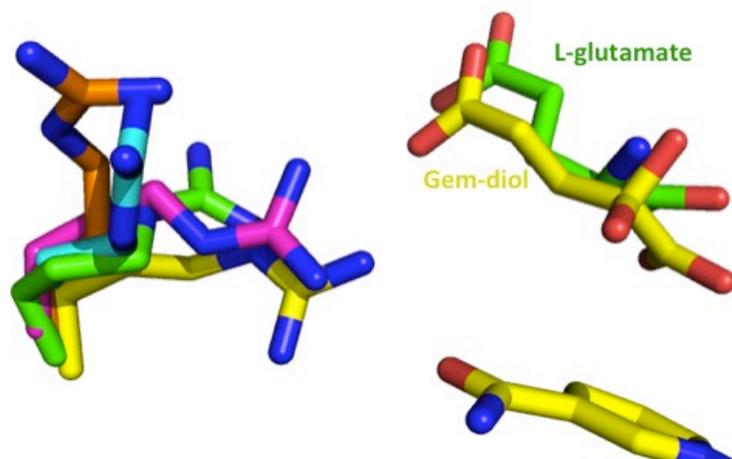
This thesis presented a high resolution (1.2 Å) structure of the *C. symbiosum* GluDH/L-glutamate complex, a binary complex of NAD⁺ with the open form of the *C. symbiosum* GluDH, and also the first bacterial GluDH ternary complex of the *M. smegmatis* GluDH/NADP⁺/gem-diol structure. Comparison of these structures has allowed a number of new observations to be made about the active site in this family of enzymes. These observations include:

1- Arg205 senses binding of L-glutamate

As well as the previously observed interactions between the γ -carboxyl of L-glutamate with S380 and K89 in *C. symbiosum* GluDH, the side chain of R205 also appears to sense the presence of glutamate or 2-oxoglutarate substrates. In the apo structures (e.g. 1HRD and 1K89) this residue interacts just with water molecule in the cleft, whereas when L-glutamate is present (1BGV and 1.2 Å complex) the side chain moves towards the glutamate-binding pocket and interacts with the side chain carboxyl of the substrate. In the ternary complex structures (e.g. diol and bovine) this side chain also interacts with the carboxyl of the carboxamide of the cofactor, aiding the positioning of the nicotinamide ring to the alpha carbon of the substrate (figure 4.14 and figure 4.15).

2- The position of the substrate in the active site is flexible

In the 1.2 Å *C. symbiosum* GluDH/L-glutamate binary complex, two conformations of the γ -carboxyl can be seen in the electron density (figure 4.15), one where the γ -carboxyl interacts with R205 and S380 and the other where the γ -carboxyl interacts with S380 and K89 (figure 4.17). In both of these conformations of the L-glutamate side chain, the α -carbon, amino, and carboxyl groups of the substrate lie in effectively the same place. In the fully closed *M. smegmatis* GluDH/NADP⁺/gem-diol ternary complex, the substrate just occupies one position, with the γ -carboxyl occupying the position that makes interactions with the serine and lysine (K93 and S381 in *M. smegmatis* GluDH) (figure 4.16). In this fully closed ternary complex, the arginine residue (R209 in *M. smegmatis* GluDH) moves even closer to the substrate, than that is seen in the closed binary complex, such that interactions are made to the γ -carboxyl. Thus in the ternary complex, R209, K93 and S381 all make stabilizing interactions with the γ -carboxyl of the substrate. However, although the position of the γ -carboxyl in the ternary complex is the same as one of those seen in the binary complex, the position of the rest of the substrate is substantially different. The gem-diol binds in an extended conformation, compared to the bent conformation of the L-glutamate. The α -carbon of the gem-diol



R205 of *C. symbiosum* GluDH apo state (1HRD)
 R205 of *C. symbiosum* GluDH single mutant K89L apo state (1K89)
 R205 of *C. symbiosum* GluDH in complex with L-glutamate (1BGV)
 R209 of *M. smegmatis* GluDH ternary complex (GluDH/NADP⁺/gem-diol)
 R211 of bovine GluDH ternary complex (GluDH/NADPH/Glutamate)

Figure 4.15. Illustrates the orientation of the arginine residue (R205) from *C. symbiosum* GluDH. R205 in *C. symbiosum* GluDH in the apo state are shown in orange and blue colour, this residue moves away from the substrate, but in the GluDH that bound to the substrate (green, yellow and pink) the arginine directs to the substrate.

occupies the position of the α -carboxylate of L-glutamate with the α -carboxylate making direct interactions to K129, K117, Q114 and N349.

This difference in the position of the α -carbon of the substrate (1.7Å) has the effect of moving the oxidized substrate away from the nicotinamide ring (figure 4.17).

3- The gem-diol provides a model for the carbinolamine

The gem-diol bound in the *M. smegmatis* GluDH/NADP⁺/gem-diol ternary complex has two hydroxyl groups attached to the sp³ coordinated α -carbon of the ligand and is thus a good model for the carbinolamine intermediate. One of these hydroxyls forms interactions to the main chain NH of G95, the carbonyl of the side chain of Q114 and the side chain of K129, with the other hydroxyl lying closer to the nicotinamide ring of NADP⁺ and interacting with the side chain of D169.

If we assume that the proposed carbinolamine binds in the same position as that seen in the diol, then the stereochemistry of the carbinolamine can be inferred. The α -amino group of the carbinolamine can only act as a hydrogen bond donor, and so must occupy the position of the hydroxyl on diol closest to D169 and the nicotinamide ring, rather than the other position, which must act as a hydrogen bond acceptor. This would then mean that the carbinolamine has the S-configuration.

Taken together, these observations enable an extension to the reaction mechanism proposed by Stillman *et al* (Stillman et al, 1993), which is summarized below (using *C. symbiosum* numbering for clarity)

i) As the majority of the interactions with the dinucleotide are made with domain II, the assumption can be made that a fully closed ternary complex with L-glutamate and NAD⁺ also occurs as part of in the enzyme mechanism, where the L-glutamate occupies the position of that seen in the ternary complex, and the nicotinamide ring occupies that seen in the ternary complex. In this proposed model the α -carbon of the L-glutamate lies 2.5 Å from the C-4 the nicotinamide, in an ideal orientation for hydride transfer (figure4.17), with D165 acting as a base to abstract a proton from the α -amino group. The hydride transfer would result in the formation of the iminioglutarate.

Glutamate dehydrogenase	Conserved residues						
<i>C. symbiosum</i>	K89	K113	Q110	K125	D165	R205	S380
<i>M. smegamtis</i>	K93	K117	Q114	K129	D169	R209	S381
<i>E. coli</i>	K92	K116	Q113	K128	D168	R207	S380
<i>P. asaccharolyticus</i>	K70	K94	M91	K106	D146	R187	S355
Bovine	K90	K114	M111	K126	N168	R211	S381
	K147	K170	M167	K183	D225	R268	S438

Table 4.3. Showing the conserved and equivalent residues that are responsible for the substrate binding for GluDHs. The residue numbering for Bovine GluDH is given for both the mature mitochondrial protein (upper line) and for the genomic sequence including the N-terminal mitochondrial targeting sequence (lower line).

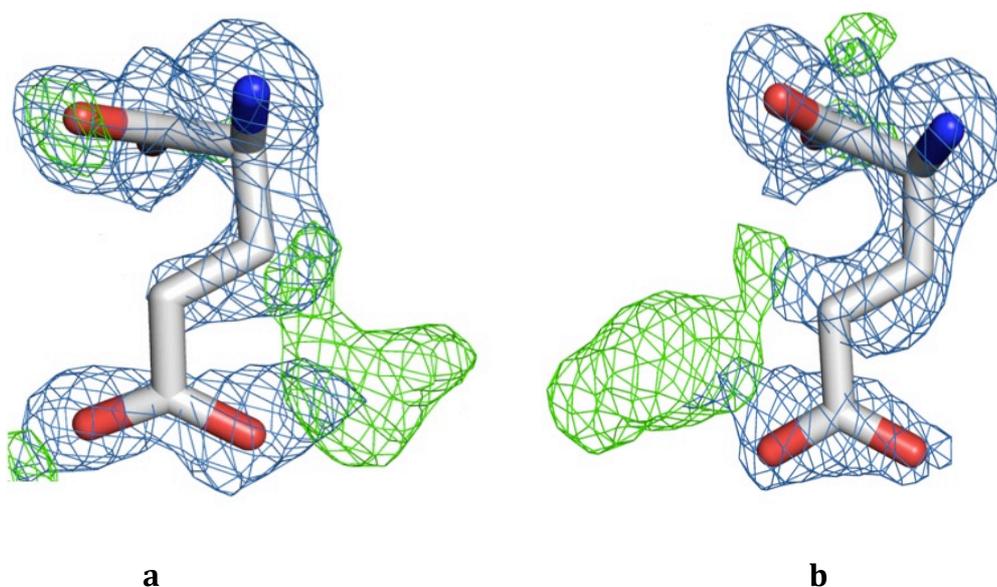


Figure 4.16. Electron density for the L-glutamate of 1.2Å binary complex of *C. symbiosum* GluDH with L-glutamate. The different density for the two conformations of the substrate side chain can be seen. In each case the map (countered at 1.0 sigma) is shown in blue, with the positive difference map (countered at 3.0 sigma) in green.

ii) With an sp^2 hybridized α -carbon, the imino group of the iminoglutarate would lie too close to the nicotinamide ring, but interconversion to the extended form of the substrate would move the α -carbon towards the position seen in the gem-diol to alleviate this potential steric clash and give a 3.9\AA distance between nicotinamide C4 and the substrate $C\alpha$.

iii) During this movement, the water molecule hydrogen bonded to K125 attacks the iminoglutarate to give the S-carbinolamine binding in a similar way to that seen for the gem-diol in the ternary complex. Subsequent collapse of the carbinolamine to the ketone and the change of sp^3 to sp^2 coordination of the $C\alpha$, would result in the carboxyl of the keto acid lying approximately parallel to the plane of the nicotinamide ring. This would be facilitated by D165 acting as a base to abstract a proton from K125, which, in turn, acts as a base to remove the proton from the hydroxyl of the carbinolamine. The enzyme can then open up, to allow release of the products and proton exchange of D165 and K125 with the bulk solution.

This proposed mechanism provides an explanation for how the enzyme avoids the abortive formation of a α -hydroxy acid when the reaction is running in the reverse direction. The α -carbon of the diol (or presumably the keto acid) lies 3.9\AA away from C4 of the nicotinamide and thus direct attack by the hydride on the keto acid is not possible. Only when NH_4^+ , stabilized by D109, has attacked, to give an sp^3 hybridized α -carbon in the S-carbinolamine, does the substrate move towards the nicotinamide ring to allow hydride transfer to form the amino acid. This change in the distance between the substrate $C\alpha$ and the nicotinamide C4 is facilitated by movement of the substrate in the active site, rather than by the domain moving, which has the consequence of keeping the cleft closed throughout the reaction and thus preventing unwanted attack by bulk water.

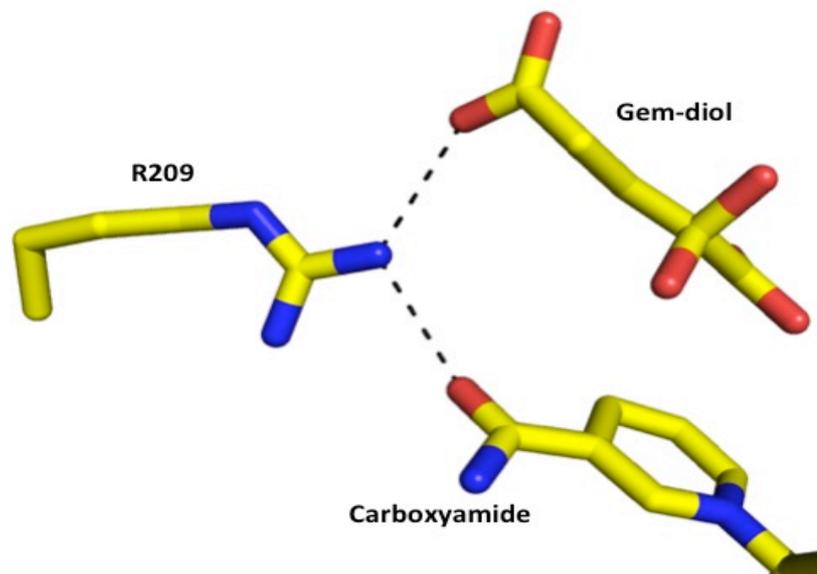


Figure 4.17. Shows the binding of the side chain of R209 with γ -carboxyl group of the gem-diol and the carbonyl group of the carboxamide of the nicotinamide of the NADP⁺ in the ternary complex of *M. smegmatis* GluDH/gem-diol/NADP⁺.

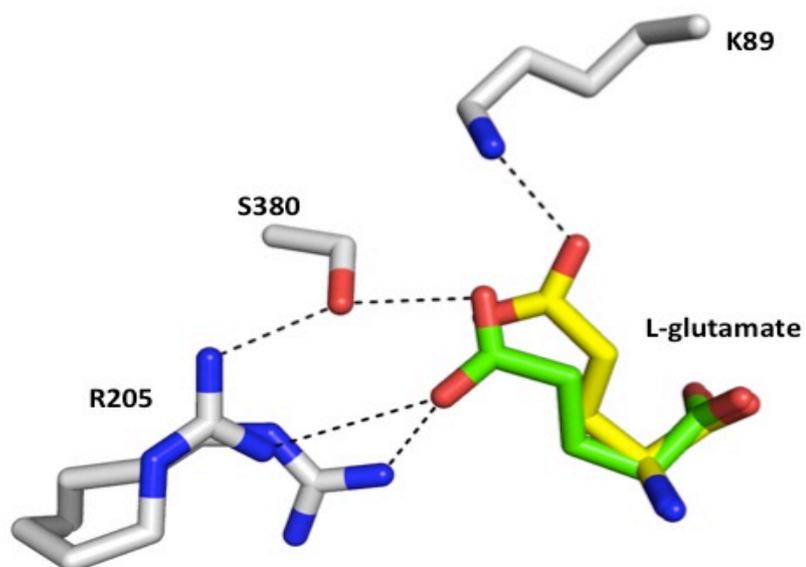


Figure 4.18. The two conformations of γ -carboxyl of the L-glutamate in the *C. symbiosum* GluDH binary complex are shown in different colors (green and yellow), also R205 has two forms.

	Hydrogen bonds of possible diol in <i>M. smegmatis</i> GluDH ternary complex	Hydrogen bonds of L-glutamate in <i>C. symbiosum</i> GluDH binary complex PDB accession number 1BVG
α-carboxyl group	i-side-chain NH2 of Gln114 3.0 Å ii- side-chain Lys117 3.17Å and 2.82Å iii- side-chain Lys129 2.80Å vi- side-chain NH2 of Asn349 2.87Å	i-side chain Gln110 3.3 Å, 3.59 Å ii- side chain Lys113 2.9 Å iii- Lys125 to water vi-no hydrogen bond Asn347
Diol	i- side-chain of Asp169 2.99 Å ii- main chain CO of Gly168 3.29 Å iii- side-chain CO of Gln114 3.10 Å vi- side-chain of Lys129 2.41Å v-main chain NH2 of Gly95 2.86Å	i- side-chain of Asp165 2.78 Å ii- main chain CO of Gly164 2.73 Å
α-amino		
γ-carboxyl group	i- side-chain of Lys93 2.73Å ii- side chain of Ser381 2.55Å iii- side-chain of Atg209 2.92 Å	i- side chain Lys89 2.56 Å ii- side chain of Ser380 2.64Å

Table 4.4. Showing the hydrogen bonds for the substrate interaction in the ternary complex of *M. smegmatis* GluDH and *C. symbiosum* GluDH binary complex (1BVG).

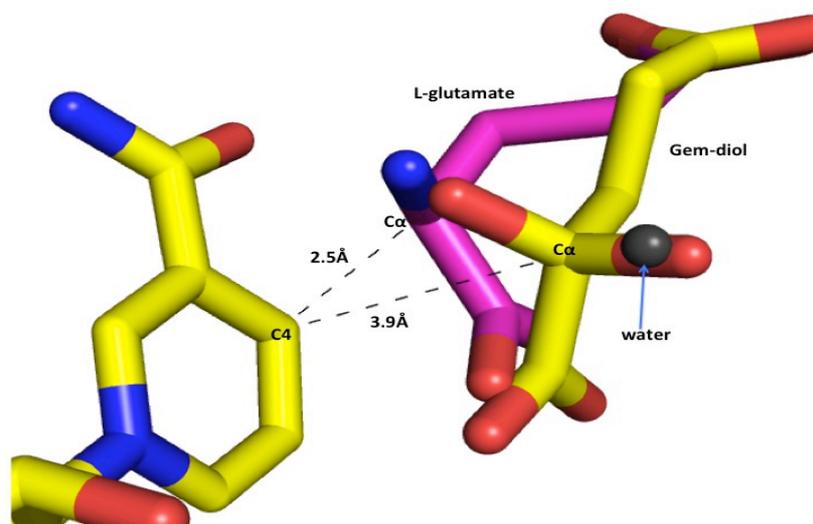


Figure 4.19. Shows the distance between the C-4 of the nicotinamide and the alpha carbon for the two substrates, L-glutamate and gem-diol. The α -carbon of the L-glutamate lies 2.5 Å from the C-4 of the nicotinamide, which allows the hydride transfer to occur.

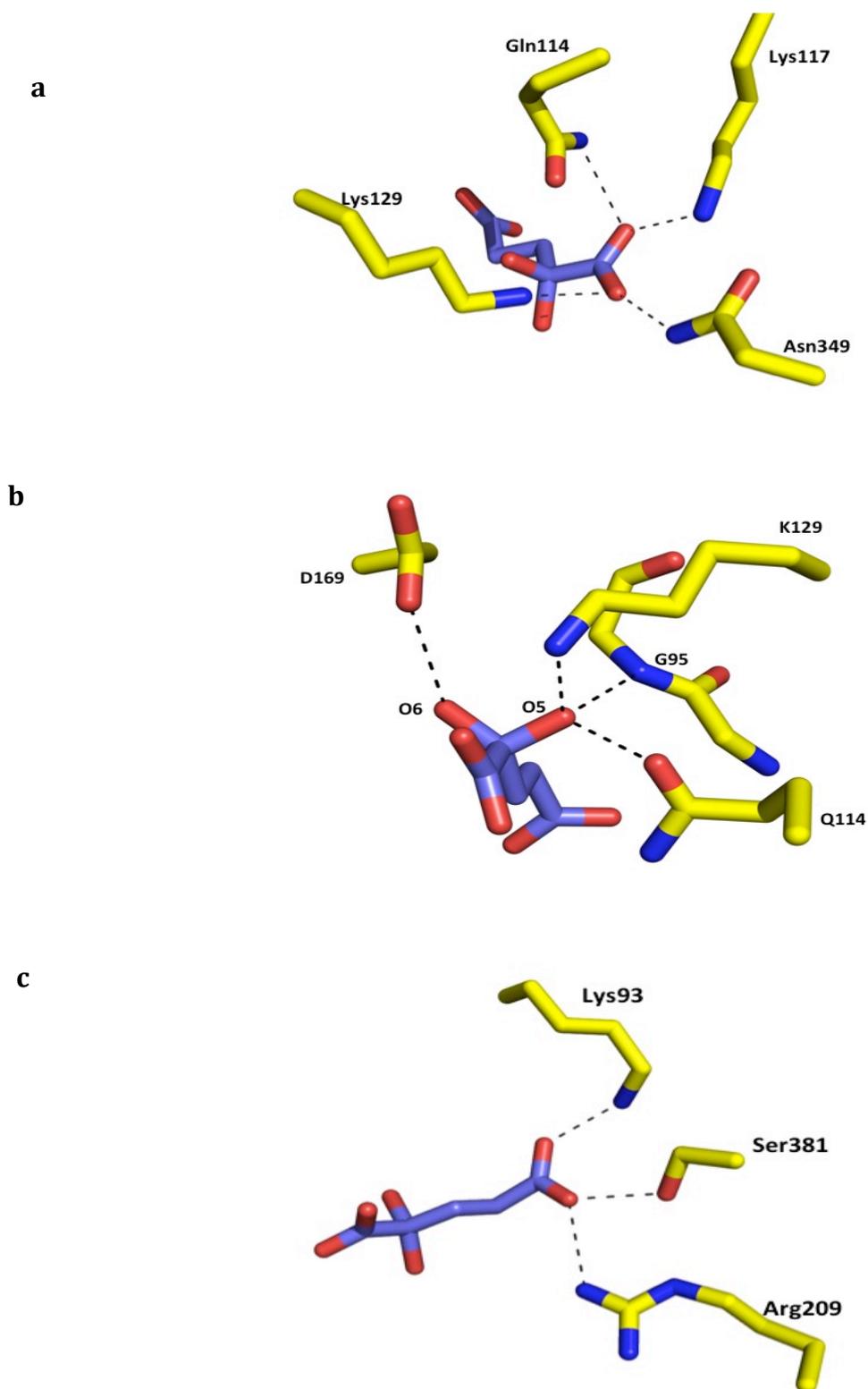


Figure 4.20. Interactions of possible diol with the adjacent groups. a) Binding of α -carboxyl group, diol and γ -carboxyl group to amino acid side chains of enzyme surface are described in pictures a, b and c respectively. The carbon atoms of the diol are shown in blue colour and of amino acids residues in yellow.

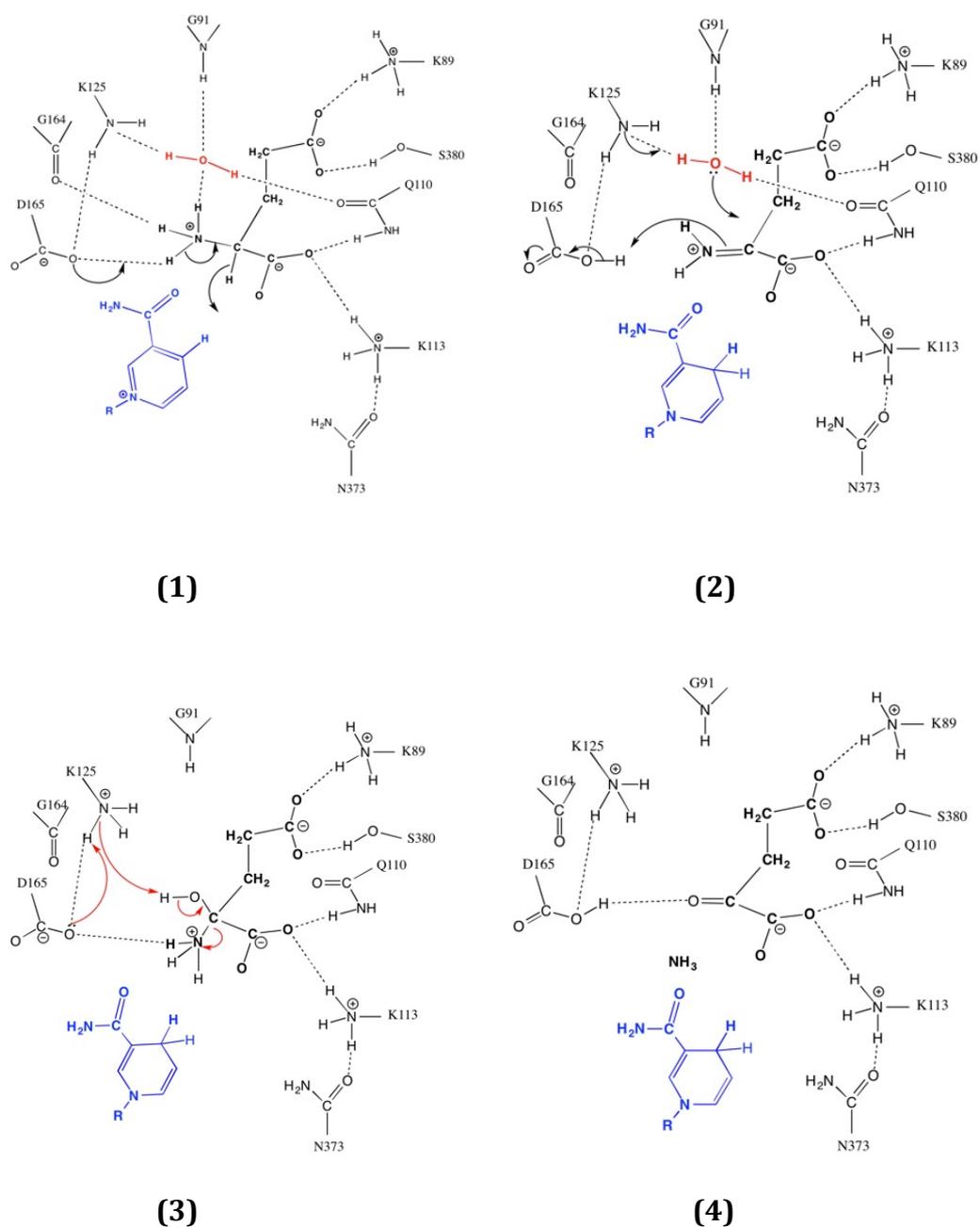


Figure 4.21. Revised chemical mechanism of GluDH in the direction of oxidative deamination of L-glutamate. Steps 1, 2 and 4 are similar to that proposed by Stillman (see figure 1.13) (Stillman et al, 1993). In the step 3 of this modified mechanism, D165 removes a proton from K125, which in turn, can act as a base to remove the hydroxyl proton of carbinolamine and thus form the ketone.

Chapter five: Purification, crystallization, data collection, structural determination and structural analysis of phenylalanine dehydrogenase from *Bacillus sphaericus*.

5.0 Introduction

This chapter describes the results for purification, crystallization, data collection, data processing, structural determination and structural analysis of phenylalanine dehydrogenase from *Bacillus sphaericus*.

5.1 Purification of phenylalanine dehydrogenase (PheDH)

5.1.1 Purification of single mutant (N145A) PheDH

Cell paste of *B. sphaericus* PheDH protein was provided by Prof Paul Engel; University College Dublin. The cell free extract was 26ml at a concentration 4.6mg/ml (total protein 120mg). 17ml of 4M-ammonium sulfate was added to bring the concentration to 1.5M. Precipitates were removed by centrifugation at 70000g for 10min. The supernatant fraction, containing 60mg of total protein was applied to a 5ml Phe-HP cartridge (GE Healthcare) for hydrophobic interaction chromatography. Proteins were eluted from the column by a reverse gradient of ammonium sulfate, from 1.5M to 0M in buffer A (50mM TRIS pH=8.0), 2.5ml fractions were collected.

All fractions were analysed for protein by the method of Bradford (Bradford, 1976) and fractions with the highest protein concentration (fr17-19) were combined, to give a total volume of 7.5ml at 3mg/ml. The volume of the sample was reduced to 1ml using a VivaSpin concentrator with a molecular weight cut off of 30kDa, and was then applied to a 16x60 Superdex200 gel filtration column, equilibrated in buffer A +0.5M NaCl. Gel filtration separation was performed with the same buffer at a flow rate at 1.5ml/min. 2ml fractions were collected after the void volume of the column had been discarded. Fractions were analysed by the method of Bradford and fractions 9-11, that had the highest protein concentration,

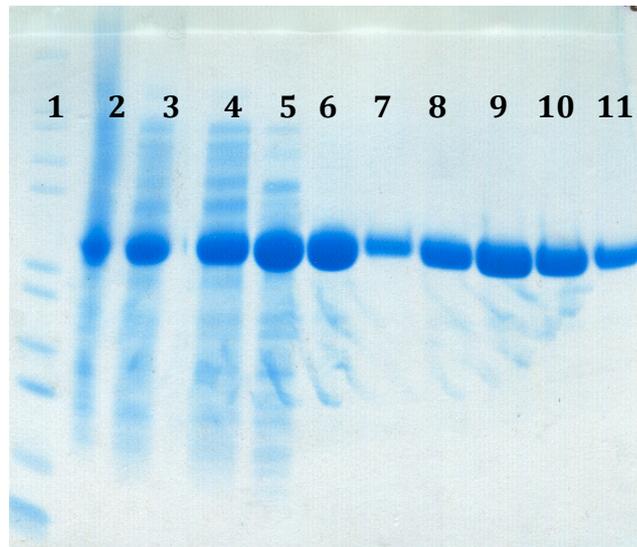
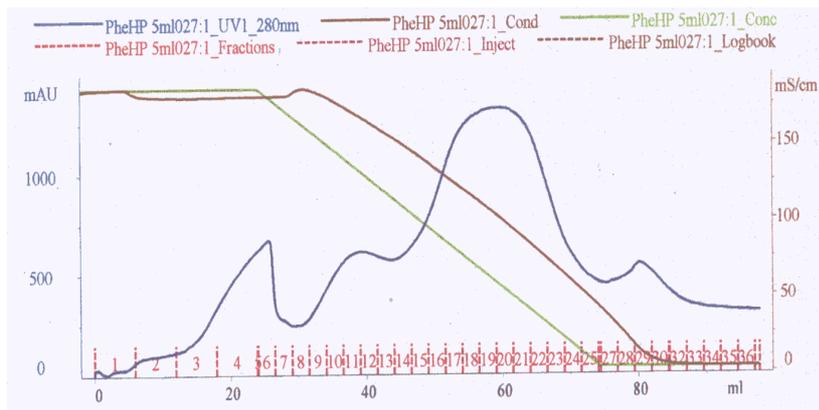


Figure 5.1. SDS-PAGE gel showing the purification progress of single mutant PheDH protein. The molecular weight of PheDH is 41.578 kDa. The lanes are arranged as following: 1= Marker 12, 2= cell debris, 3= cell free extract, 4= sample applied on a Phe-HP column, 5= sample applied on a gel filtration column, 6-11 fractions 8-12 from gel filtration column.

e



q

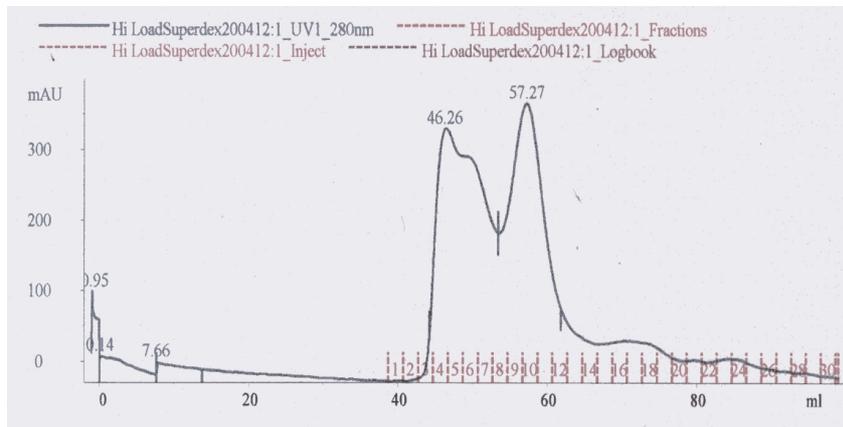


Figure 5.2. Chromatogram pictures for the purification of single mutant PheDH protein. **(a)** Hydrophobic chromatography (Phe-HP) purification stage showing the most active fractions 8-10. **(b)** Gel filtration purification step showing the peak of elutes at 57 ml.

were combined to give a total volume of 6ml at 1.5mg/ml. PheDH elutes from the column at 57ml, which corresponds to apparent MW of 370kDa, which suggests that PheDH is probably an octamer (subunit molecular weight=41.6 kDa).

The purification progress was analysed by SDS-PAGE using 4-12% Bis-Tris NOVAX gel and is shown in figure 5.1.

To prepare PheDH for crystallisation, a VivaSpin concentrator with a dialysis cup was used to change the buffer in the sample to 40mM Glycine-KOH pH8.0, 0.1M KCl and finally protein was concentrated to 10-16mg/ml.

5.1.2 Purification of wild type protein

Wild type *B. sphaericus* PheDH was purified the same way as the single mutant. 3g of cells were taken; 243 mg of total protein. After ammonium sulphate precipitation 205mg of total protein was applied to a Phe-HP cartridge. Fractions 15-17 were combined to give a total protein of 60mg, which was concentrated and applied to the gel filtration column. Fractions 7-9 were combined to give 21mg of purified PheDH. The samples for crystallization were prepared the same way as for the N145A protein.

5.2 Crystallization of *Bacillus sphaericus* PheDH

5.2.1 Crystallization of single mutant (N145A) PheDH with NAD⁺

The protein sample of the N145A mutant PheDH from *B. sphaericus* at 10 mg/ml was mixed with NAD⁺ (to a final concentration of the cofactor of 5mM) before placing the sample in crystallization trials (see section 2.3.1 for details). Four initial screens, PACT, JCSG, Classics and PEGs suites, were carried out using a hydra II plus one robot, by mixing 0.2 µl of the protein sample with 0.2 µl of crystallization solution in each well of the 96 conditions, prior to incubating the trays at 17°C. After 3 days, hits were found in the JCSG condition wells B8 and C5, and crystals appeared in the conditions of PHClear well G8 and JCSG well H7 (figure 5.5). In an attempt to get higher quality crystals, these conditions were optimized using the

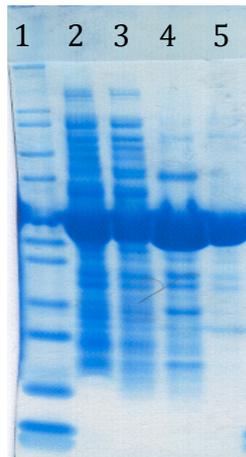


Figure 5.3. SDS-PAGE gel showing the purification steps of wild type PheDH (MW= 41.578 kDa). Lane 1 shows marker 12. Lane 2 shows cell free extract. Lane 3 shows sample applied on Phe-HP column. Lane 4 shows sample after Phe-HP. Lane 5 sample after gel filtration.

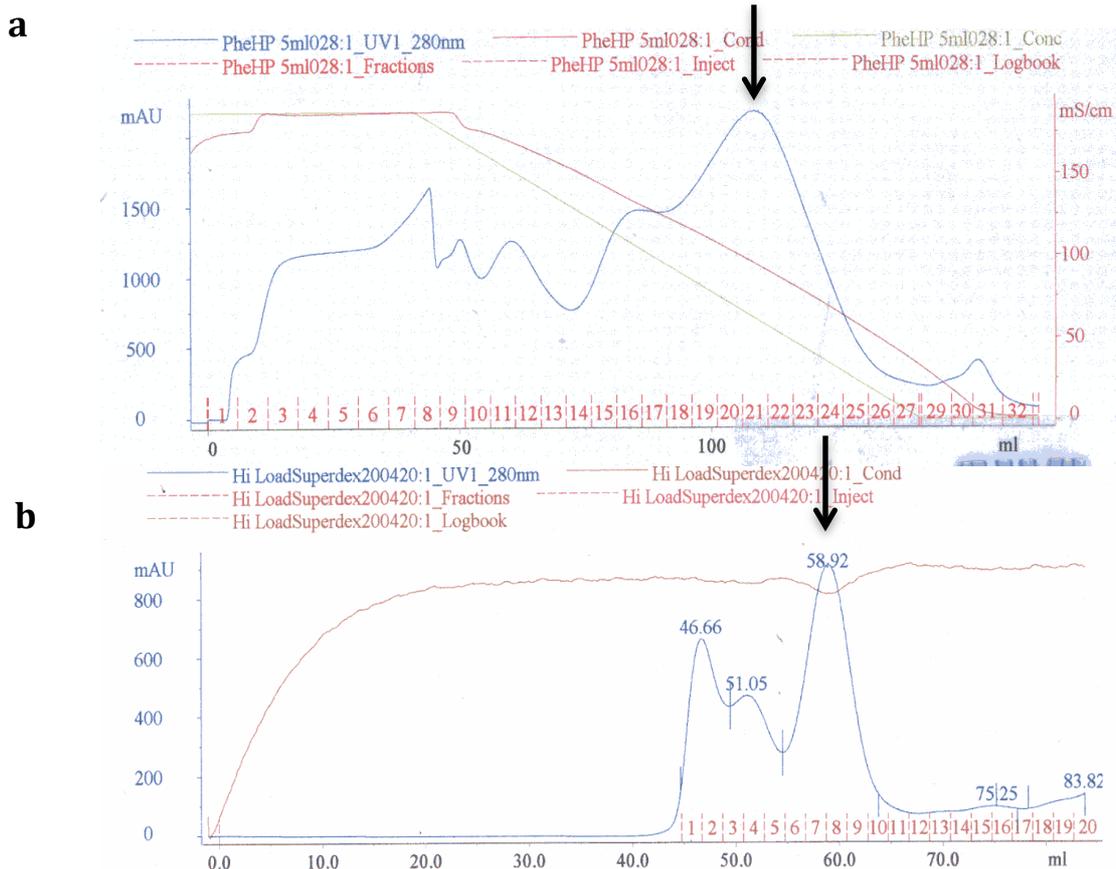
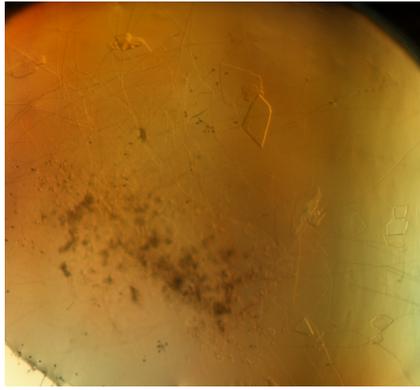
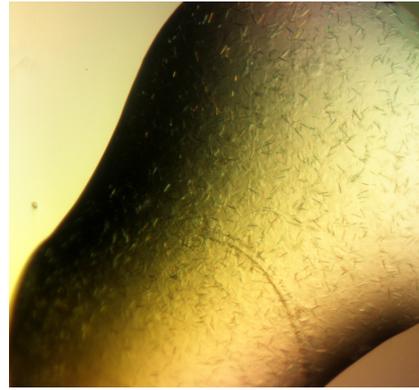


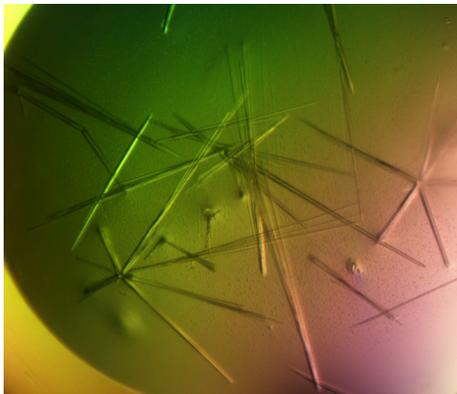
Figure 5.4. Chromatogram traces for the purification of wild type PheDH.(a) hydrophobic chromatography (Phe-HP) purification stage and (b) gel filtration purification step. Arrows indicate peak containing PheDH protein.



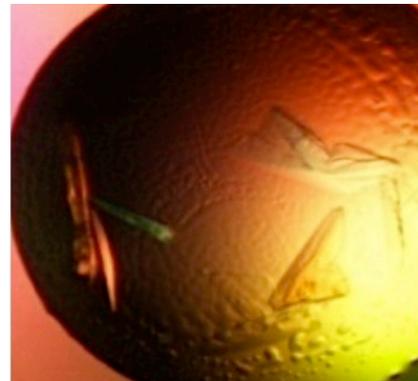
Classics B8
0.2M Magnesium Chloride
0.1M Tris pH 7.0
10% (w/v) PEG 8000



Classics C5
0.1M Sodium acetate pH4.6
2.M Ammonium sulphate



PHClear G8
0.1M Sodium acetate pH 5.0
20% (v/v) MPD



JCSG H7
0.2M Ammonium sulphate
0.1M Bis-Tris pH 5.5
25% (w/v) PEG3350

Figure 5.5. Photographs of Robot Screen crystals of single mutant *B. sphaericus* PheDH with NAD⁺.

hanging drop vapour diffusion method as described in section 2.3.2. For optimization the JCSG-H7 condition, the concentration of PEG3350 was altered from 15% to 30% and the pH also varied from 4.5 to 6.0 and the tray was incubated at 17°C. The pH was changed from 4.0 to 6.0 for optimization of PHClear B8. Unfortunately, no crystals grew in this trial.

5.2.2 Crystallization of wild-type PheDH with NADH

Co-crystallization of wild-type PheDH with cofactor was undertaken using the same method established above (section 5.1.1). A number of hits were found in the pHClear condition wells C2, C3 and H3, JCSG condition wells E10 and H4, PEG condition well C12 and Classics condition well B2 (Figure 5.6). Efforts were made to optimize the crystallization conditions for the wild type protein crystals by using a hanging drop vapour diffusion method as described in section 2.3.2. Optimization of the condition PHClear C3 was the only condition producing good quality crystals. In this experiment the concentration of isopropanol was varied from 1% to 15%. The best crystals were obtained under the conditions with 0.1 M MES pH=6.0 and 12% isopropanol, after two weeks of incubation at 17°C (figure 5.7b).

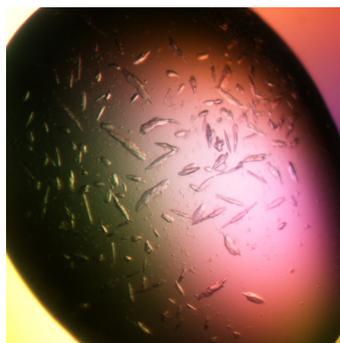
5.3 Data collection and processing of Phenylalanine Dehydrogenase from *Bacillus Sphaericus*

5.3.1 Single mutant (N145A) of phenylalanine dehydrogenase

Crystals of N145A mutant of PheDH with NAD⁺ from the JCSG-H7 robot screen were loop mounted and placed in a cryo-protectant solution consisting of 0.2M ammonium sulphate, 0.1M Bis-Tris pH=5.5, 25%(w/v) PEG3350 and 30% ethylene glycol prior to flash cooling in liquid nitrogen. Data were collected on beamline I02 at the Diamond Light Source (DLS). 120 images were recorded from one crystal with rotation angle of 0.75° and an exposure time of 1.8 s per image. The data were processed to a resolution of 3.1Å using Xia2 system in the 3dii mode (Winter, 2010) at DLS. The data were indexed to space group P4 with unit cell parameters $a=b=123.90$, $c=143.68\text{Å}$, $\alpha=\beta=\gamma=90^\circ$. Data-collection and processing statistics are shown in Table 5.1.



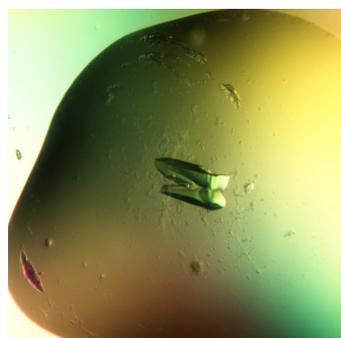
PEG C12
0.1 M MES pH 6.5
15% (w/v) PEG 20000



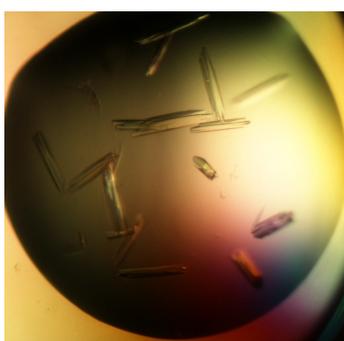
JCSG E10
0.1M BICINE Ph 9.0
10% (w/v) PEG6000



PHClear H3
0.1M MES pH 6.0
40% (v/v) MPD



PHClear C2
0.1M Citric acid pH 5.0
5% (w/v) PEG6000

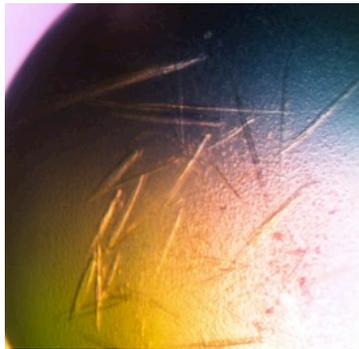


Classics B2
25% (v/v) Ethylene glycol

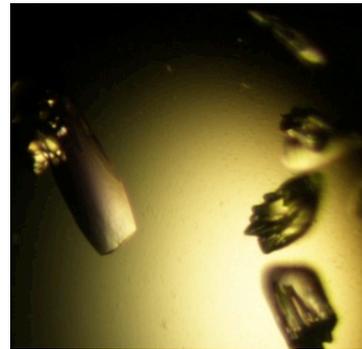


JCSG H4
0.2M Calcium chloride
0.1M Bis-Tris pH 5.5
45% (v/v) MPD

Figure 5.6. Pictures of robot screen crystals of wild-type *B. sphaericus* PheDH with NADH.



a



b

Figure 5.7. Photographs of crystals for wild-type *B. sphaericus* PheDH with NADH.

- a) pHClear C3- Robot Screen: 0.1M MES pH=6.0 and 5% Isopropanol; Drop size 0.2 μ l of protein + 0.2 μ l precipitant.
- b) pHClear C3- Hanging Drop: 0.1M MES pH=6.0 and 12% Isopropanol; drop size 1.0 μ l of protein sample + 1.0 μ l crystallization solution.

5.3.2 Binary complex of wild type PheDH with NADH

Crystals were harvested for data collection from the optimization experiment (0.1M MES pH=6.0 + 12% isopropanol) based on their size and overall definition. These crystals were transferred to a cryoprotectant solution containing the mother liquor and 30% ethylene glycol prior to flash cooling in liquid nitrogen. Initial diffraction analysis was performed using a Rigaku007 rotating anode generator and a MAR345 image plate. 3 images were collected with 5 min exposure time per image. One of these crystals diffracted to beyond 3.5Å resolution on the home source and so it was saved in liquid nitrogen and taken to the beamline IO4 of the DLS synchrotron, Oxford. A total 1350 images were collected from this crystal to the resolution of 2.5Å with a rotation angle of 0.15° and an exposure time of 0.5 s per image. The X-ray diffraction data were processed through Xia2 system using the 3dii mode (Winter, 2010) at the diamond light source. The crystal indexed to space group C2 with cell dimensions; a=215.6 Å, b=97.4Å, c=202.6, $\alpha = \gamma = 90^\circ$ and $\beta = 121.69^\circ$. The volume of the unit cell suggested that the asymmetric unit contained 8 subunits of the PheDH, to give a V_m of 2.73 Å³ Da⁻¹, which is in the range for Matthews Coefficients (Kantardjieff & Rupp, 2003). Data-collection and processing statistics are shown in Table 5.2.

5.4 Molecular replacement and model building

In order to get a suitable search model for molecular replacement, the amino acid sequence of PheDH from *B. sphaericus* was queried against the protein data bank using the Basic Local Alignment Search Tool (BLAST) (Altschul et al, 1997). The best homology was found with the leucine dehydrogenase from *B. sphaericus* (PDB accession number 1LEH), where the identity between two enzymes is about 50%. The octameric *B. sphaericus* LeuDH (no water) was trimmed to a tetramer using *Pdbset* in CCP4i and four subunits A, B, C and D in the form of the two dimers (A, B and C, D) were selected for an automated search through *Phaser* (McCoy et al, 2007) in CCP4i. As V_m calculations indicated that the asymmetric unit contained 8 subunits and the space group includes a 2-fold axis, there could

	Overall	Low	High
High resolution limit Å	3.14	14.04	3.14
Low resolution limit Å	123.91	123.91	3.22
Completeness %	99.8	98.9	100.0
Multiplicity	3.7	3.5	3.8
I/sigma	12.0	38.5	2.5
Rmerge	0.124	0.022	0.733
Rmeas(I)	0.174	0.032	0.981
Rmeas(I+/-)	0.166	0.031	0.984
Rpim(I)	0.088	0.017	0.491
Rpim(I+/-)	0.110	0.021	0.650
Wilson B factor	75.918		
Partial bias	0.000	0.000	0.000
Anomalous completeness	94.7	95.2	97.3
Anomalous multiplicity	1.9	1.9	1.9
Anomalous correlation	0.127	0.003	0.037
Anomalous slope	1.134	0.000	0.000
Total observations	141765	1585	10832
Total unique	37897	451	2828

Table 5.1. Data-collection and processing statistics for the single mutant PheDH crystallized with the cofactor.

Dataset	Over all	Low	High
High resolution limit Å	2.54	11.37	2.54
Low resolution limit Å	85.81	85.81	2.61
Completeness %	98.9	99.5	96.3
Multiplicity	3.9	3.9	3.5
I/sigma	14.0	68.5	1.4
Rmerge	0.074	0.017	0.860
Rmeas (I)	0.102	0.032	1.176
Rmeas (I+/-)	0.099	0.023	1.162
Rpim (I)	0.050	0.018	0.612
Rpim (I+/-)	0.066	0.015	0.776
Wilson B factor	61.752		
Anomalous completeness	95.3	96.5	90.6
Anomalous multiplicity	1.9	2.1	1.7
Anomalous correlation	0.072	0.314	0.015
Anomalous slope	1.058	0.000	0.000
Total observations	4522 22	5463	2882 5
Total unique	1157 51	1397	8279

Table 5.2. Data-collection and processing statistics for the binary complex of the wild-type *B. sphaericus* PheDH with NADH.

be either one complete octamer of PheDH or two half-octamers in the asymmetric unit. Therefore the X-ray dataset was entered and *Phaser* was run in the C2 space group, searching for two copies, of the tetramer. *Phaser* found a solution with a Z-score 15.12 for rotation and 20.15 for translation for tetramer one and the Z-score of rotation function=12.1 and translation function=53.2 for tetramer two with an Rfactor=0.40 and Rfree=0.47 after refinement of the resultant model by REFMAC5 (Murshudov et al, 2011). This solution from *Phaser* gave two half-octamers in the asymmetric unit, with the crystallographic two-fold being used to assemble the octameric PheDH oligomer, suggesting a correct solution. However, inspection of the resultant structure and the electron density map for all subunits showed that the electron density was poor for this initial model. Therefore, to get the C-terminal domains correct in the density map for all subunits, the model was subjected to a rigid body refinement using REFMAC5 (Murshudov et al, 2011) in the CCP4i suite at 6.0Å and 2.6Å. The Phenix software suite was used to fit loops that were disordered, before restrained refinement using REFMAC5 (Murshudov et al, 2011). The resultant electron density map was checked for this model and it was found that the electron density for domain I was better than domain II for the whole structure. Examining the electron density within the active site for all subunits showed that there was a positive different peak associated with the nucleotide-binding site of all 8 subunits, which could be assigned to the cofactor. Different features of this density could be seen for the eight chains in the asymmetric unit, but the best density was linked with the chain D. Thus, the D subunit was used for rebuilding and the sequence of the *B. sphaericus* LeuDH search model was changed to the *B. sphaericus* PheDH sequence. All disordered residues were removed from the model, including the first 11 residues and the loops at the residues 145-155 and 223 to 234. NADH was then modeled within the electron density in the active site. The rebuilt subunit D was superimposed onto other subunits using secondary structure matching superpose (SSM superpose) through COOT (Emsley et al, 2010). Several cycles of model building and refinement were then

conducted using COOT (Emsley et al, 2010) and REFMAC5 (Murshudov et al, 2011). There are a number of areas in the electron density, which are of poor quality and thus building a model with good geometry is difficult. Nevertheless, the electron density around the active site and the NADH binding site is quite good and some useful points can be made about the structure.

5.5 Analysis of the *B. sphaericus* PheDH structure

B. sphaericus PheDH is an octameric enzyme that consists of eight subunits of 381 amino acid residues, organized into 42-point group symmetry. The polypeptide chain is formed from 13 α -helices and 10 β -sheets with 163 and 46 residues in these secondary structures, respectively, which represents \approx 55% of the whole subunit. In *B. sphaericus* LeuDH the monomer contains 14 α -helices and 12 β -strands, which amounts to 75% of the subunit (Baker et al, 1995).

5.5.1 Tertiary structure of PheDH subunit

Like GluDH the PheDH monomer comprises two domains, linked by flexible loops, and separated by a deep groove in which the substrates bind. Domain I is relatively larger than domain II and it consists of amino acid residues 1-140 and 344-380. The polypeptide chain in this domain is organized into a cluster of 4 α -helices and 4 β -strands. The secondary structure units begin with a short helix (residues 16-20), followed directly by three segments of antiparallel β -strands. The remaining residues are organized into two α -helices, one β -strand and one α -helix. This domain is comprised of four parallel/antiparallel β -strands, surrounded by four α -helices on both sides. Domain II or C-terminal domain (residues 41-343), is similar to the Rossmann fold and contains five parallel β -strands (β e- β h) flanked by seven α -helices (α 5- α 9). One domain of the binary complex of *B. sphaericus* PheDH is illustrated in figure 5.8a, and compared to the GluDH in figure 5.8b.

5.5.2 Quaternary structure

Two subunits of PheDH assemble into a dimer, in a similar way to the

```

PheDH 1 -----MAKQLEKSSKIGNEDVFQ--KIANHEQIVFCNDPV 33
LeuDh 1 -----MEIFKYMEKYDYEQLVFCQDEA 22
msgdh 1 MSELHPKLQAVFEVSHRNGGEKEFHQAVFEVLQSLGPPVVEKHPEYADNAIR--RLCEPERQIIFRVPW 68

                hhhhhh      hhhhh  hh  eeeeeeeee

PheDH 34 -----SGLQAIIAIHDTTLGPALGGTRMYPYKNVDEALEDLRLSEGPTYKCAAADIDFGGGKAVIIG 96
LeuDh 23 -----SGLKAVIAIHDTTLGPALGGARMWNTYNAEEEEIEDALRLARGMTYKNAAGLNLGGGKTVIIG 85
msgdh 69 VDDTGHVQINRGRFVFNALGPYKGGRLFHPSVY----LGIVKFLGFEQIFKNALTGLPIGGGKGGSD 134

                eeeeeeeee      eee      hhhhhhhhhhhhhhhhhhhhh      eeeee

PheDH 97 DPEKDKSPALFRAFQGFVESLN-----GRFYTGDMGTTMDDFVHAQKET-----NFINGIPEQYG 152
LeuDh 86 DPFADKNEDMFRALGRFIQGLN-----GRYITAEDVGTVDMDLIHQET-----DYVTGISPAFG 141
msgdh 135 DPKGRSDAEIMRQCQSFMTELYRHIGEYDVPAGDIGVGTREIGYLFQYKRI TNRYESGLT GKGLTWG 204

                hhhhhhhhhhhhhhhhhhh      eee      hhhhhhhhhhh      ee  hh

PheDH 153 GSGDSSIPTAQGVIIYALKATNQYLFSDSLSGKTYAIQGLGKVGKYVAEQLLKAGADLFV-TDIHENVLN 221
LeuDh 142 SSGNPSPVTA YGVYRGMKAAAKEAFGSDSLEGLAVSVQGLGNVAKALCKKLNTEGAKLVV-TDVNKAAS 210
msgdh 205 GSRVRTEATGYGTVFFVDEILQSRGQ--SFDGKRVVVSGSNVAIYAIEKVHALGGIVVACSDSGVYVD 272

                hhhhhhhhhhhhhhhhhhh      eeeee      hhhhhhhhhhh      eeee ee  hh

PheDH 222 SIKQKSEELG-----GSVTI VKSDDIYSVQADIFVPCAMGGIINDKTIPKL---KVK 270
LeuDh 211 AAVAE-----EGADAVAPNAIYGVTCDFAPCALGAVLNDFTIPQL---KAK 254
msgdh 273 EKGIDL DLLKEVKEVQRARIDAYAEARGGATQFVSGGSVMNVACDIALPCATQNEVSGDDARALIESGCR 342

                hhhhhhhhhhh      ee      eeee      hhhhhh      e

PheDH 271 AVVGSANNQLKDLRHNANVLEKGI LYAPDYIVNAGGLIQVADELYGPNKERVLLKTKEIYRSLEIFNQA 340
LeuDh 255 VIAGSADNQLKDRHGKYLHELGI VYAPDYVINAGGVINVADELYGYNRTRANKRVDGIYDSIEKIFAIS 324
msgdh 343 IVAEGANMP-CTPEAVKLFDAQGVTFAPGKAVNAGGVATSALEMQ-NASRDSWTFDDTEGRLAGIMRRI 410

                eeee      hhhhhhhhh      eee  hhhh  hhhhhhhhhhh      hh  hhhhhhhhhhhhhhhhh

PheDH 341 ALDCITVEAA-----NRKCQKTIEGQOTRNSFFSRGRRPKWNIKE-- 381
LeuDh 325 KR DGVPSYVAA-----DRMAEERIAKAVAKARSQFLQDQR---NILNGR 364
msgdh 411 HDRCLTTADEYDQPGNYVAGANIAGFTQVADAMLALGLI----- 449

                hhh  hhhhh      hhhhhhhhhhhhhhhhhhhhh

```

Figure 5.8. Structural sequence alignment of the *B. sphaericus* PheDH, *B. sphaericus* LeuDh and *M. smegmatis* GluDH. Secondary structure segments are shown under the sequence. This figure was generated using PROMALS3D (Pei et al, 2008).

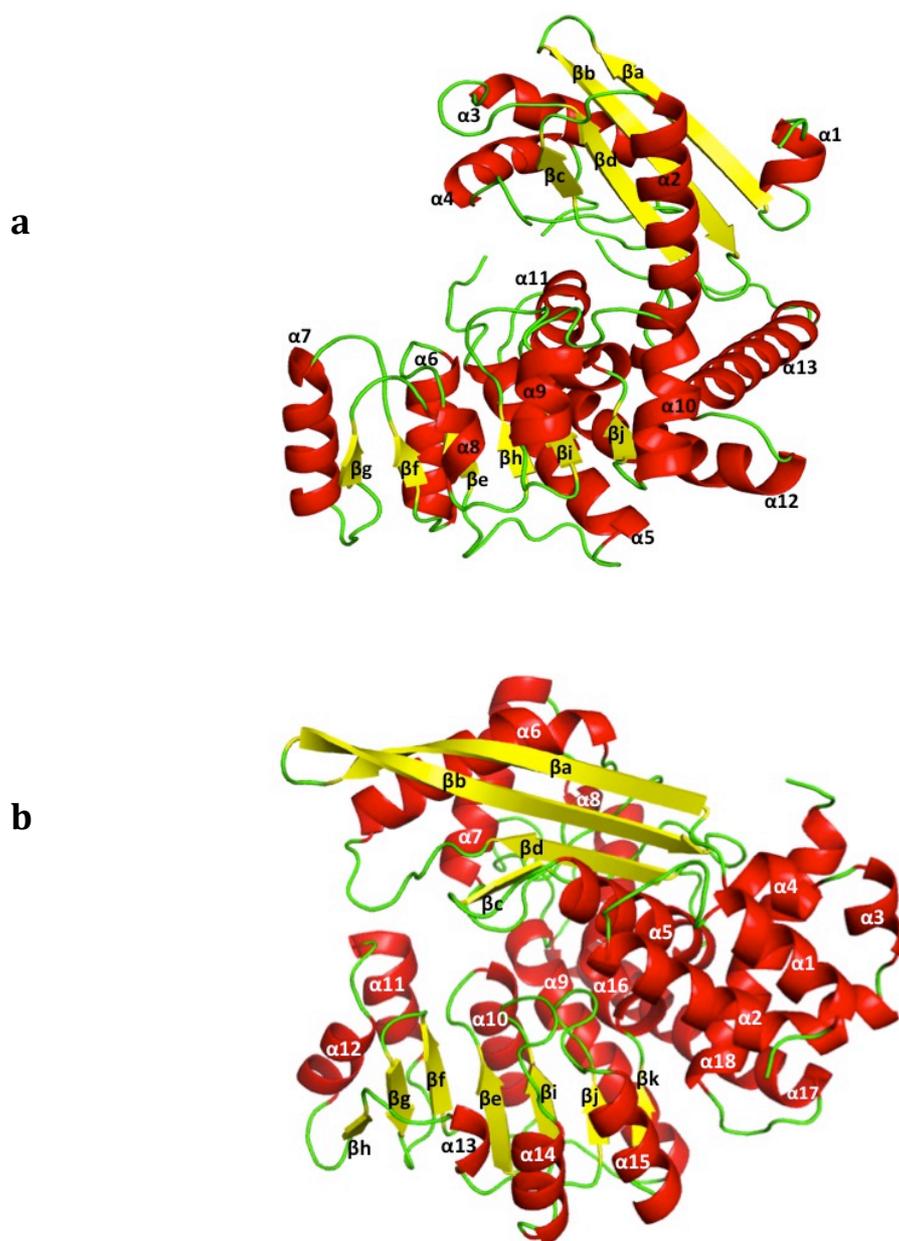


Figure 5.9. Schematic diagram shows the overall organization of one chain of *B. sphaericus* PheDH (a) and *M. smegmatis* GluDH (b). The PheDH subunit is viewed in the four-fold axis and the chain of GluDH is viewed in the three-fold axis. The single subunits of both enzymes are organized into two domains separated by a deep cleft. In PheDH, the α -helices numbered from $\alpha 1$ - $\alpha 13$ and the strands numbered βa - βj , whereas in GluDH the α -helices labeled from $\alpha 1$ - $\alpha 18$ and the strands labeled βa - βk .

dimer assembly in GluDH. However, the four dimers of *B. sphaericus* PheDH assemble around the four-fold axis to give an octamer with 4₂-symmetry compared to the GluDH hexamer. The main contacts between these monomers arise from the amino acid residues in the N-terminal domain, where the interactions through the twofold axis make the dimers and the contacts across the fourfold axis form the octamer. The domains I are packed in the centre of the oligomer and the closest to the 4₂-symmetry point. N- and C-terminal domains in each monomer meet on top of one another in the side of the fourfold axis. The PheDH octamer is very similar to the *B. sphaericus* LeuDH octamer (figure 5.10).

5.5.3 Dinucleotide binding site

All phenylalanine dehydrogenases use NAD⁺ as a cofactor in their reactions (Brunhuber & Blanchard, 1994). Although the structure for the binary complex of *B. sphaericus* PheDH with NADH has been determined only to medium resolution (2.5Å), the nature of the interactions between the dinucleotide and this enzyme can be determined. The PheDH from *B. sphaericus* displays substantial similarity for the secondary structure and fold of the dinucleotide-binding domain to the other amino acid dehydrogenases GluDH and LeuDH. The interaction between the PheDH and the cofactor (NADH) is also similar. In addition, primary structural analysis for this binary complex shows that the dinucleotide-binding region includes βe-α6-βf and that the amino acid sequence of the glycine rich loop, which is connected the end of βe and the start of α6, is GLGKVG (P1-P6). From this sequence it can be noted that the residue at the P6 is glycine-196, in agreement with general pattern of glycine at P6 in most NAD⁺-dependent amino acid dehydrogenases.

5.5.3.1 Mode of binding of NADH with *B. sphaericus* PheDH enzyme

As mentioned above, the residue at the P6 position is glycine which mean a type I hydrogen bond can be formed in this complex (see section 5.3.2 for details about type I hydrogen bond)(Baker et al, 1992c). The carbonyl group of G191 (P1) points towards the α-carbon of G196 (P6). The side-

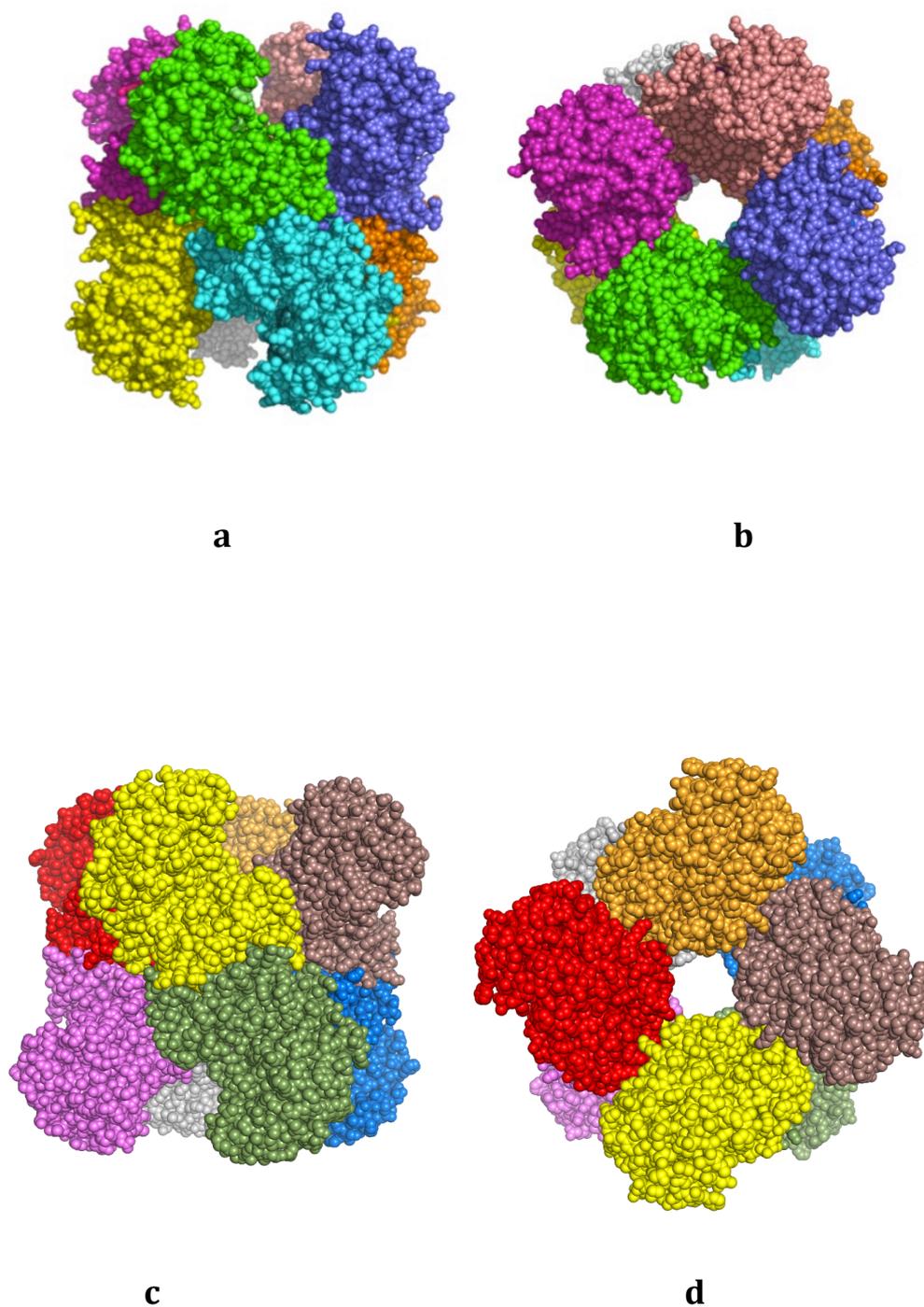


Figure 5.10. Space filling illustration of the octameric structures, of PheDH (a and b) and LeuDH (c and d) from *B. sphaericus*, with each subunit individually coloured. a and c view down the two-fold axis applied to form the dimer, and the four-fold axis being perpendicular. b and d view down the four-fold axis.

chain carboxyl group of the acidic residue at P7 (Asp214) forms two hydrogen bonds with the 2' and 3' hydroxyl groups of adenine ribose, and also makes a hydrogen bond to the backbone amide nitrogen of L192 (P2). The adenine ring lies in the pocket lined by Gln190, Ile215, Ala254 and Met255. In this binary complex the side chain sulphur atom of Met255 is within 3.03 Å of the NH₂ of the adenine ring and could perhaps form a hydrogen bond. As has been seen in the ternary complex of *M. smegmatis* GluDH (see section 5.3.4 for details) the pyrophosphate moiety of the NADH of the *B. sphaericus* PheDH binary complex lies at the end of the dinucleotide binding helix ($\alpha 6$), which aids in the stabilization of the phosphate group with its helix dipole. An oxygen atom of the adenine phosphate (O2A) interacts with the backbone amide nitrogen of Lys194 by a hydrogen bond (3.2Å). The main N-H of Val195 hydrogen bonds to the O2N of the pyrophosphate moiety, and O1N forms a hydrogen bond with the N-H of the nicotinamide moiety, (2.76Å).

By examining the electron density map corresponding to the cofactor (NADH) it can be seen that the nicotinamide moiety adopts the *syn* conformation for the glycosidic bond between the nicotinamide moiety and linked ribose as is seen in GluDH. As a result of this conformation, the oxygen atom of carbonyl group of nicotinamide ring makes a hydrogen bond (2.76Å) to the O1N of pyrophosphate. In addition, two hydrogen bonds can be made between the amide group of the nicotinamide moiety and the side-chain hydroxyl group and backbone amide group of Ser157, where the lengths of these bonds are 2.48Å and 2.42Å respectively. Furthermore, the 2'hydroxyl group of the nicotinamide ribose forms a hydrogen bond (2.79Å) to the side-chain amide group of Asn277 and the 3'hydroxyl group hydrogen bonds (2.53Å) to the backbone carbonyl group of Ala254.

5.5.4 Domain motion

As has been described at length in chapter 4 (section4.5), domain motion during the catalytic cycle is a feature of the amino acid dehydrogenase superfamily. In addition to the movement described for GluDH, X-ray

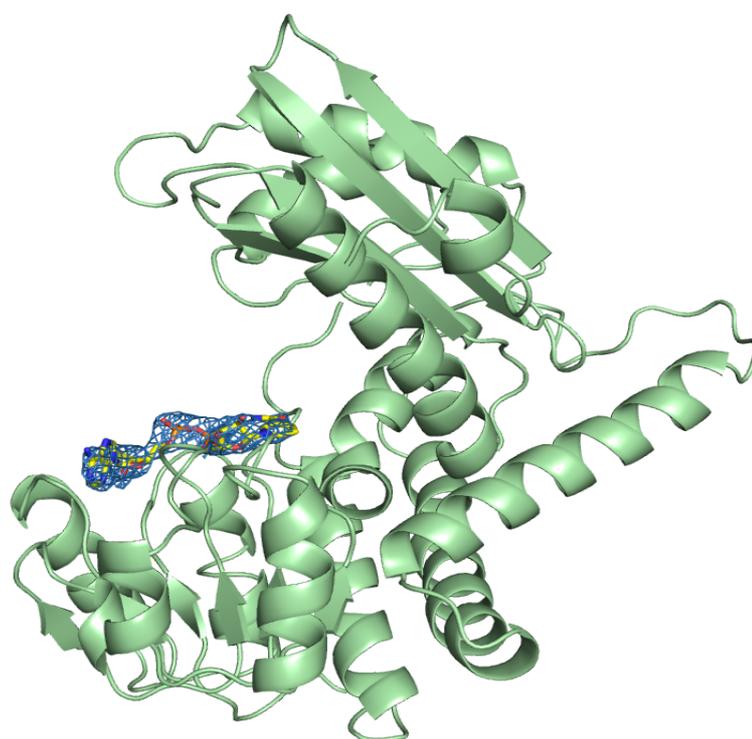


Figure 5.11. A schematic diagram of one subunit of *B. sphaericus* PheDH showing the cofactor location.

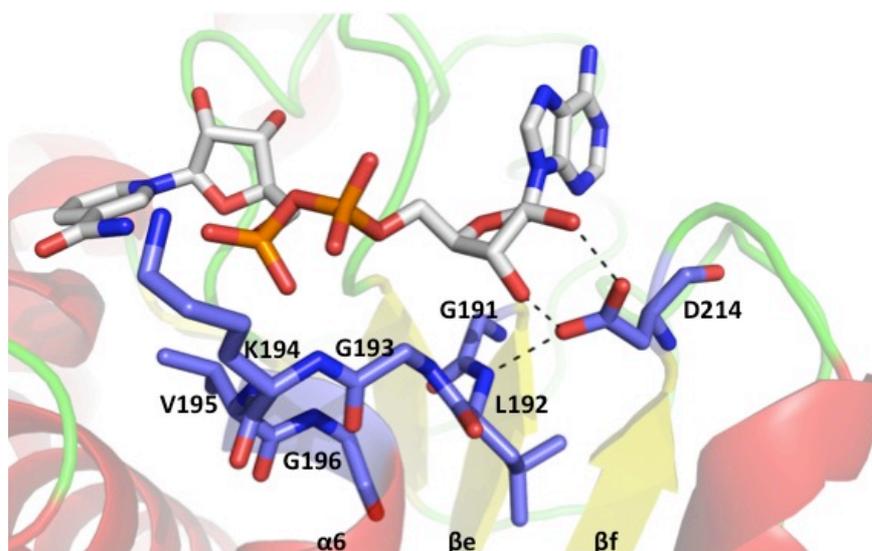


Figure 5.12. A schematic depicts the $\beta\alpha\beta$ motif (βe - $\alpha 6$ - βf). P1-P6 residues that connect the βe and the $\alpha 6$ are shown in blue colour (G91-G96) and D214 is an acidic residue P7. Type I hydrogen bond can be seen between the side chain of D214 (P7) and NH2 main chain of L192 (P2), and 2' and 3' hydroxyl groups of adenine ribose.

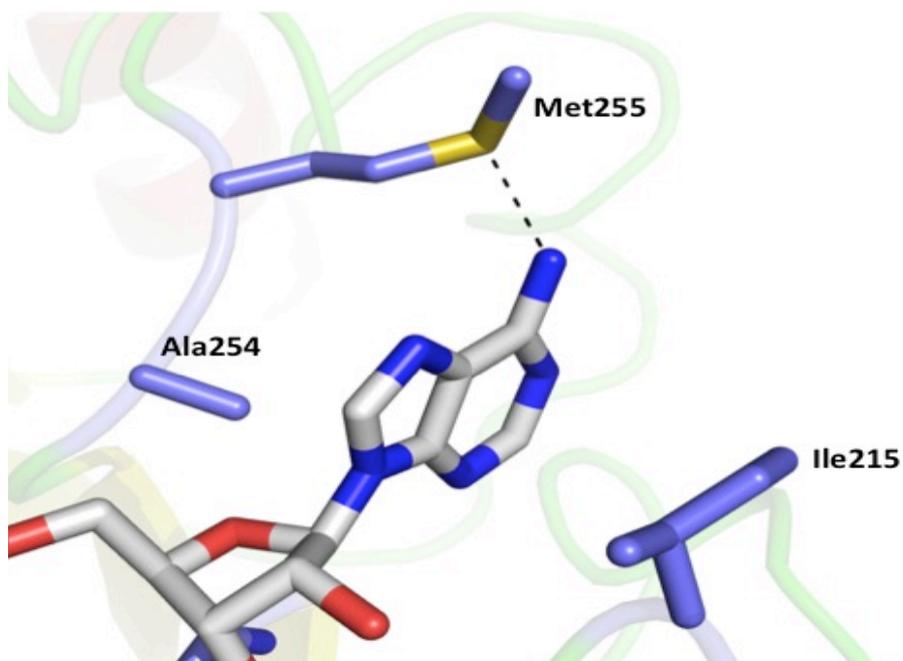


Figure 5.13. A schematic diagram of an interaction of adenine ring in the binary complex structure of *B. sphaericus* PheDH.

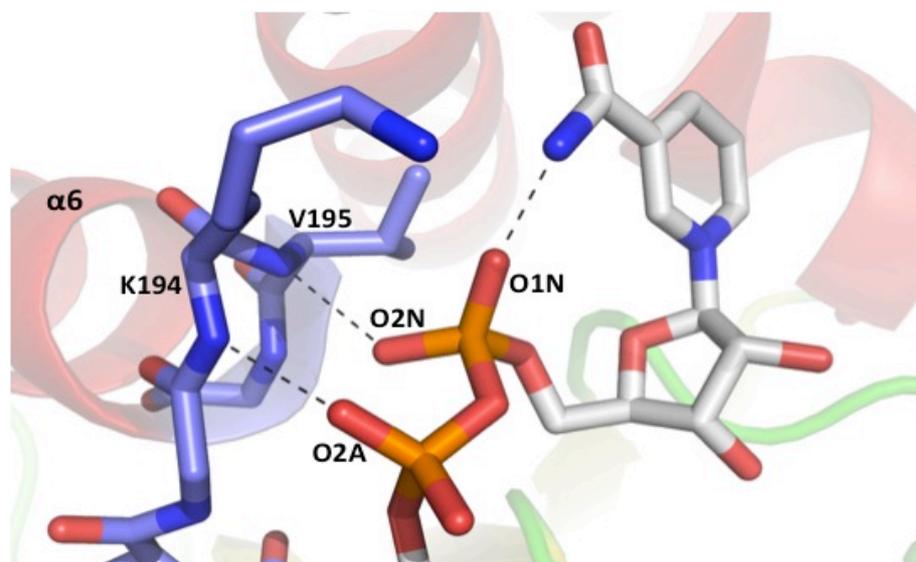


Figure 5.14. Schematic illustrating the pyrophosphate binding with the adjacent groups. Two oxygen atoms (O2A and O2N) make hydrogen bonds to the glycine rich loop and the O1N forms a hydrogen bond to NH₂ of nicotinamide.

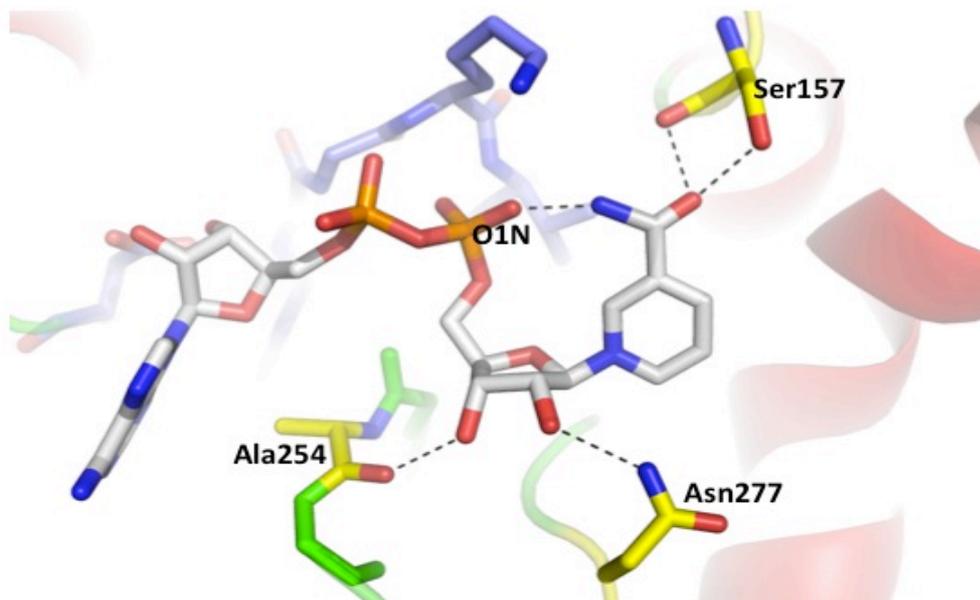


Figure 5.15. The interaction of nicotinamide ring and associated ribose to the neighbouring groups in the binary complex structure of *B. sphaericus* PheDH with NADH.

NADH	Hydrogen bonds of NADH in <i>B. sphaericus</i> PheDH binary complex
Nicotinamide ring	
Carbonyl group	i- Side-chain of Ser157 2.48 Å ii- NH ₂ -main chain Ser157 2.42 Å
Amino group	i- Oxygen of pyrophosphate 2.76 Å
Nicotinamide ribose	
2'OH	i- NH ₂ side-chain of Asn277 2.79 Å
3'OH	i- Main chain CO of Ala254 2.53 Å
Pyrophosphate	
O1N	i- NH ₂ of carboxyamide 2.76 Å
O2N	i- Main chain NH ₂ of Val195 2.74 Å
O2A	ii- Main chain NH ₂ of Lys194 3.2 Å
Adenine ribose	
2'OH	i- Side chain of Asp214 2.83 Å

Table 5.3. Shows hydrogen bonds between NADH and *B. sphaericus* PheDH.

structures of PheDH from *Rhodococcus sp.* M4 have been determined for a number of cofactor/ or substrate or inhibitor ternary complexes. In each one of these structures the PheDH enzyme was in a closed conformation (Brunhuber et al, 2000; Vanhooke et al, 1999). A number of structures of *Nocardia sp.* 239 PheDH have also been determined (Turnbull, A. P *et al.*, personal communication). In these structures the free enzyme and the binary complex with NADH adopt the open form with the structure of the non-productive complex (PheDH/NAD⁺/Phenylpyruvate) representing a closed form for this protein (Turnbull, A. P *et al.*, personal communication). The structure of *B. sphaericus* LeuDh has been reported as the apo enzyme, and the crystal form of this protein has a dimer in the asymmetric unit (Baker et al, 1995). Superimposing the two subunits showed that there is a variation in the relative orientations of domains I and II in the two subunits, with one subunit in an open form, and the other in a conformation between the open and the closed form. Analysis of the crystal structure for the binary complex (PheDH-NADH) of *B. sphaericus* PheDH determined here shows that this complex adopts the open conformation, which is similar to other superfamily members (figure 5.10), and again shows that the domain positions are not dependent on the presence of substrates.

5.5.5 Active site

Although no substrate is present in the *B. sphaericus* PheDH structure described here, phenylalanine can be modelled into the active site, based on the position of L-glutamate in GluDH, and L-phenylalanine in *Rhodococcus sp.* M4 PheDH (figure 5.15). These models, in conjunction with the sequence alignment, (figure 5.16 and table 5.4) show conservation of the crucial residues involved in the active site (Britton et al, 1993b). For example; five glycines (residues 90, 91, 122, 123 and 376) that are critical in the active site design in GluDH are conserved in *B. sphaericus* PheDH (glycine residues 52, 53, 88, 89 and 305). The lysine residue (K113) that interacts with the α -carboxyl group of L-glutamate in GluDH, lysine125 (K125) that enhances the nucleophilicity of the water molecule and

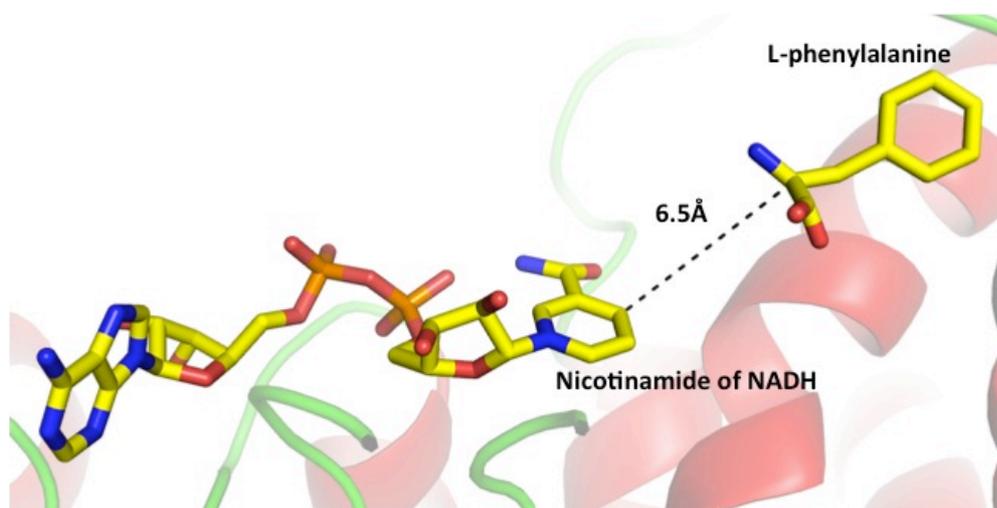


Figure 5.16. Shows the distance between the C4 of the nicotinamide and the alpha carbon of the modelled L-phenylalanine. The distance between these contributors involved in hydride transfer is about 6.5Å, which is incompatible with catalysis.

aspartate165 (D165), which is involved in proton transfer in the catalytic mechanism (Stillman et al, 1993), are also conserved in *B. sphaericus* PheDH (residues K79, K91 and D133 respectively). Conservation of these crucial residues between the GluDH, LeuDH and PheDH enzymes suggests that they share a similar catalytic mechanism. In contrast the residues involved in the main interactions with the substrate vary between GluDH, LeuDH and PheDH. In *C. symbiosum* GluDH, the γ -carboxyl group of L-glutamate interacts with K89 and S380, which are substituted to L51 and V310 in PheDH from *B. sphaericus*, but in *Rhodococcus sp. M4* PheDH are A38 and L296, respectively. These changes make the substrate-binding site more hydrophobic for the phenylalanine substrate. In the *Rhodococcus sp. M4* PheDH ternary complex A38 and L296 make direct interactions with phenylalanine (3.8Å and 3.9 Å, respectively) and phenylpyruvate (3.8 Å for A38 and 4.20Å for L296). In the binary and ternary complex structures of PheDH from *Nocardia sp239*, described in section 5.3.4, the interactions between the phenylalanine and phenylpyruvate with the enzyme are similar, with distances of 4.6Å and 4.9Å, to Ala31 and Leu285, for phenylalanine, respectively. For the phenylpyruvate these distances are 4.4Å and 5.3Å, respectively. This suggests that the substrate will bind to *B. sphaericus* PheDH in a similar way. In the amino acid dehydrogenases (AADH) the distance between the α -carbon of the modeled substrate and the C4 of nicotinamide moiety of NADH (the contributors implicated in hydride transfer) is about 6.5Å; this distance is incompatible with catalysis. It is likely that the active site is closed by moving the domain II, excluding the solvent from catalytic pocket and bringing the groups involved in hydride transfer into a suitable position for catalysis.

The structure of *B. sphaericus* PheDH described in this thesis has been determined only at medium resolution (2.54 Å) and has a number of loops missing. Furthermore, only the binary complex of PheDH with NADH was solved in this project. Therefore, the information that could be derived from this structure was limited. A full understanding of the structure/function relationship in the PheDH enzyme family needs further work to produce better diffracting crystals of the complex determined, and

also crystals of other binary and ternary complexes of both the wild-type and mutant PheDH enzymes are required.

5.5.6 Mutant of *B. sphaericus* PheDH

Numerous attempts have been to vary substrate specificity in *B. sphaericus* PheDH using either site-directed mutagenesis or random directed mutagenesis (Chen & Engel, 2009; Seah et al, 2003). These experiments have been designed and explained based on either the *Rhodococcus sp.* M4 PheDH structure, or modelled on other AADH structures for example *C. symbiosum* GluDH. Although the *B. sphaericus* PheDH structure determined here is only medium resolution some structural analysis into these mutant experiments can be made.

5.5.6.1 Single mutant PheDH (N145A)

As PheDH from *B. sphaericus* is relatively stable it is a candidate for use in the design of a biosensor for measuring the serum levels of phenylalanine, that are elevated in the inherited disorder disease called Phenylketonuria (PKU) (Dooley, 1992). However, the activity of *B. sphaericus* PheDH towards L-phenylalanine and L-tyrosine is quite similar, which makes this enzyme inappropriate for use in monitoring the levels of L-phenylalanine in diagnosis the Phenylketonuria (PKU). Nevertheless, homology modeling of PheDH from a variety of sources, in the light of the determined structures of the homologous enzymes LeuDH from *B. sphaericus* and GluDH from *C. symbiosum* have shown that an asparagine residue at site 145 in *B. sphaericus* PheDH is equivalent to a valine or alanine in PheDHs from other species. This has been suggested to be the key reason for the poor differentiation by *B. sphaericus* PheDH between L-phenylalanine and L-tyrosine. Substitution of alanine or valine instead of the asparagine (N145), using site directed mutagenesis, increases discrimination by the enzyme, to about 50-fold, between L-phenylalanine and L-tyrosine (Seah et al, 2002). To better understand the effect of this mutation, a tyrosine substrate molecule was modeled into the *B. sphaericus* PheDH structure, using the coordinates of the Norcadia PheDH/NADH/Phe ternary complex structure (provided by Dr A.P Turnbull) as a guide (figure 5.18). In this

modeled structure, it can be seen that the side chain hydroxyl of the putative tyrosine substrate can hydrogen bond to the carboxamide group of the side chain of N145. In the N145A mutant enzyme, this potential hydrogen bond between the tyrosine substrate and the enzyme is lost, perhaps explaining why the *B. sphaericus* N145A mutant PheDH shows a much increased discrimination between Phe and Tyr as the substrate, compared to the wild type.

5.5.6.2 Triple mutant proteins N145A, G124A and L307V

A number of site directed mutants of PheDH have been produced to engineer the protein to produce non-natural amino acids. For example; the triple mutant N145A, G124A and L307V shows increased activity towards aliphatic amino acid substrates and decreased catalytic activity towards L-phenylalanine (Seah et al, 2003).

	Conserved residues									
GluDH <i>C. symbiosum</i>	K89	K113	Q110	K125	D165	R205	S380	G90	A163	V377
GluDH <i>M. smegamatis</i>	K93	K117	Q114	K129	D169	R209	S381	G94	A167	V378
GluDH <i>E. coli</i>	K92	K116	Q113	K128	D168	R207	S380	G93	A166	V377
GluDH <i>P. asaccharolyticus</i>	K70	K94	M91	K106	D146	R187	S355	G71	A144	V352
Bovine	K90	K114	M111	K126	D225	R211	S381	G91	A223	V378
PheDH <i>B. sphaericus</i>	L51	K79	M75	K91	D133	S157	V310	G52	M131	L307
LeuDH <i>B. sphaericus</i>	L40	K68	M65	K80	D122	P146	V294	G41	V120	V291

Table 5.4. Shows the residues responsible for the substrate binding of GluDHs, PheDH and LeuDH.

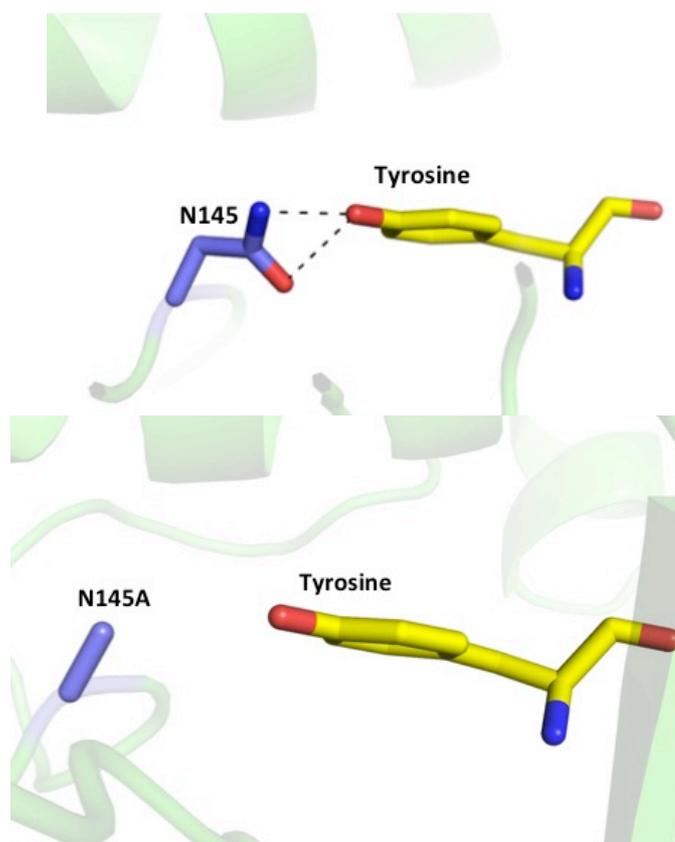


Figure 5.18. A tyrosine substrate modeled into the active site of *B.sphaericus* PheDH, based on the position of the phenylalanine substrate in *Nocardia* PheDH/Phe/NADH complex. In the wild type PheDH (upper picture), OH of the tyrosine side chain of the substrate hydrogen bonds to the side chain of N145 residue. No hydrogen bond can be made between the side chain of the substrate and the side chain of N145A in the case of the single mutant PheDH (lower picture).

Chapter Six: Cloning, Over-Expression and Purification of NDH-2 from *M. smegmatis*, a Model of *M. tuberculosis*

6.0 Introduction

In addition to the work described in this thesis on amino acid dehydrogenases, other dehydrogenase enzymes were investigated for potential biotechnological or therapeutic use. One such enzyme identified is the type II NADH dehydrogenase: menaquinone oxidoreductase (NDH-2) of *Mycobacterium tuberculosis*. This enzyme is part of the electron transport chain of *M. tuberculosis* and forms part of this system that regulates the menaquinone pool (Teh et al, 2007). This enzyme has been identified as a potential drug target (Warman et al, 2013; Weinstein et al, 2005; Yano et al, 2006), and its structure is unknown. This chapter gives an introduction to attempts to try to elucidate the three dimensional structure of NDH-2.

6.1 Tuberculosis:

Tuberculosis is defined as the infectious disease caused by *Mycobacterium tuberculosis* complex, which includes, *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti* and *M. canettii*. Although all these organisms (except *M. microti*) may cause disease to humans, *M. tuberculosis* is the most common cause of tuberculosis (Palomino et al, 2007). *M. tuberculosis* is a bacillus bacterium measuring about 0.53 μm , and requires a special stain, so it is classified as acid-fast bacilli (Wilson et al, 2011). The cell wall of *M. tuberculosis* is rich in lipid complexes such as mycolic acid and sulpholipids (figure 6.1). *M. tuberculosis* is an aerobe, it needs free oxygen for growth. However, the division time of this bacteria is longer than for other bacteria, it takes from 12 to 20 hours to double and may require many weeks to grow visible colonies (Madkour, 2001).

6.2 Epidemiology of tuberculosis:

About one third of the population of the world is infected by mycobacterium tuberculosis, and roughly 8 million new cases are diagnosed, and approximately two million deaths are estimated every year. In addition, 95%

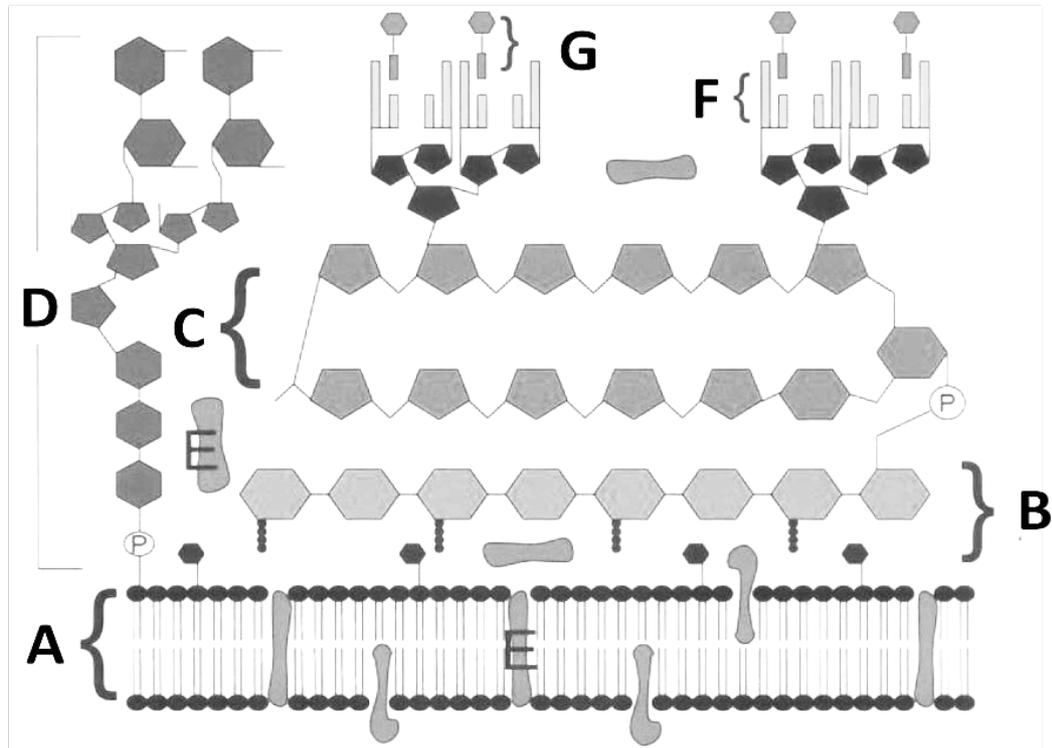


Figure 6.1 The cell wall of *Mycobacterium tuberculosis*.(A) plasma membrane (B)peptidoglycan, (C) arabinogalactan, (D)mannose-capped lipoarabinomannan, (E)plasma membrane and cell wall-integrated proteins, (F) mycolic acid and (G) glycolipid surface molecules associated with the mycolic acids. This figure has been adapted from (Karakousis et al, 2004).

of the tuberculosis cases arise in developing countries, because of poverty and the spread of HIV infection. The two central reasons for the rapid dissemination of this disease rapidly are crowded living conditions, and a population with little native resistance. Nearly all infections of *M. tuberculosis* are a result inhalation of the infected particles with tubercle bacilli. Moreover, these particles are spread by the airborne way from the infected persons with pulmonary tuberculosis during coughing or sneezing to individuals who become infected after inhalation these particles (Malani, 2010).

6.3 Treatment of Tuberculosis:

There are number of special antibiotics which are considered as the most effective drugs against mycobacterium tuberculosis, and are used as the backbone in the treatment of this disease. These antibiotics are isoniazid, ethambutol and pyrazinamide. Furthermore, the medicines of anti-tuberculosis are separated into a number of groups depending on the effectiveness of these drugs; isoniazid, rifampicin, ethambutol and pyrazinamide are used as the frontline drugs to cure the tuberculosis infections (Blumberg et al, 2003). Patients need second line drugs if drug resistant TB appears. These drugs consist of: fluoroquinolone, cycloserine, ethionamide and para-aminosalicylic acid, and some intravenous agents like streptomycin, capreomycin, kanamycin and amikacin. Third line drugs are required when the infected person has an extremely drug resistant form (Blumberg et al, 2003).

6.4 Mechanism of action of anti-tuberculosis agents:

The most potent drugs, which currently are used to treat *M. tuberculosis*, are isoniazid, rifampin, ethambutol and pyrazinamide. Isoniazid and ethambutol work by preventing synthesis of mycolic acid, a main constituent of the mycobacterium cell wall, and rifampin acts by inhibiting the synthesis of RNA polymerase. The mode of action of pyrazinamide is not known (Wilson et al, 2011)

6.4.1 Mechanism of action of isoniazid and rifampin:

Isoniazid (INH) is a pro-drug and converted to the active form (the isonicotinic acyl radical) by the catalase-peroxidase enzyme, which is encoded by the KatG

gene. Activated isoniazid binds covalently to NADH, to produce isonicotinic acyl NADH complex, which attaches to a long-chain enoyl-acyl carrier protein reductase (InhA), leading to the inhibition of this enzyme and thus mycolic acid biosynthesis, a major component of mycobacterium cell wall, which leads to death of the cell (Rozwarski et al, 1998; Vilcheze & Jacobs, 2007). Rifampicin works by binding to the prokaryotic RNA polymerase, hence inhibiting transcription to RNA and thus protein translation. This inhibition is due to the tight binding between the RNA polymerase β -subunit and Rifampicin. Structural studies of the rifampicin-RNA polymerase complex have shown that this antibiotic prevents the synthesis of RNA by interference with the initiation stage (Floss & Yu, 2005).

6.5 Drug resistant tuberculosis:

Although the drugs described above are a successful treatment in many tuberculosis infections, there is a worrying trend of drug resistant forms of the disease appearing (World-Health-Organization, 2013). In multi-drug resistant TB (MDR-TB), the mutant bacteria are resistant to isoniazid or rifampicin. In extensively drug resistant TB (XDR-TB), the bacterial strains are also resistant to fluoroquinolone antibiotics and the second-line, intravenous antibiotics, cryptomycin, kanamycine and amikacin (Gillespie, 2002).

M. tuberculosis resistant to isoniazid INH is associated with mutations in the KatG gene, which encodes the activator of INH catalase-peroxidase enzyme that results in a reduction the activity of this enzyme. Thus strains that have a mutation in KatG are expected resistant to INH due to a decreased ability to produce the INH-NAD adduct that inhibits InhA (Vilcheze and Jacobs, 2007). Another mechanism of resistance is the alteration of NADH/NAD⁺ ratio. A mutation in the *ndh* gene that encodes NADH dehydrogenase (NdhII) causes a decrease in NdhII activity, leading to accumulation of NADH, the substrate for NdhII. This accumulation of NADH works as a competitive inhibitor for binding to InhA and protects InhA against the inhibitory effect of the INH-NAD adduct (Vilcheze et al, 2005).

The most common resistance of rifampicin is attributed to a mutation in the *rpoB* gene, which encodes the β -subunit of RNA polymerase, which is

responsible for the activity of prokaryotic RNA-polymerase and is the target for rifampicin. Therefore, a mutation in this gene leads to a reduction in the affinity of RNA-polymerase to the antibiotic, which usually attaches tightly to the wild-type enzyme. This reduction in affinity between rifampicin and RNA polymerase is associated with decreasing susceptibility of the bacteria to inhibition by this antibiotic (Floss & Yu, 2005).

6.6 Development of new antibiotics:

Three approaches can be directed towards overcoming this increased resistance. Firstly, the use of good hygiene in a hospital setting is paramount. Secondly, the proper use of current antibiotics should be encouraged and regulated. Thirdly the development of novel antimicrobial agents could overcome the antibiotic resistance problem. There are several methods to develop new antibiotics including structure-based drug design.

6.6.1 Structure-based drug design:

Structure-based drug design requires a high resolution, three-dimensional structure of the protein molecule, to which small molecule drugs are targeted. Typically, a structure will be available of one molecule in complex with its target. Using this knowledge it is possible to design the best possible inhibitor to interact with the selected target, in a process of rational drug design (Congreve et al, 2005). The application of structure-based drug design extends from helping in validation of a protein target to molecular modeling and design, to organic synthesis, and the drug optimization stage. This method has succeeded in discovering some drugs that have been manufactured and arrived at the market, AIDS drugs for example: Agenerase and Viracept were designed using the crystal structure of HIV protease (Lapatto et al, 1989; Miller et al, 1989). The flu drug Relenza was developed using the crystal structure of neuraminidase (Varghese, 1999). Furthermore, development of experimental approaches and bioinformatics research has caused a leap of activity in structure-based drug design over the last years (Arcus et al, 2006). This has led to the discovery of more than 40 compounds, which have entered clinical trials (Congreve et al, 2005).

6.7 Organization of respiratory chain of *Mycobacterium tuberculosis*:

Although TB is the cause of death about 2 million people annually (World-Health-Organization, 2013) there is no clear description of the pathways and components of central energy metabolism. Indeed, most information concerning the organization of the electron transport chain of *M. tuberculosis* is mainly dependent on genomic data (Cole et al, 1998; Galagan, 2014).

6.7.1 Respiratory chain (Electron Transport Chain):

The respiratory chain is a series of protein complexes bound to the bacterial cell membrane, and is the energy factory of respiratory organisms. The electron transport chain catalyzes the movement of electrons from NADH or succinate (reducing substrates) to an electron acceptor. In aerobic conditions oxygen acts as the electron acceptor, while sulfate or nitrate receive the electrons under anaerobic conditions. The movement of electrons through the respiratory chain is coupled to protons pumping across the membrane, when more electrons flow, more protons move outside the membrane causing a proton gradient and a membrane potential, that allows the ATP synthase enzyme to produce ATP from ADP and Pi (Hatefi, 1985; Kana et al, 2009)

6.7.2 Electron transport chain of *M. tuberculosis*:

The respiratory chain of *M. tuberculosis* is similar to those found in other bacteria. It catalyzes transmission of electrons from NADH or succinate to oxygen, which is preferred as the electron acceptor, because *M. tuberculosis* is categorized as an aerobe. This pathway can produce substantial energy for exponential growth in oxygen rich conditions (Teh et al, 2007).

According to genomic data, the electron transport chain of *M. tuberculosis* is a branched pathway terminating in two terminal oxidases (Weinstein et al, 2005). The first oxidase enzyme is cytochrome *bd* oxidase (quinol oxidase), which is induced under low-oxygen states in *Mycobacterium smegmatis*, and it is necessary for bacterial growth under aerobic conditions (Kana et al, 2001). The second enzyme is aa3-type cytochrome oxidase that is essential for exponential growth in rich oxygen milieu (Weinstein et al, 2005). Both pathways chains are supplied by the menaquinol pool (MQH₂), that is produced by 4 oxidoreductase enzymes, these are: succinate menaquinone

oxidoreductase (SQR), and two types of NADH:menaquinone oxidoreductase, a multimeric type I NADH dehydrogenase (complex I), where redox activity is linked to proton pumping, and a type II NADH:menaquinone oxidoreductase (NDH-2), which exists in two forms, *ndh* and *ndhA*. The redox activity of this enzyme is not linked to proton pumping (Fisher et al, 2009). The cytochrome bc1 complex works as a linking enzyme, oxidizing menaquinone and, in turn, reducing cytochrome c. The genome of *M. tuberculosis* also encodes two anaerobic respiratory complex genes, nitrate reductase *narGHJI*, and fumarate reductase *frdABCD* and, lastly, the ATP synthase super complex which uses the proton motive force through the bacterial membrane to generate ATP (figure 6.2) (Fisher et al., 2009)(Weinstein et al, 2005). Of the 4 enzymes that contribute to the menaquinol pool, the type II NADH: menaquinone oxidoreductase (NDH-2) has been validated as a drug target for tuberculosis. This enzyme plays a significant role in feeding MQH₂ during infection, as well as one of the essential adaptation processes of *M. tuberculosis* against extracellular stress, which results in an upregulation of NDH-2 over NDH-1 (Teh et al., 2007, Fisher et al., 2009, (Weinstein et al, 2005).

6.7.3 Type II NADH: Menaquinone oxidoreductase (NDH-2) of *M. tuberculosis*:

The structure of the NDH2 from *M. tuberculosis* is unknown and the data concerning structure and function of this is limited to genomic data and modeling studies depend on the sequence similarity with solved structure of dehydrogenase enzymes for example, lipoamide dehydrogenase (Fisher et al., 2009). These enzymes share 24% sequence similarity. According to those studies, the *M. tuberculosis* NDH2 structure can be predicted. It is made up from a single polypeptide chain containing 463 amino acids in length with molecular weight 49,619 Da. It has two GXGXXG motifs in β - α - β structures, which are used for binding the dinucleotides.

6.7.4 Type II NADH: menaquinone oxidoreductase catalyzes the first reaction in electron transport chain of *M. tuberculosis*

Results obtained from transposon-insertion mutation experiments suggest that type I NDH and the second type II NADH dehydrogenase *ndhA* are not

essential for *M. tuberculosis* growth *in vitro*. However, transposon disruption of the type II NDH gene (NDH-2) is lethal and it is necessary for optimal growth of *Mycobacterium tuberculosis* *in vitro* (Sasseti & Rubin, 2003). In addition, mutations in the NDH-2 gene give conditional mortality on *Mycobacterium smegmatis* (Miesel et al, 1998). Furthermore, inhibition of NDH-2 enzyme by the antibiotic phenothiazine leads to inhibition of *M. tuberculosis* growth both *in vitro* and *in vivo* (Weinstein et al, 2005).

It is likely that NDH-2 acts as the major NADH oxidoreductase of the electron transport chain in exponential growth in environments rich in oxygen and nutrients (Teh et al, 2007). However, because human mitochondria possess just type I NADH dehydrogenase, while NDH-2 is the unique NADH dehydrogenase in the *M. tuberculosis* respiratory chain, this enzyme can be a valuable drug target.

6.8 Aims of project

As stated above, the type II oxidoreductase acts as the major dehydrogenase enzyme in the electron transport chain of *M.tuberculosis*. However, because human mitochondria possess just the type I NADH dehydrogenase, while NDH-2 is the unique NADH dehydrogenase in the Mtb respiratory chain, the NDH-2 enzyme could be a valuable drug target. Knowledge about 3D structure is a vital tool in the discovery of possible lead inhibitors and ultimately in the design and optimization of possible drug compounds. There is currently no high resolution structure known for NDH-2. The overall aim of this project therefore, was to determine the structure of NDH-2 to analyze inhibitor binding to this protein. Therefore, attempts were made to:

- 1- Clone the gene which encodes NDH-2 protein from *M. smegmatis* a model of tuberculosis (the identities = 371/454 (81%)).
- 2- Express and purify NDH-2 protein.
- 3- Crystallize this protein.
- 4- Determine the structure of NDH-2 protein.
- 5- Analysis of inhibitor binding to NDH-2.

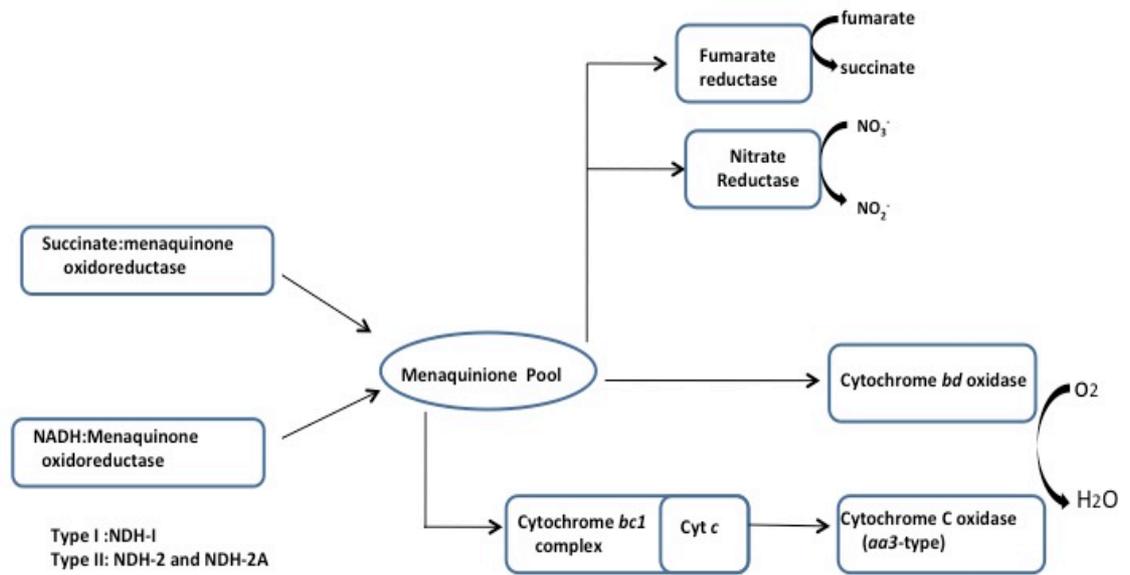


Figure 6.2. Putative organization of the *M. tuberculosis* respiratory chain. NADH: Menaquinone oxidoreductase includes type I and type II NADH. This figure has been adapted from Teh *et al.* (2007).

6.9 Identification of *Mycobacterium tuberculosis* *ndh* gene, which encodes NDH-2 protein:

The *ndh* gene was identified by searching the tuberculist website (<http://genolist.pasteur.fr/TubercuList>) to obtain both the DNA and protein sequences. A BLAST search, at the NCBI (<http://blast.ncbi.nlm.nih.gov/>) was used for aligning the (NDH-2) protein sequence of *M. tuberculosis* with *M. smegmatis* proteins. The alignment results, which are presented in figure 6.4, have shown that there is 82% identity between NDH-2 of *M. tuberculosis* and *M. smegmatis* NDH-2. Therefore, the NDH-2 of *M. smegmatis* is a good model of *M. tuberculosis* NDH-2.

6.10 PCR for amplification of *M. smegmatis* NDH2 gene (*M. smeg3621*)

Polymerase Chain Reaction (PCR) was used to amplify the *M. smegmatis* NDH2 gene (*msmeg3621*) from *M. smegmatis* genomic DNA (strain ATCC 700084 / mc (2)155). Primers were designed based on the nucleotide sequence and included restriction sites for NdeI and BamHI. For the PCR reaction the following materials were added to the PCR tube:

Sterile H ₂ O	27.5µl
DNA Polymerase Buffer 5x	10µl
dNTPs (5mM)	2.5µl
Forward primer (5µM)	4.0µl
Reverse primer (5µM)	4.0 µl
Genomic DNA (<i>M. smegmatis</i>)	2.0µl
DNA polymerase	1.0 µl
Total volume	50 µl

The tube was then placed in a PCR machine and the reaction was set up under the following conditions:

Initial denaturation	95°C	2 min	
Denaturation	95°C	1 min	25 cycles
Annealing	60°C	1:30 min	
Elongation	72°C	2:00 min	
Final Elongation	72°C	7.0 min	

The PCR products were run on a 1% agarose gel, and a 1.4 kb band was seen, as expected for the *msmeg3621* gene (1400bp). The PCR products were purified using QIAquick® gel extraction kit (Qiagen) (see section 2.1.2.1) and were stored at -20°C prior use in cloning. Figure 6.5 shows the PCR products at the expected size.

6.11 Cloning of *M. smegmatis* NDH2 gen (*M. smeg3621*)

The *msmeg3621* gene PCR product was digested with the restriction enzymes NdeI and BamHI, before insertion into pET14b plasmid, which was cleaved with the same enzymes. After incubating the digestion mixture for 2 hours at 37 °C, the samples were analyzed on a 1% agarose gel and the digested insert and vector were purified using a QIAquick® gel extraction kit (Qiagen) prior to setting up the ligation reaction (see section 2.1.5). The ligation mixes were transformed into competent DH5α *E. coli* cells (Novagen) (see section 2.1.5.1). As plasmid pET14b has an ampicillin resistance gene, the resulting transformants were spread on LB-agar plates supplemented with 50µg/mL ampicillin, and the plates were incubated overnight at 37°C. Some colonies could be seen on the plate and were screened for successful cloning.

6.12 Identification of successful clones

A first screen of the possible clones was done by restriction digestion. A number of colonies were picked up from the plated clones and the possible constructs were extracted using a miniprep kit (Qiagen) (see section 2.1.3). The resultant miniprep samples were digested with the restriction enzymes

NdeI and BamHI (see section 2.1.4) and again the samples were run on the 1% AGE to verify successful ligation. From the gel picture (figure 6.6) it can be seen that some clones gave bands at the expected size (1400 bp) showing the existence of the desired gene in these clones.

6.13 Expression of *M. smegmatis* NDH2

The confirmed construct was then transformed into a competent *E. coli* strain, BL21 (DE3) (Novagen), and plated onto an Ampiciline agar plate. A small-scale overexpression experiment was done (see section 2.1.6.2 for details) to find the appropriate conditions to produce a soluble protein. The suitable condition was found to be 37°C, 1mM IPTG and induction time 5 hours. Two 2L Erlenmeyer flasks containing 500 ml LB, supplemented with 50µg/mL Amp, and inoculated with the 10ml of starter culture were used. The flasks were left to grow at 37°C with shaking at 200rpm until the A_{600} reached 0.6, before adding IPTG for induction (final concentration 1mM). The cells were allowed to over-express for five hours, under the same conditions. The cells were centrifuged at 1844 g for 20 min for harvesting and the cell pastes were stored at -80°C prior to use in purification experiments.

6.14 Purification of NDH2 from *M. smegmatis*

The cells thawed on ice and suspended in 3ml BugBuster (Novagen) with 5µl of benzonase and the mixture was mixed by vortex. The suspension was kept on ice for 30min after that it was centrifuged at 4°C and 4500 rpm for 30min and the proteins in the supernatant were collected. His-tagged NDH-2 protein was purified by metal-chelating chromatography using Ni-NTA resin (Qiagen) at 4°C. 1ml of Ni-NTA resin Qiagen was pre-equilibrated with phosphate buffered saline PBS (50mM NaH₂PO₄, 400mM NaCl, pH=8.0) and it was mixed with the supernatant. After that the mixture was rotated at 4°C for 30min before packing onto a gravity column. Contaminants were washed out using PBS buffer and the target protein was eluted with elution buffer (50mM NaH₂PO₄, 400mM NaCl, pH=8.0 and 250mM Imidazole). The elution was collected in 0.5 ml fractions. After cell lysis the greater part of NDH-2 protein was in soluble form. Further purification was carried out using gel filtration, the volume of the sample was reduced to 1ml using a VivaSpin concentrator

a

Mtb	5	EPTAQPPRRHRVVIIGSGFGGLNAAKCLKRADVDIKLIARTTHHLFQPLLYQVATGIISE	64
		P A RH+VVIIGSGFGGL AAK LKRADVD+KLIARTTHHLFQPLLYQVATGIISE	
Msmeg	3	HPGATASDRHKVVIIGSGFGGLTAAKTLKRADVDVLIARTTHHLFQPLLYQVATGIISE	62
Mtb	65	GEIAPPTRVLRKQRNVQVLLGNVTHIDLQCVVSELLGHTYQTPYDSLIVAAGAGQSY	124
Msmeg		GEIAP TRV+LRKQ+N QVLLG+VTHIDL + V S LLGHTY TPYDSLII+AAGAGQSY	63
		GEIAPATRVILRKQKNAQVLLGDVTHIDLLENKTVDSVLLGHTYSTPYDSLIIAAGAGQSY	122
Mtb	125	FGNDHFAEFAPGMKSIDDALELRGRILSAFEQAERSSDPERRAKLLTFTVVGAGPTGVEM	184
		FGNDHFAEFAPGMKSIDDALELRGRIL AFEQAERSSDP RRAKLLTFTVVGAGPTGVEM	
Msmeg	123	FGNDHFAEFAPGMKSIDDALELRGRILGAFEQAERSSDPVRRAKLLTFTVVGAGPTGVEM	182
Mtb	185	AGQIAELAEHTLKGAFRHIDSTKARVILLDAAPAVLPPMGAKLQRAAARLQKLGVEIQ	244
		AGQIAELA+ TL+G+FRHID T+ARVILLDAAPAVLPPMG KLG++A ARL+K+GVE+QL	
Msmeg	183	AGQIAELADQTLRGSFRHIDPTEARVILLDAAPAVLPPMGKLGKARARLEKMGVEVQL	242
Mtb	245	GAMVTDVDRNGITVKDSGDGTVRRIESACKVWSAGVSASRLGRDLAEQSRVELDRAGRVQV	304
		GAMVTDVDRNGITVKDSGDG+RRIESACKVWSAGVSAS LG+DLAEQS VELDRAGRV+V	
Msmeg	243	GAMVTDVDRNGITVKDSGDGTVRRIESACKVWSAGVSASPLGKDLAEQSGVELDRAGRVKV	302
Mtb	305	LPDLSIPGYPNVFVVGDMAAVEGVPVGAQGAIQGAKYVASTIKAELAGANPAEREPFQYF	364
		PDL++PG+PNVFVVGDMAAVEGVPVGAQGAIQG +Y A IK E++G +P R PF+YF	
Msmeg	303	QPDLTLPGHPNVFVVGDMAAVEGVPVGAQGAIQGGRYAANKIKREVSGETSPKIRTPFEYF	362
Mtb	365	DKGSMATVSRFSAVAKIGPVFEFSGFIAWLILVLHLYLIGFKTKITLLSWVTFLSTR	424
		DKGSMATVSRFSAVAK+GPVEF+GF AWL WLVLHL YL+GFKTKI TLLSW VTFLST+	
Msmeg	363	DKGSMATVSRFSAVAKVGPVEFAGFFAWLVCWLVLHLYLVGFKTKIVTLLSWGVTFLSTK	422
Mtb	425	RGQLTITDQAFARTRLEQLAEIAAEAQGSAASA	458
		RGQLTIT+QQA+ARTR+E+L E+AA Q + +A	
Msmeg	423	RGQLTITEQQAYARTRIEELEEIAAAVQDTEKAA	456

b

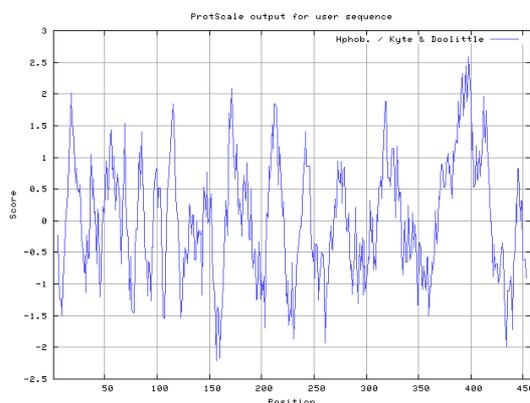


Figure 6.3. a) Sequence alignment of *M. tuberculosis* (*Mtb*) NDH-2 and its homologues from *M. smegmatis* (*Msmeg*) NDH-2. Identities = 371/454 (82%), Positives = 409/454 (90%), Gaps = 0/454 (0%). b) Hydrophobicity plot of the NDH-2 protein from *M. smegmatis*.

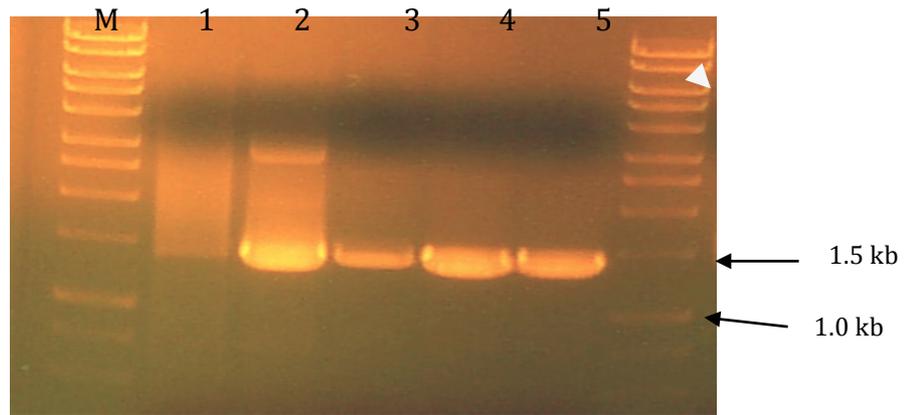


Figure 6.4. Electrophoretic pattern of agarose gel. PCR products of *ndh* gene from *M. smegmatis*. The numbers at the top represents the five different PCR reaction products. The agarose gel picture shows that there were bands at the expected size (1.4 kb), and they might be the gene of interest.

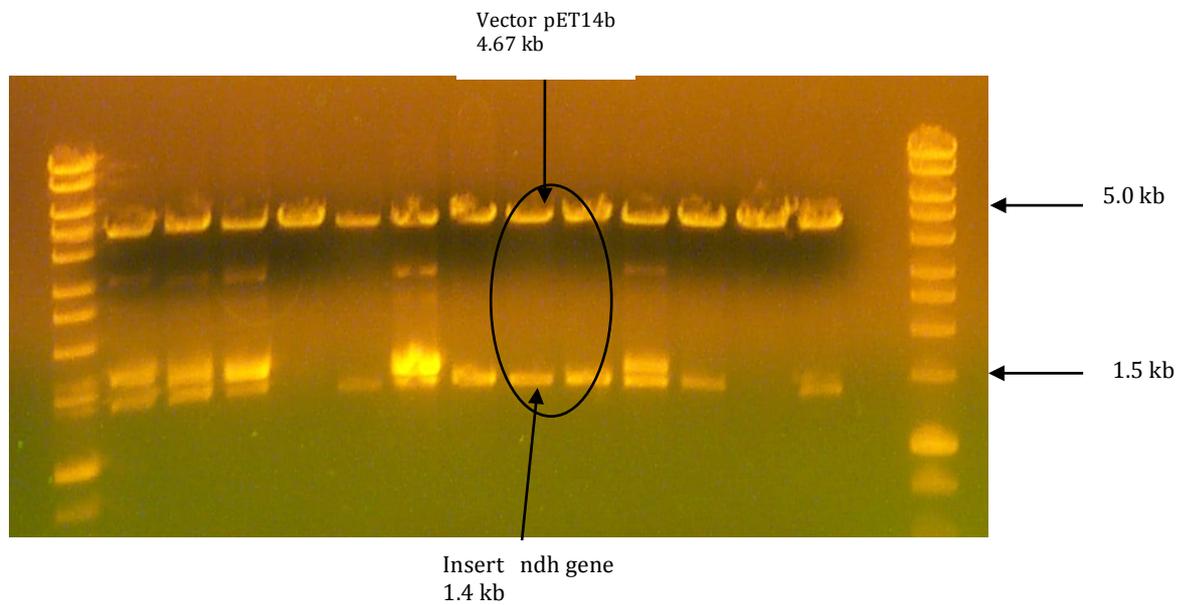


Figure 6.5. A picture of a 1% AGE gel showing the screening results for the recombinant pET14b with the *Mycobacterium smegmatis* *ndh* gene that had been digested by NdeI and BamHI. Some of these lanes represent the correct cloning.

with a molecular weight cut off of 30kDa, and was then applied to a Superdex200 gel filtration column, equilibrated in buffer A +0.5M NaCl. Gel filtration separation was performed with the same buffer at a flow rate at 1.5ml/min. 2ml fractions were collected after the void volume of the column had been discarded. Fractions were analysed by 10% SDS-PAGE (figure6.7 and 6.9). It can be seen from figure 6.9 that NDH-2 is present in all the fractions from the gel-filtration column, perhaps indicating that the protein is not correctly folded, and not interacting with the column in a specific manner. As NDH-2 is believed to be a monotopic membrane protein, interacting with the membrane at a C-terminal helix, it is perhaps not surprising that the full length protein behaved anomalously.

One strategy to overcome this apparent misfolding would be to design a truncated construct in an attempt to improve the solubility. However, at this point in the project the structure of the NDH-2 from *Caldalkalibacillus thermarum* was determined (Heikal et al, 2014). This protein shares 31% sequence identity of *M. tuberculosis* NDH-2 (figure6.4), whereas the identity between *M. smegmatis* NDH-2 and of *M. tuberculosis* NDH-2 is 82% (figure6.3). The structure reported is of the full length enzyme, but expression and purification were undertaken in the presence of the detergent n-octyl- β -D-glucopyranoside. Analysis of the structure of the *C. thermarum* NDH-2 shows that a C-terminal domain comprising of three antiparallel β -strands and two amphipathic α -helices is responsible for anchoring to the membrane (Heikal et al, 2014). Indeed if the protein is truncated at residue Ile379, removing this domain, overexpressed NDH-2 appears in the cytoplasmic fraction, rather than with the membrane, indicating importance of these residues as the membrane anchor (Heikal et al, 2014). Although the structure of a bacterial NDH-2 is now known, it would still be useful to obtain the structure of this enzyme from mycobacterial source. To this end the use of detergents in the expression and purification of *M. smegmatis* NDH-2 should be investigated, as well as producing possible truncated constructs, removing the C-terminal membrane-anchoring domain.

```

Query 17  VVIIGSGFGGLNAA----KCLKRADVDIKLIARTTHHLFQPLLYQVATGIISEGEIAPPT 72
+VI+G+G+GG+ AA  K+L  + DI L+ +  +H  L+Q A G +  +
Sbjct 6  IVILGAGYGGIVAAALGLQKRLNYNEADITLVNKNDYHYITTELHQPAAGTMHHDQ----A 61

Query 73  RVVLRK---QRNVQVLLGNVTHIDLAGQCQVSELLGHTYQTPYDSLIVAAGAGQSYFGND 129
RV +++  ++ ++ +  V ID  Q V +  +  YD L+V G+  FG +
Sbjct 62  RVGIKELIDEKKIKFVKDQVVAIDREQQKVTLQ----NGELHYDYLVVGLGSEPETFGIE 117

Query 130  HFAEFAPGKMSIDDALELRGRILSAFEQAERSSDPERRAKLLTFTVVGAGPTGVEMAGQI 189
E A + SI+  +R I  F A+ +++PER  LT V GAG TG+E G++
Sbjct 118  GLREHAFSINSINSVRIIRQHIEYQF--AKFAAEPEP-TDYLTIVVGGAGFTGIEFVGEL 174

Query 190  AELAEHTLKGAFRHIDSTKARVILLDAAPAVLPPMGAKLGQRAAARLQKLGVEIQLGAMV 249
A+  A  +D  R+I ++AAP VLP  L A L GVE ++G +
Sbjct 175  ADRMPELC--AEYDVPKLVRIINVEAAPTVLPGFDPALVNYAMDVLGGKGVFVKIGTPI 232

Query 250  TDVDRNGITVKDSGTVRRIESACKVNSAGVSASRLGRDLAEQSRVELDRAGRVQVLPDL 309
G+ + + DG  I++A VW+ GV  G + E+S E R GR++V P L
Sbjct 233  KRCTPEGVVI-EVDGEEEEIKAATVVWTGGVR----GNSIVEKSGFETMR-GRIKVDPYL 286

Query 310  SIPGYPNVVFVGDMAAV-----EGVPGVAQGAIQGAKYVASTIKAELAGANPAEREPPQ 363
PG+ N+F+VGD A +  P AQ AIQ + VA+ + A + G +  PF+
Sbjct 287  RAPGHENIFIVGDCALIINEENRPPPTAQIAIQHGENVAANLAALIRGGS---MTPFK 343

Query 364  YFDKGSMTVSRFSAVAKIGPVEFSGFIA-WLIWLV-LHLAYLIG 406
+G++A++ R A+ +G + G A WL L+ +  YLIG
Sbjct 344  PHIRGTVASLGRNDAIGIVGGRKVVYGHAAASWLKKLIDMRYLYLIG 388

```

Figure 6.6. Sequence alignment of *M. tuberculosis* NDH-2 (Query) with *Caldalkalibacillus thermarum* NDH-2 (subject). Identities = 124/405 (31%), Positives = 198/405 (48%), Gaps = 37/405 (9%).

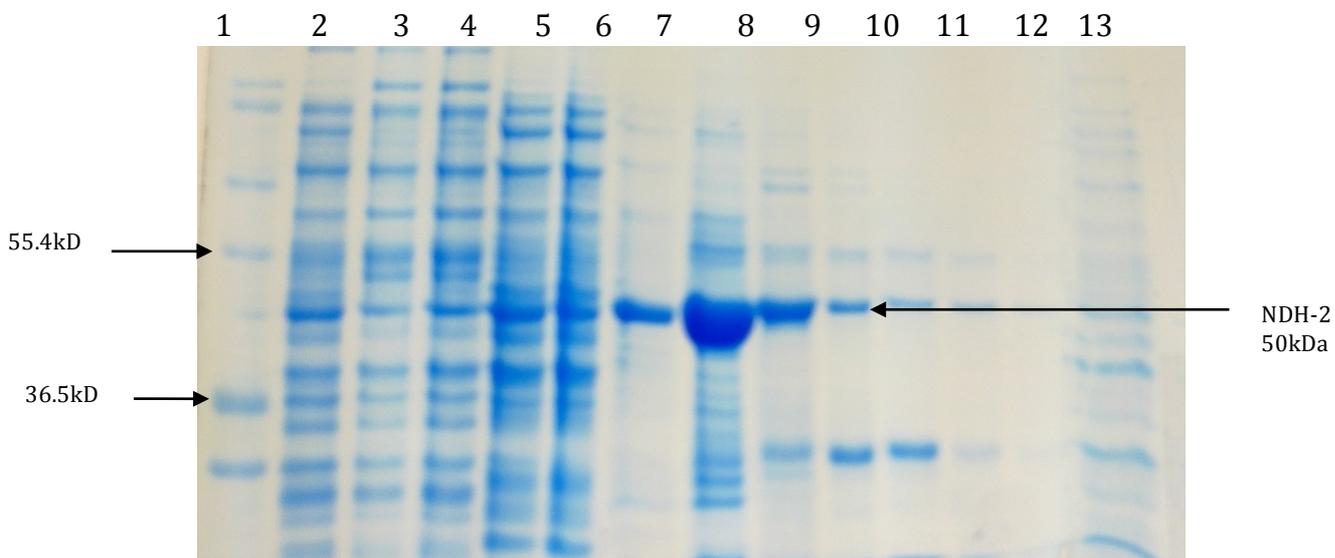


Figure 6.7. 10% SDS.PAGE analysis of purification of NDH-2. Lane1=(M12) molecular weight markers (kDa), lane2=pre-induction, lane3=pellet and lane5-14 eluted fractions from Ni-HP 5ml.

```

No 1: Sample Volume and Type
      NDH2 full lenth 116mg
No 2: Column
      Ni-HP 5ml
No 3: Eluent A
      tris 0.5M NaCl
No 4: Eluent B
      A+0.5M Im
No 5: Remarks
      0-100% 15CV 10.2.11
  
```

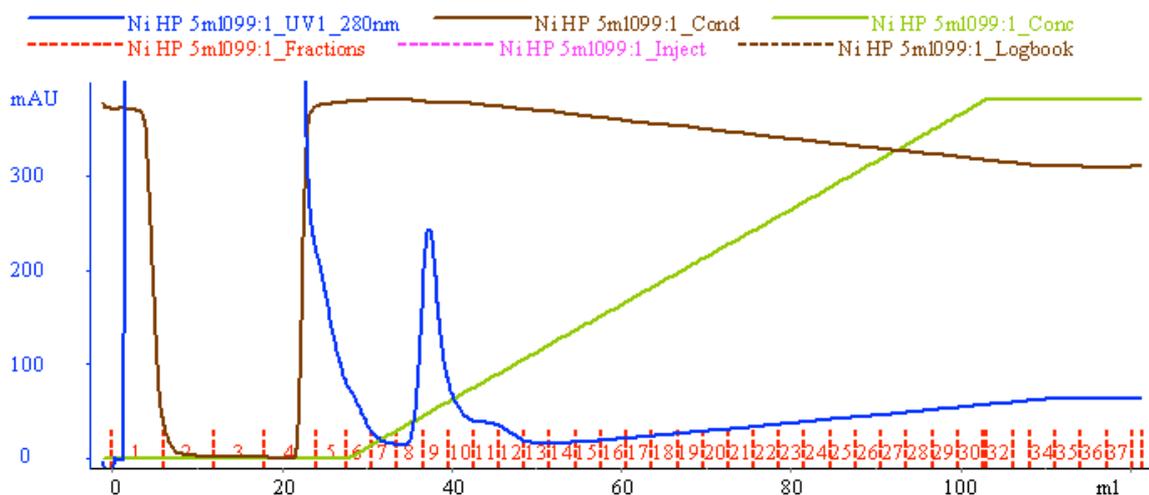


Figure 6.8. Chromatogram picture for the purification of NDH-2 from *M. smegmatis* using Ni-NTA column.

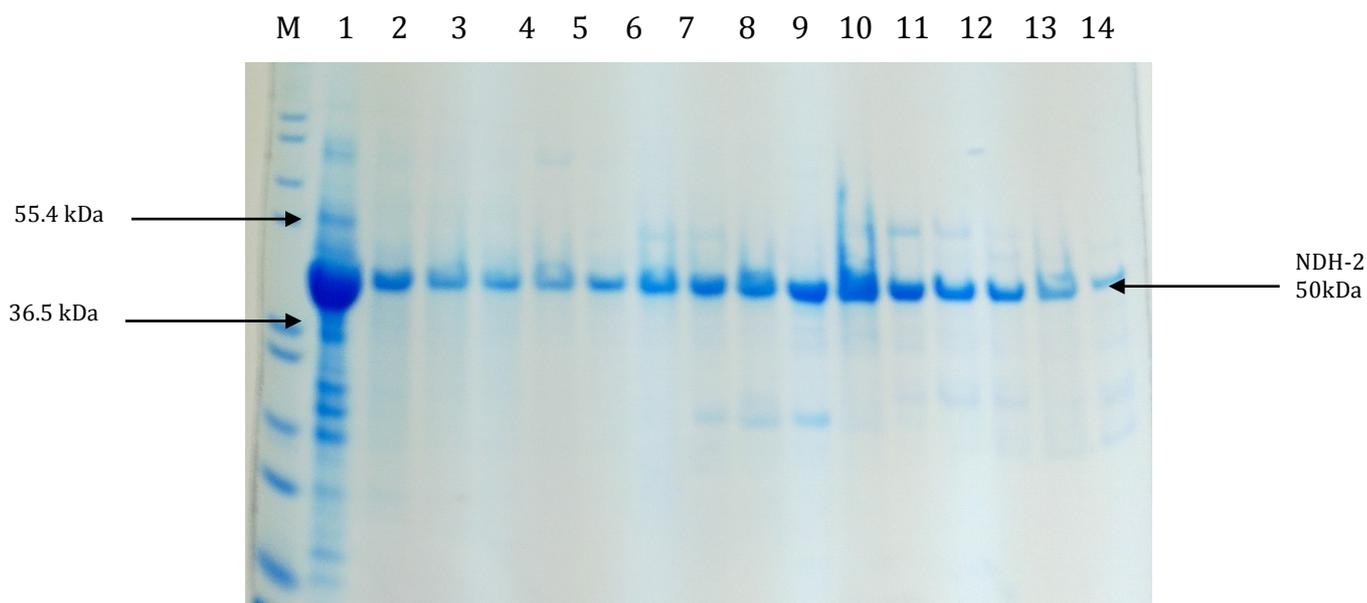


Figure 6.9. 10% SDS.PAGE analysis of purification of NDH-2 protein, lane M (M12) molecular weight markers (kDa). Lane1= sample applied on gel filtration column and lanes 2-16 eluted fractions from gel filtration (superdex200).

```
No 1: Sample Volume and Type
      NDH2 fl 3mg
No 2: Column
      Superdex200 10
No 3: Eluent A
      0.5M NaCl tris
No 4: Eluent B
No 5: Remarks
      10.2.11
```

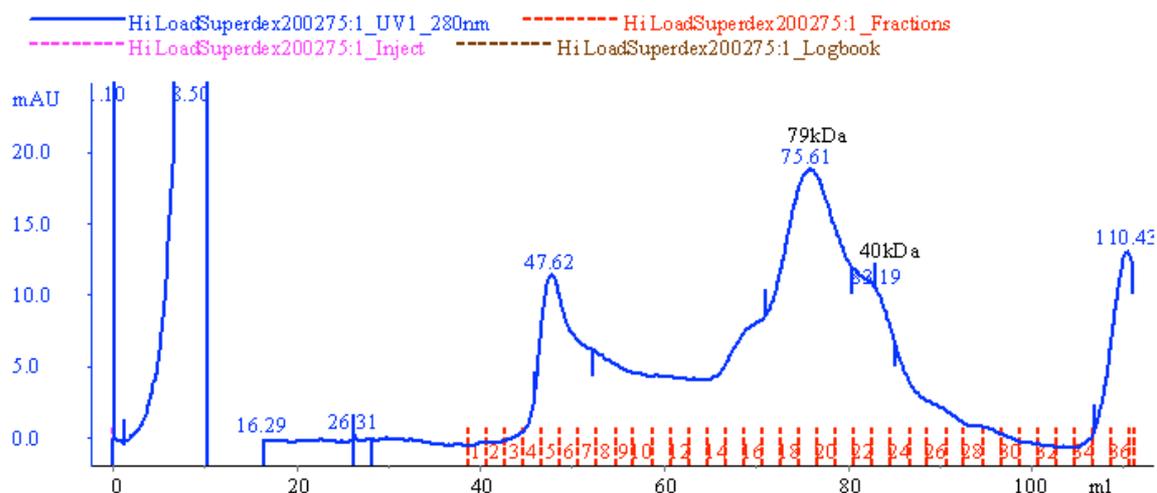


Figure 6.10. Chromatogram picture for the purification of NDH-2 from *M. smegmatis* using gel filtration column.

Chapter seven: Conclusion and future work

7.1 Conclusion

7.1.1 Introduction

The work presented in this thesis had the aim of providing a greater understanding of the structure/ function relationships in the amino acid dehydrogenase family as an aid to help in the biotechnological exploitation of these enzymes as, for example, biosensors or in the production of chiral intermediates for the pharmaceutical industry. To this end, the first structure of *M. smegmatis* GluDH was determined at high resolution, in two different forms, the GluDH/NADP⁺/2oxoglutarate ternary complex at 1.64Å and the GluDH/NADPH binary complex at 1.78 Å. In addition, the binary complex structures of *C. symbiosum* GluDH, with NAD⁺ or L-glutamate, have also been determined, at 1.80 Å and 1.24Å, respectively. Analysis of these structures has provided new information about the conformational changes of the protein during the reaction, the movement of the substrate within the active site during the reaction mechanism and also structural aspects of cofactor specificity among this enzyme family. In addition, the first structure of the PheDH from *B. sphaericus* was determined at medium resolution (2.5Å) and preliminary experiments were undertaken on NDH2 from *M. smegmatis*.

7.1.2 Conformational changes in GluDH

The level of domain motion in the GluDHs family is described in section 4.5 of this thesis. The 1.24 Å binary complex of *C. symbiosum* GluDH with L-glutamate was in a closed form, similar to that seen for the structure determined at 1.9 Å, previously ((Stillman et al, 1993) and PDB code: 1BGV). However, despite the fact that these two structures were derived from crystals grown in very similar conditions, the alignment of the two domains in the subunit of the two structures is different, with the 1.24 Å structure determined here adopting a slightly more closed structure than that of 1BGV (17.7° for 1.24 Å structure and 15.4° for 1BGV structure). These differences are also reflected in slightly different cell dimensions between these two structures, with 1BGV crystallizing in space group R32 with a=162.8 Å and c=102.8 Å, whereas the

1.24 Å structure crystallizes in the same space group, but with $a=158.7$ Å and $c=101.4$ Å, a change of some 3% in the a direction and 1% in the c direction. This change in the relative domain closure also affects the position of the glutamate in the active site of the enzyme. In the 1.24 Å structure the electron density clearly shows that the glutamate binds in two different conformations at the side chain carboxyl end, with one interacting with K89 and S380, and the other interacting with S380 and R205. This latter residue (R205) also adopts a different conformation to that seen in 1BGV, swinging closer to the active site in the 1.24 Å structure (figure7.1). In addition, although the alpha carbon of the glutamate in the two structures adopts a similar position, the alpha carboxyl group of the glutamate in the 1.24 Å structure moves closer to Gly91, with one of the carboxyl oxygens occupying the same space as the catalytic water in 1BGV (figure7.2). These structures show that the glutamate is not fixed within the active site of the enzyme, and can adopt different conformations and interactions with the enzyme.

As might have been expected, the *C. symbiosum* GluDH/NAD⁺ complex was in an open form, similar to that seen for the apo form of the enzyme ((Baker et al, 1992a), pdb code,1HRD). The ternary complex of *M. smegmatis* GluDH with NADP⁺ and 2-oxoglutarate adopted a hyper closed form, which is similar to that seen in the ternary complex of bovine GluDH (NADPH/Glu.). In contrast, the structure of the wild-type apo form of the *E. coli* GluDH adopts a closed conformation ((Smith et al, 2001), PDB, 1HWZ). Analysis of these structures shows that the domain motion between the open and closed structure is not necessarily associated with the binding of substrate, but so far, the hyper closed form has only been observed when both the substrate and cofactor are bound to the enzyme.

7.1.3 Catalytic pocket of GluDH

A comparison between the GluDHs structures that have been determined in this project, has allowed some new annotations to be made about the active site in these enzymes. In the *M. smegmatis* GluDH ternary complex structure, the electron density clearly indicated that the 2-oxoglutarate that had been added to the crystallization solution had bound to the enzyme in its gem diol

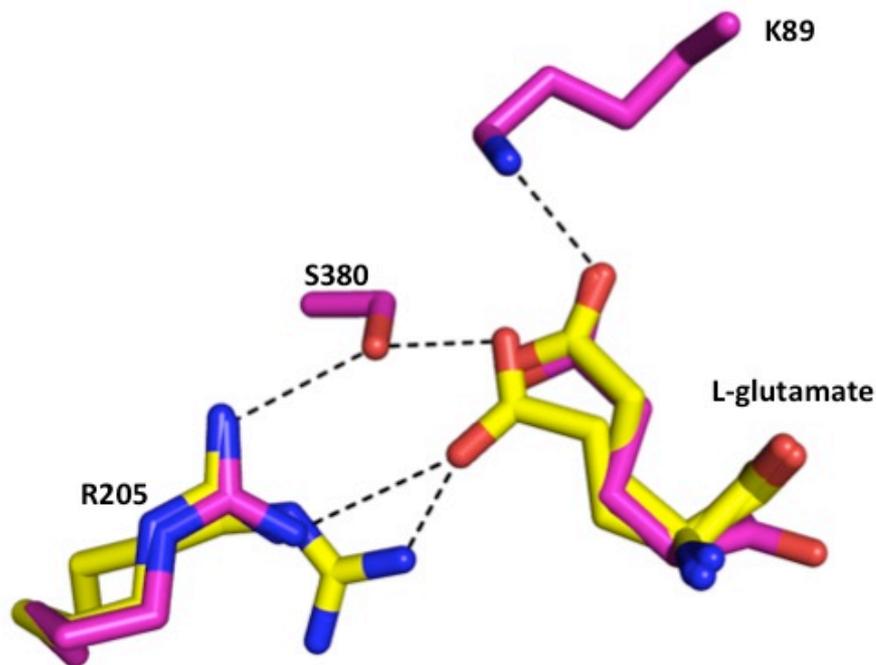


Figure 7.1. A schematic diagram shows the interaction of the side chain carboxyl group of L-glutamate in binary complex (1.24\AA) of the *C. symbiosum* GluDH with glutamate and 1BGV. The two conformations of γ -carboxyl and R205 of the L-glutamate in *C. symbiosum* GluDH binary complex (1.24\AA) are shown in yellow color and L-glutamate, R205 of 1BGV are shown in pink color.

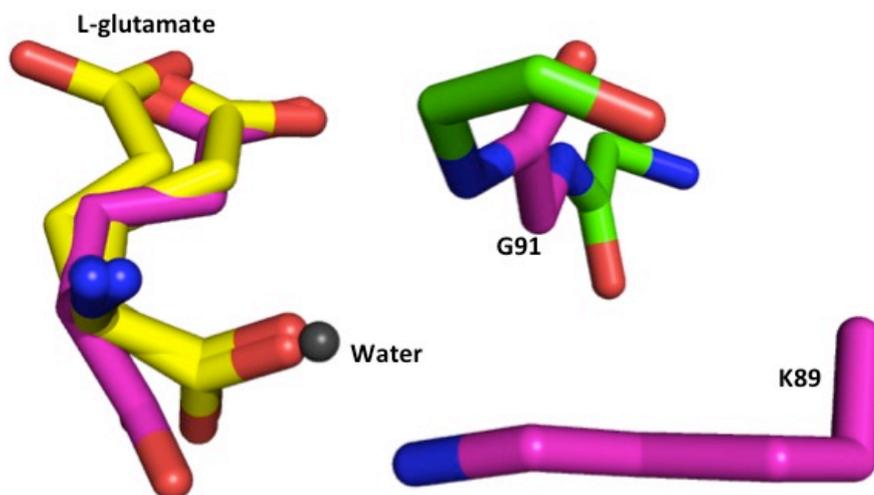


Figure 7.2. Shows the different position of the α -carboxyl group of L-glutamate in *C. symbiosum* GluDH/L-glutamate (1.24\AA), yellow color and 1BGV, pink color. The α -carboxyl group of the substrate of 1.24\AA binary complex (yellow color) moves toward G91 residue and one oxygen atom occupies the water position in 1BGV.

form. The side chain of R209 (*M. smegmatis* GluDH – R205 in *C. symbiosum* GluDH) seems to sense the presence of the substrate. In the presence of L-glutamate or 2-oxoglutarate, this arginine side chain moves towards the binding cleft and interacts with the γ -carboxyl group of the substrate (figure 7.3). In addition, in the ternary complex structures, it also makes a hydrogen bond to the carboxamide oxygen of the nicotinamide ring of NADP⁺ (figure 7.5). This residue is conserved in most of the GluDHs from prokaryotic and eukaryotic species, for example in the ternary complex of bovine-GluDH the side chain of R211 directs to the substrate γ -carboxyl group and forms a hydrogen bond. In the *E. coli* GluDH apo enzyme structure, R207 points in the opposite direction to the empty substrate binding pocket suggesting that this residue can sense the presence of substrate (figure 7.4).

In the *M. smegmatis* GluDH/NADP⁺/2-oxoglutarate ternary complex the gem-diol binds in an extended form, which is considered to be a good model for the proposed carbinolamine intermediate. Indeed the O5 hydroxyl of the gem-diol adopts the same position as the catalytic water molecule in the 1.9 Å *C. symbiosum* GluDH/L-glu complex (1BGV) (figure 7.6). As the interactions made to this hydroxyl of the gem diol include the hydrogen bond donors Lys129 and the main chain NH of Gly95, then this hydroxyl must be in the position of the hydroxyl of the carbinolamine, with the NH₂ of the carbinolamine occupying the position of the second hydroxyl (O6), on the nicotinamide ring side of the substrate. If we assume that the gem-diol is a good model for the carbinolamine, then the attack by ammonia to the keto acid must come from the *pro-S* side of the ketone, to generate the *S*-carbinolamine.

In another unexpected twist, with the gem-diol binding in the extended form observed, the alpha carbon of the gem-diol lies in a different position to that of the L-glutamate in the binary structures, some 1.4 Å further away from the C4 of the nicotinamide ring. In this conformation, Asp169 (*M. smegmatis* numbering) is too far away to act as a base to remove the proton from the carbinolamine hydroxyl to form the keto acid, as proposed in the mechanism of Stillman (Stillman et al, 1993) (figure 7.7). However, Asp 169 can remove a proton from the amine of Lys 129, which in turn, can then act as a base to remove the hydroxyl proton and thus form the ketone.

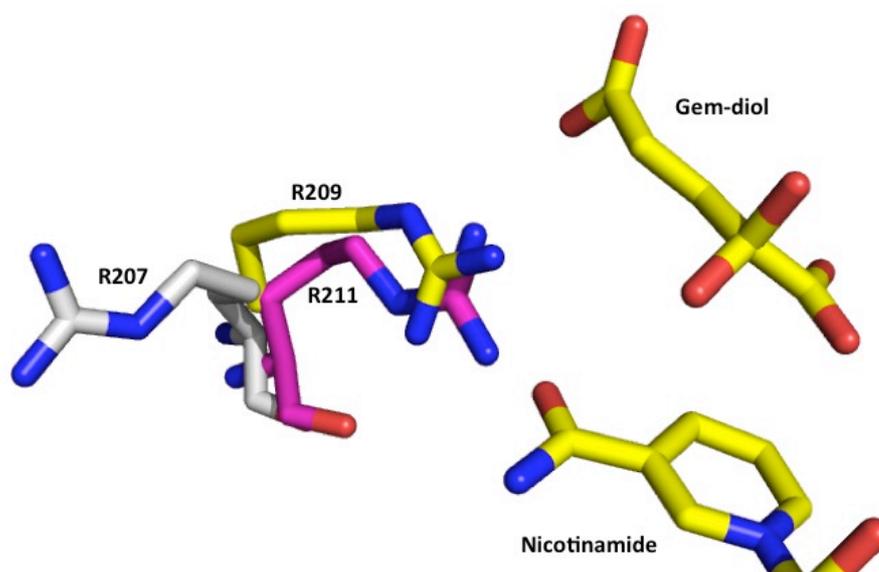


Figure 7.3. Illustrates the different orientations of arginine residue in GluDHs. In *M. smegmatis* GluDH (R209, yellow color) and bovine GluDH (R211, pink color) this residue directs to the active side and bind to the γ -carboxyl of the substrate. In *E. coli* GluDH in the apo state the arginine (R207, gray color) points away from the substrate.

<i>Ms</i>	199	KGLTWGGSRV	R TEATGYGTVFFVDEILQSRG	-----	QSFDGKR	V VVSGSGNVAIYAI	IEKVHALGGIV	260
<i>Ec</i>	197	KG L SFGGSLIR	PEATGYGLVYFTEAMLKRHG	-----	MGFEGMR	V SVSGSGNVAQYAI	EKAMEFGARV	258
<i>Pf</i>	258	KN I KWGGSNIR	AEATGYGVVYFAENVLKDLN	-----	DNLENKK	L VS	SGSGNVAQYLV	EKLEKGAIV
<i>Cs</i>	196	KAR S FGGSLVR	PEATGYGSVYYVEAVMKHEN	-----	DTLVGKT	V ALAGFGNVA	WGAACKLAE	LGAKA
<i>Pa</i>	177	KPVAFGGSEGR	NEATGFGVAVVRESAKRFG	-----	IKMEDAK	I AVQGF	GNVGTFTV	KNIERQGGKV
<i>Bo</i>	201	KP I SQGGIHR	I SATGRGVFHGIENFINEASYMS	ILG	MTPGFGDKT	F VVQGF	GNVGLHSMRYL	HRRFGAKC
								270

Figure 7.4. Sequence alignment of a number of GluDHs. Highlighted Arginine residue is conserved in these enzymes. *Mycobacterium smegmatis* (*Ms*), *Escherichia coli* (*Ec*), *Plasmodium falciparum* (*Pf*), *Clostridium symbiosum* (*Cs*), *Peptoniphilus asaccharolyticus* (*Pa*) and *Bovine* (*Bo*).

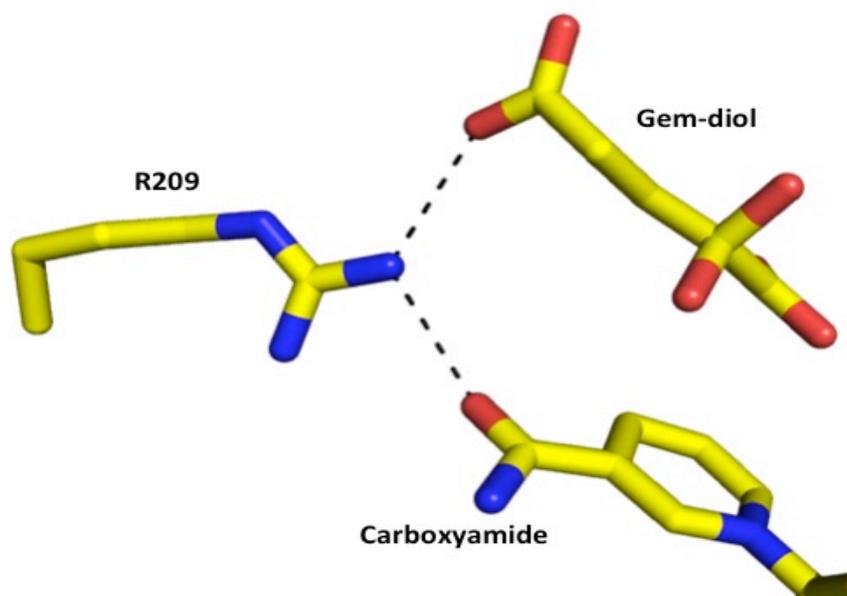


Figure 7.5. Shows the binding of the side chain of R209 with γ -carboxyl group of the gem-diol and carbonyl group of the carboxyamide of nicotinamide of the NADP⁺.

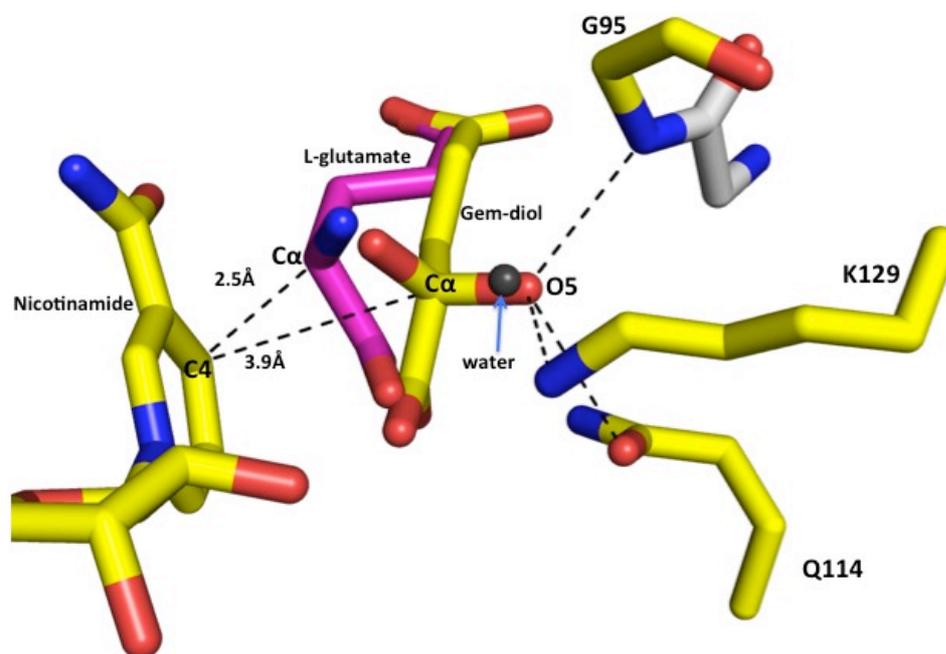


Figure 7.6. Shows the distance between the C4 of the nicotinamide and the alpha carbon for the two substrates, L-glutamate (1BGV) and gem-diol. The α -carbon of the L-glutamate lies 2.5 Å from the C-4 of the nicotinamide, which allows the hydride transfer to occur. The adjacent residues and water molecule can be seen.

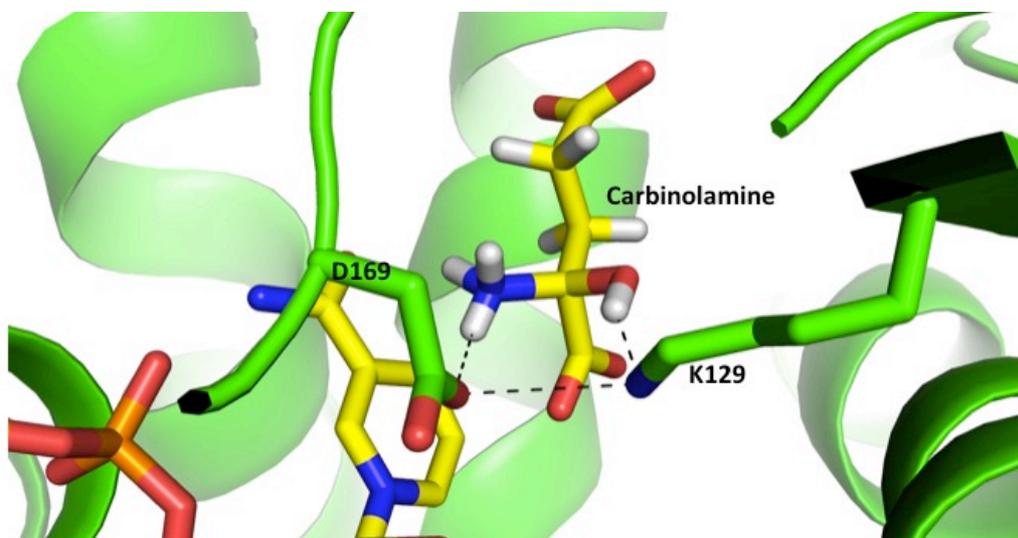


Figure 7.7. A schematic diagram highlighting the position of D169 in the *M. smegmatis* GluDH ternary complex. This residue is in the wrong orientation to abstract a proton from the hydroxyl group of the carbinolamine (O5). However, this side chain could abstract a proton from the amino group of K129, which, in turn, can take the proton from the carbinolamine hydroxyl.

In the reductive amination reaction direction, the next step in the reaction after the formation of the carbinolamine is the dehydration to the iminoglutarate intermediate. This can occur, again using Asp 169 and Lys 129. However, the variation in the position of the substrate observed in the structures determined here would suggest that, as the iminoglutarate is formed, the alpha carbon would move towards the nicotinamide ring, to a suitable position for hydride transfer, again from the *pro-S* side of the iminoglutarate, to produce the product, S-glutamate (L-glutamate). This alteration in the distance between the C α of the substrate and C4 of the nicotinamide of the cofactor is assisted by the substrate movement in the catalytic pocket, rather than by the domain motion.

This variation in the position of the substrate and product of the reaction in the active site of the hyper-closed form of the enzyme, also suggests a mechanism whereby the enzyme can protect against the formation of a hydroxyl-acid, rather than the amino acid, in the reduction direction, as the alpha carbon of the 2-oxoacid is too far from the C4 for hydride transfer to occur, before attack of ammonia has happened.

The similarity in structure and conservation of the crucial residues involved in the active site of GluDHs from both prokaryotes and eukaryotes would suggest that this proposed movement of the substrate in the active site is a universal mechanism. Indeed, given the similarity in the active site of GluDH to PheDH and LeuDH (Britton et al, 1993a) it is probable that a similar mechanism also occurs in the wider amino-acid dehydrogenase superfamily.

7.2 Phenylalanine Dehydrogenase

The structure of *B. sphaericus* PheDH in complex with NADH has been determined at medium resolution (2.5 Å). The essential residues involved in the catalysis are conserved between GluDH, LeuDH and PheDH, which suggests that they share a similar catalytic mechanism. However, the residues responsible for recognition of the substrate are varied between these enzymes. In the *C. symbiosum* GluDH, the side chain carboxyl group of the L-glutamate binds to K89 and S380, which make the active site pocket more hydrophilic, and these residues are changed to hydrophobic residues in *B. sphaericus*

PheDH (L51 and V310) and *Rhodococcus sp.* M4 PheDH (A38 and L296) and also in *B. sphaericus* LeuDH (L40 and V294). The interaction of phenylalanine and phenylpyruvate with the PheDHs enzymes from *Rhodococcus sp.* M4 and *Nocardia sp.* 239 is similar, which suggests that the binding of the substrate to *B. sphaericus* PheDH would follow the same pattern. Therefore, phenylalanine was modelled into the active site of *B. sphaericus* PheDH. The distance between the hydride transfer partners in this model is 6.5 Å, which is too far for the transfer to occur. Indeed, in the ternary complex structures of bovine-GluDH, *Rhodococcus sp.* M4 PheDH and *Nocardia sp.* 239 PheDH the nicotinamide C4 to alpha carbon distance is around 4 Å, the van der Waals packing distance, and still not close enough for catalysis, which supports the idea that the substrate moves within the active site during the reaction to be in the suitable position for hydride transfer.

7.3 Cofactor specificity

Determination the binary complex structure of *C. symbiosum* GluDH/NAD⁺ and the ternary complex of *M. smegmatis* GluDH/NADP⁺/2oxoglutarate has allowed the nature of the cofactor specificity to be examined. A comparison between these structures with the structures of other GluDHs (for example the NADP(H) dependent *E. coli* GluDH) has supported the suggestion that the residues at P7 and P8 at the C-terminal end of the β-strand following the GXGXXG/A helix in domain 2 (the Rossmann fold) determine the NADP⁺/NAD⁺ specificity (Oliveira et al, 2012; Sharkey et al, 2013) In *M. smegmatis* GluDH, these residues are Asp265 (P7) and Ser266 (P8), which are in agreement with most of the NADP(H) dependent GluDHs (figure7.8).

The interaction of the 2' phosphate of the cofactor to *M. smegmatis* GluDH is similar to that seen in bovine-GluDH, a dual-cofactor dependent enzyme (Smith et al, 2001). In both enzymes the residues that stabilize the 2'phosphate group are conserved. These residues are Lys137, Lys285 and Ser266 in *M. smegmatis* GluDH and Lys134, Lys295 and Ser276 in bovine- GluDH. In the NAD⁺-dependent *C. symbiosum* GluDH, these residues are substituted by Asn133, Arg285 and Pro262, which make the 2'phosphate binding pocket too small to accommodate this group and also less positively charged.

Structural analysis of *B. sphaericus* PheDH in complex with NADH shows that the interaction of the cofactor with domain II of the enzyme is similar to that observed in *B. sphaericus* LeuDH (Baker et al, 1995). However, the nicotinamide ring binding is quite different between the two proteins. In LeuDH the side-chain hydroxyl groups of Thr150 and Ser147 form hydrogen bonds to the carboxamide moiety of the nicotinamide ring. These two residues are conserved in *B. sphaericus* PheDH as Thr161 and Ser158, but they do not bind to the nicotinamide ring. Instead, Ser157 (equivalent of Pro146 in LeuDH) makes a hydrogen bond to the carboxamide in PheDH. The nicotinamide ring takes the *syn* conformation for the glycosidic bond between the nicotinamide moiety and the associated ribose. This is similar to that observed in all GluDHs and also resembles the conformation for the cofactor seen in *Nocardia sp.* 239 PheDH and *B. sphaericus* LeuDH.

7.4 Further Work

The work described in this thesis has provided further insights into the reaction mechanism of the amino acid dehydrogenase superfamily. However, a number of extra experiments can be proposed that could further explain this enzyme superfamily.

- Higher resolution structure of *B. sphaericus* PheDH (with and without substrates) – Although the first structure of this enzyme has been determined, the medium resolution (and the fact that a number of loops are disordered) has meant that only limited information can be gleaned from this enzyme structure. Despite numerous attempts to produce better diffracting crystals of both the wild type and two site directed mutants, further experiments using different conditions, substrate combinations and perhaps inhibitors could produce a better crystal form.
- Ternary complex of *M. smegmatis* GluDH. The NADP⁺/2oxoglutarate structure determined here has given crucial insights into the reaction mechanism. However, this structure is of a non productive ternary complex. It might be possible to produce crystals of *M. smegmatis* GluDH in complex with NADPH and 2oxoglutarate, as without ammonia, this complex should also be non-productive and thus a single species should be present in the crystal. In such a

complex, the cofactor would be reduced and the substrate oxidised, in the same oxidation states as that seen in the reaction.

- In addition, crystallizations could be attempted for productive complexes of the enzyme, as perhaps the density for the substrates would not be ambiguous, if the crystals diffracted to a similar resolution to those described in this thesis.
- Clarification that the 2-oxoglutarate binds as a gem-diol in the ternary complex of *M. smegmatis* GluDH. Although it is well known that ketones with an electron withdrawing group adjacent to the carbonyl have appreciable amounts of the gem-diol present in aqueous solution (Berg et al, 2002) direct experiments could be undertaken to show this is the case for the structure described here. Firstly, NMR and IR spectroscopy can be used on a solution of 2-oxoglutarate to show the proportion of the gem-diol present. Secondly, it may be possible to use mass spectrometry on the crystals to measure the enzyme-substrate complex and show the presence directly.
- Mutants of GluDH. A number of residues in the active site of GluDH have been produced and characterized, for example D165N for *C. symbiosum* GluDH (Paradisi et al, 2005) and structures of ternary complexes of the equivalent mutants in *M. smegmatis* GluDH would further explain the mechanism of this enzyme.
- New constructs of the Type II NADH: Menaquinone oxidoreductase (NDH-2) of *M. tuberculosis*. Following the successful structure determination of the NDH2 from *Caldalkalibacillus thermarum* (Heikal et al, 2014), this structure can be used to design alternative constructs of NDH2 that might crystallize. As this enzyme is a possible target for novel antimicrobial agents, a structure of the mycobacterial enzyme could be very useful in designing potential inhibitors. This work can be extended to other species, such as *Plasmodium falciparum*, another pathogenic organism where NDH2 has potential to be a target for antimicrobials (Fisher et al, 2009).

References

- Adams MJ, Buehner M, Chandrasekhar K, Ford GC, Hackert ML, Liljas A, Rossmann MG, Smiley IE, Allison WS, Everse J (1973) Structure-function relationships in lactate dehydrogenase. *Proceedings of the National Academy of Sciences* **70**: 1968-1972
- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography* **66**: 213-221
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**: 3389-3402
- Arcus VL, Lott JS, Johnston JM, Baker EN (2006) The potential impact of structural genomics on tuberculosis drug discovery. *Drug Discovery Today* **11**: 28-34
- Asano Y, Nakazawa A (1985) Crystallization of phenylalanine dehydrogenase from *Sporosarcina-urea*. *Agricultural and Biological Chemistry* **49**: 3631-3632
- Asano Y, Nakazawa A, Endo K (1987a) Novel phenylalanine dehydrogenases from *Sporosarcina ureae* and *Bacillus sphaericus*. Purification and characterization. *The Journal of biological chemistry* **262**: 10346-10354
- Asano Y, Nakazawa A, Endo K, Hibino Y, Ohmori M, Numao N, Kondo K (1987b) Phenylalanine dehydrogenase of *Bacillus badius*. Purification, characterization and gene cloning. *European journal of biochemistry / FEBS* **168**: 153-159
- Bailey J, Bell ET, Bell JE (1982) Regulation of bovine glutamate dehydrogenase. The effects of pH and ADP. *The Journal of biological chemistry* **257**: 5579-5583
- Bailey J, Powell L, Sinanan L, Neal J, Li M, Smith T, Bell E (2011) A novel mechanism of V-type zinc inhibition of glutamate dehydrogenase results from disruption of subunit interactions necessary for efficient catalysis. *The FEBS journal* **278**: 3140-3151
- Baker PJ, Britton KL, Engel PC, Farrants GW, Lilley KS, Rice DW, Stillman TJ (1992a) Subunit assembly and active site location in the structure of glutamate dehydrogenase. *Proteins* **12**: 75-86
- Baker PJ, Britton KL, Engel PC, Farrants GW, Lilley KS, Rice DW, Stillman TJ (1992b) Subunit assembly and active-site location in the structure of glutamate dehydrogenase. *Proteins-Structure Function and Genetics* **12**: 75-86
- Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ (1992c) Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold. Implications for nucleotide specificity. *Journal of molecular biology* **228**: 662-671
- Baker PJ, Turnbull AP, Sedelnikova SE, Stillman TJ, Rice DW (1995) A role for quaternary structure in the substrate specificity of leucine dehydrogenase. *Structure* **3**: 693-705

Baker PJ, Waugh ML, Wang XG, Stillman TJ, Turnbull AP, Engel PC, Rice DW (1997) Determinants of substrate specificity in the superfamily of amino acid dehydrogenases. *Biochemistry* **36**: 16109-16115

Banerjee S, Schmidt T, Fang J, Stanley CA, Smith TJ (2003) Structural studies on ADP activation of mammalian glutamate dehydrogenase and the evolution of regulation. *Biochemistry* **42**: 3446-3456

Berg JM, Tymoczko JL, Stryer L, Clarke ND *Biochemistry*, 2002. *Stryer, Lubert*

Bhuiya MW, Sakuraba H, Ohshima T, Imagawa T, Katunuma N, Tsuge H (2005) The first crystal structure of hyperthermostable NAD-dependent glutamate dehydrogenase from *Pyrobaculum islandicum*. *Journal of molecular biology* **345**: 325-337

Birktoft J, Banaszak L (1984) Structure-function relationships among nicotinamide-adenine dinucleotide dependent oxidoreductases. *Peptide and protein reviews* **4**: 1-47

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids research* **7**: 1513-1523

Blumberg H, Burman W, Chaisson R, Daley C, Etkind S, Friedman L, Fujiwara P, Grzemska M, Hopewell P, Iseman M (2003) American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. *American journal of respiratory and critical care medicine* **167**: 603

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248-254

Braunstein AE (1957) [Principal ways of assimilation & dissimilation of nitrogen in animals]. *Advances in enzymology and related subjects of biochemistry* **19**: 335-389

Britton KL, Baker PJ, Engel PC, Rice DW, Stillman TJ (1993a) Evolution of substrate diversity in the superfamily of amino acid dehydrogenases: prospects for rational chiral synthesis. *Journal of Molecular Biology* **234**: 938-945

Britton KL, Baker PJ, Engel PC, Rice DW, Stillman TJ (1993b) Evolution of substrate diversity in the subfamily of amino acid dehydrogenases, prospects for rational chiral synthesis. *Journal of Molecular Biology* **234**: 938-945

Britton KL, Baker PJ, Rice DW, Stillman TJ (1992) Structural relationship between the hexameric and tetrameric family of glutamate dehydrogenases. *European journal of biochemistry / FEBS* **209**: 851-859

Britton KL, Yip KS, Sedelnikova SE, Stillman TJ, Adams MW, Ma K, Maeder DL, Robb FT, Tolliday N, Vetriani C, Rice DW, Baker PJ (1999) Structure determination of the glutamate dehydrogenase from the hyperthermophile *Thermococcus litoralis* and its comparison with that from *Pyrococcus furiosus*. *Journal of molecular biology* **293**: 1121-1132

Brunhuber NM, Blanchard JS (1994) The biochemistry and enzymology of amino acid dehydrogenases. *Critical reviews in biochemistry and molecular biology* **29**: 415-467

Brunhuber NMW, Thoden JB, Blanchard JS, Vanhooke JL (2000) Rhodococcus L-phenylalanine dehydrogenase: Kinetics, mechanism, and structural basis for catalytic specificity. *Biochemistry* **39**: 9174-9187

Burgi HB, Dunitz JD (1983) From Crystal Statics to Chemical-Dynamics. *Accounts of Chemical Research* **16**: 153-161

Chavez S, Candau P (1991) An NAD-specific glutamate dehydrogenase from cyanobacteria. Identification and properties. *FEBS letters* **285**: 35-38

Chen S, Engel PC (2009) Efficient screening for new amino acid dehydrogenase activity: directed evolution of *Bacillus sphaericus* phenylalanine dehydrogenase towards activity with an unsaturated non-natural amino acid. *Journal of biotechnology* **142**: 127-134

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* **66**: 12-21

Clayden G, Warren W, Greeves N, Wothers P. Organic Chemistry, 2001. Oxford University Press.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaiia F, Badcock K, Basham D, Brown D, Chillingworth T, Conner R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence (vol 393, pg 537, 1998). *Nature* **396**: 190-198

Collaborative CP (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica Section D, Biological crystallography* **50**: 760

Congreve M, Murray CW, Blundell TL (2005) Keynote review: Structural biology and drug discovery. *Drug Discovery Today* **10**: 895-907

Corman L, Kaplan NO (1967) Kinetic studies of dogfish liver glutamate dehydrogenase with diphosphopyridine nucleotide and the effect of added salts. *The Journal of biological chemistry* **242**: 2840-2846

Crowther Rt, Blow D (1967) A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallographica* **23**: 544-548

Dawson R C, Elliott, D. C, Elliott, WH, and Jones, KM (1986) *Data for Biochemical Research*: Clarendon Press, Oxford.

Deboer L, Vanrijssel M, Euverink GJ, Dijkhuizen L (1989) Purification, Characterization and Regulation of a Monomeric L-Phenylalanine Dehydrogenase from the Facultative Methylophilic *Nocardia* Sp-239. *Archives of Microbiology* **153**: 12-18

Dooley KC (1992) Enzymatic method for phenylketonuria screening using phenylalanine dehydrogenase. *Clinical biochemistry* **25**: 271-275

Drenth J (2007) *Principles of protein X-ray crystallography*: Springer.

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta crystallographica Section D, Biological crystallography* **66**: 486-501

Fahien LA, Wiggert BO, Cohen PP (1965) Crystallization and Kinetic Properties of Glutamate Dehydrogenase from Frog Liver. *The Journal of biological chemistry* **240**: 1083-1090

Fisher HF, Conn EE, Vennesland B, Westheimer FH (1953) The enzymatic transfer of hydrogen. I. The reaction catalyzed by alcohol dehydrogenase. *The Journal of biological chemistry* **202**: 687-697

Fisher N, Warman AJ, Ward SA, Biagini GA (2009) Type II NADH: Quinone Oxidoreductases of Plasmodium Falciparum and Mycobacterium Tuberculosis: Kinetic and High-Throughput Assays. *Methods in Enzymology, Vol 456* **456**: 303-320

Floss HG, Yu TW (2005) Rifamycin-mode of action, resistance, and biosynthesis. *Chemical reviews* **105**: 621-632

Galagan JE (2014) Disease Mechanisms Genomic Insights into Tuberculosis. *Nature Reviews Genetics* **15**: 307-320

Gerstein M, Anderson BF, Norris GE, Baker EN, Lesk AM, Chothia C (1993) Domain closure in lactoferrin: Two hinges produce a see-saw motion between alternative close-packed interfaces. *Journal of Molecular Biology* **234**: 357-372

Gillespie SH (2002) Evolution of drug resistance in Mycobacterium tuberculosis: Clinical and molecular perspective. *Antimicrobial Agents and Chemotherapy* **46**: 267-274

Goldin BR, Frieden C (1971) L-Glutamate dehydrogenases. *Curr Top Cell Regul* **4**: 77-117

Höfner E, Müller T, Palumbo K, Patek M, Brocker M, Krüger R, Burkovski A (2009) A game with many players: Control of *gdh* transcription in *Corynebacterium glutamicum*. *J Biotechnol* **142**: 114-122

Hansler E, Müller T, Palumbo K, Patek M, Brocker M, Kramer R, Burkovski A (2009) A game with many players: control of *gdh* transcription in *Corynebacterium glutamicum*. *Journal of biotechnology* **142**: 114-122

Harper CJ, Hayward D, Kidd M, Wiid I, van Helden P (2010) Glutamate dehydrogenase and glutamine synthetase are regulated in response to nitrogen availability in *Mycobacterium smegmatis*. *BMC microbiology* **10**: 138

Hatefi Y (1985) The Mitochondrial Electron-Transport and Oxidative-Phosphorylation System. *Annual Review of Biochemistry* **54**: 1015-1069

Heikal A, Nakatani Y, Dunn E, Weimar MR, Day CL, Baker EN, Lott JS, Sazanov LA, Cook GM (2014) Structure of the bacterial type II NADH dehydrogenase: a monotopic membrane protein with an essential role in energy generation. *Molecular microbiology* **91**: 950-964

Hol W, Van Duijnen PT, Berendsen H (1978) The alpha-helix dipole and the properties of proteins. *nature* **273**: 443

- Hornby DP, Engel PC (1984) Characterization of *Peptostreptococcus asaccharolyticus* glutamate dehydrogenase purified by dye-ligand chromatography. *Journal of general microbiology* **130**: 2385-2394
- Hornby DP, Engel PC, Hatanaka S (1983) Beef liver glutamate dehydrogenase: a study of the oxidation of various alternative amino acid substrates retaining the correct spacing of the two carboxylate groups. *The International journal of biochemistry* **15**: 495-500
- Hudson RC, Daniel RM (1993) L-glutamate dehydrogenases: distribution, properties and mechanism. *Comparative biochemistry and physiology B, Comparative biochemistry* **106**: 767-792
- Hummel W, Kula MR (1989) Dehydrogenases for the synthesis of chiral compounds. *European journal of biochemistry / FEBS* **184**: 1-13
- Hummel W, Weiss N, Kula M-R (1984) Isolation and characterization of a bacterium possessing L-phenylalanine dehydrogenase activity. *Archives of Microbiology* **137**: 47-52
- Kana B, Machowski E, Schechter N, Teh J, Rubin H, Mizrahi V (2009) Electron transport and respiration in mycobacteria. *Parish T, Brown A, eds*: 35-64
- Kana BD, Weinstein EA, Avarbock D, Dawes SS, Rubin H, Mizrahi V (2001) Characterization of the *cydAB*-encoded cytochrome bd oxidase from *Mycobacterium smegmatis*. *Journal of bacteriology* **183**: 7076-7086
- Kantardjieff KA, Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein science : a publication of the Protein Society* **12**: 1865-1871
- Karakousis PC, Bishai WR, Dorman SE (2004) *Mycobacterium tuberculosis* cell envelope lipids and the host immune response. *Cellular microbiology* **6**: 105-116
- Knapp S, de Vos WM, Rice D, Ladenstein R (1997) Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. *Journal of molecular biology* **267**: 916-932
- Laemmli U (1970) Most commonly used discontinuous buffer system for SDS electrophoresis. *nature* **227**: 680-685
- Lapatto R, Blundell T, Hemmings A, Overington J, Wilderspin A, Wood S, Merson JR, Whittle PJ, Danley DE, Geoghegan KF, Hawrylik SJ, Lee SE, Scheld KG, Hobart PM (1989) X-Ray-Analysis of Hiv-1 Proteinase at 2.7 Å Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **342**: 299-302
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck, a program to check the stereochemical quality of protein structure. *Journal of applied crystallography* **26**: 283-291
- Lesk AM, Chothia C (1984) Mechanisms of domain closure in proteins. *Journal of Molecular Biology* **174**: 175-191
- Li C, Li M, Chen P, Narayan S, Matschinsky FM, Bennett MJ, Stanley CA, Smith TJ (2011) Green tea polyphenols control dysregulated glutamate dehydrogenase in transgenic mice by hijacking the ADP activation site. *The Journal of biological chemistry* **286**: 34164-34174

- Li M, Smith CJ, Walker MT, Smith TJ (2009) Novel inhibitors complexed with glutamate dehydrogenase: allosteric regulation by control of protein dynamics. *The Journal of biological chemistry* **284**: 22988-23000
- Madkour MM (2001) *Madkour's brucellosis*: Springer Verlag.
- Malani PN (2010) Mandell, Douglas, and Bennett's principles and practice of infectious diseases. *JAMA* **304**: 2067-2071
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *Journal of applied crystallography* **40**: 658-674
- Miesel L, Weisbrod TR, Marcinkeviciene JA, Bittman R, Jacobs WR, Jr. (1998) NADH dehydrogenase defects confer isoniazid resistance and conditional lethality in *Mycobacterium smegmatis*. *Journal of bacteriology* **180**: 2459-2467
- Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH, Wlodawer A (1989) Structure of Complex of Synthetic Hiv-1 Protease with a Substrate-Based Inhibitor at 2.3-Å Resolution. *Science* **246**: 1149-1152
- Misono H, Yonezawa J, Nagata S, Nagasaki S (1989) Purification and characterization of a dimeric phenylalanine dehydrogenase from *Rhodococcus maris* K-18. *Journal of bacteriology* **171**: 30-36
- Mitchell E, Kuhn P, Garman E (1999) Demystifying the synchrotron trip: a first time user's guide. *Structure* **7**: R111-R122
- Moon K, Smith EL (1973) Sequence of bovine liver glutamate dehydrogenase. 8. Peptides produced by specific chemical cleavages; the complete sequence of the protein. *The Journal of biological chemistry* **248**: 3082-3088
- Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica Section D, Biological crystallography* **67**: 355-367
- Nakamura K, Fujii T, Kato Y, Asano Y, Cooper AJ (1996) Quantitation of L-Amino Acids by Substrate Recycling between an Aminotransferase and a Dehydrogenase: Application to the Determination of L-Phenylalanine in Human Blood. *Analytical biochemistry* **234**: 19-22
- Nakasako M, Fujisawa T, Adachi S, Kudo T, Higuchi S (2001) Large-scale domain movements and hydration structure changes in the active-site cleft of unligated glutamate dehydrogenase from *Thermococcus profundus* studied by cryogenic X-ray crystal structure analysis and small-angle X-ray scattering. *Biochemistry* **40**: 3069-3079
- Ohshima T, Misono H, Soda K (1978) Properties of Crystalline Leucine Dehydrogenase from *Bacillus-Sphaericus*. *Journal of Biological Chemistry* **253**: 5719-5725
- Ohshima T, Nagata S, Soda K (1985) Purification and Characterization of Thermostable Leucine Dehydrogenase from *Bacillus-Stearothermophilus*. *Archives of Microbiology* **141**: 407-411
- Ohshima T, Nishida N, Bakthavatsalam S, Kataoka K, Takada H, Yoshimura T, Esaki N, Soda K (1994) The purification, characterization, cloning and sequencing of the gene for a

halostable and thermostable leucine dehydrogenase from *Thermoactinomyces intermedius*. *European journal of biochemistry / FEBS* **222**: 305-312

Ohshima T, Soda K (1990) Biochemistry and biotechnology of amino acid dehydrogenases. *Advances in biochemical engineering/biotechnology* **42**: 187-209

Ohshima T, Takada H, Yoshimura T, Esaki N, Soda K (1991) Distribution, purification, and characterization of thermostable phenylalanine dehydrogenase from thermophilic actinomycetes. *Journal of bacteriology* **173**: 3943-3948

Oliveira T, Panjikar S, Carrigan JB, Hamza M, Sharkey MA, Engel PC, Khan AR (2012) Crystal structure of NAD⁺-dependent *Peptoniphilus asaccharolyticus* glutamate dehydrogenase reveals determinants of cofactor specificity. *Journal of structural biology* **177**: 543-552

Palomino JC, Leão SC, Ritacco V (2007) Tuberculosis 2007; from basic science to patient care.

Paradisi F, Collins S, Maguire AR, Engel PC (2007) Phenylalanine dehydrogenase mutants: efficient biocatalysts for synthesis of non-natural phenylalanine derivatives. *Journal of biotechnology* **128**: 408-411

Paradisi F, Woolfson R, Geoghegan KF, Engel PC (2005) Identification of the residue responsible for catalysing regeneration of activity in the inactive glutamate dehydrogenase mutant D165N. *FEBS letters* **579**: 2830-2832

Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research* **36**: 2295-2300

Peterson PE, Smith TJ (1999) The structure of bovine glutamate dehydrogenase provides insights into the mechanism of allostery. *Structure* **7**: 769-782

Rhodes G (2010) *Crystallography made crystal clear: a guide for users of macromolecular models*: Academic press.

Rife JE, Cleland WW (1980a) Determination of the chemical mechanism of glutamate dehydrogenase from pH studies. *Biochemistry* **19**: 2328-2333

Rife JE, Cleland WW (1980b) Kinetic mechanism of glutamate dehydrogenase. *Biochemistry* **19**: 2321-2328

Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* **250**: 194-199

Rozwarski DA, Grant GA, Barton DH, Jacobs WR, Jr., Sacchettini JC (1998) Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*. *Science* **279**: 98-102

Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual (3-volume set)*, Vol. 999: Cold spring harbor laboratory press Cold Spring Harbor, New York:.

Sarada KV, Rao NA, Venkatasubramanian TA (1980) Isolation and characterisation of glutamate dehydrogenase from *Mycobacterium smegmatis* CDC 46. *Biochimica et biophysica acta* **615**: 299-308

Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 12989-12994

Schwacha A, Bender RA (1993) The product of the *Klebsiella aerogenes* nac (nitrogen assimilation control) gene is sufficient for activation of the hut operons and repression of the gdh operon. *Journal of bacteriology* **175**: 2116-2124

Seah SY, Britton KL, Rice DW, Asano Y, Engel PC (2002) Single amino acid substitution in *Bacillus sphaericus* phenylalanine dehydrogenase dramatically increases its discrimination between phenylalanine and tyrosine substrates. *Biochemistry* **41**: 11390-11397

Seah SY, Britton KL, Rice DW, Asano Y, Engel PC (2003) Kinetic analysis of phenylalanine dehydrogenase mutants designed for aliphatic amino acid dehydrogenase activity with guidance from homology-based modelling. *European journal of biochemistry / FEBS* **270**: 4628-4634

Sharkey MA, Oliveira TF, Engel PC, Khan AR (2013) Structure of NADP(+)-dependent glutamate dehydrogenase from *Escherichia coli*--reflections on the basis of coenzyme specificity in the family of glutamate dehydrogenases. *The FEBS journal* **280**: 4681-4692

Smith TJ, Peterson PE, Schmidt T, Fang J, Stanley CA (2001) Structures of bovine glutamate dehydrogenase complexes elucidate the mechanism of purine regulation. *Journal of molecular biology* **307**: 707-720

Smith TJ, Schmidt T, Fang J, Wu J, Siuzdak G, Stanley CA (2002) The structure of apo human glutamate dehydrogenase details subunit communication and allostery. *Journal of molecular biology* **318**: 765-777

Srinivasan R, Fisher HF (1985) Reversible reduction of an alpha-imino acid to an alpha-amino acid catalyzed by glutamate dehydrogenase: effect of ionizable functional groups. *Biochemistry* **24**: 618-622

Stillman TJ, Baker PJ, Britton KL, Rice DW (1993) Conformational flexibility in glutamate dehydrogenase. Role of water in substrate recognition and catalysis. *Journal of Molecular Biology* **234**: 1131-1139

Stillman TJ, Migueis AM, Wang XG, Baker PJ, Britton KL, Engel PC, Rice DW (1999) Insights into the mechanism of domain closure and substrate specificity of glutamate dehydrogenase from *Clostridium symbiosum*. *Journal of Molecular Biology* **285**: 875-885

Struck J, Jr., Sizer IW (1960) The substrate specificity of glutamic acid dehydrogenase. *Archives of biochemistry and biophysics* **86**: 260-266

Stryer L (1995) *Biochemistry*, 1995. *Chapter 14*: 385

Syed SE, Engel PC, Parker DM (1991) Functional studies of a glutamate dehydrogenase with known three-dimensional structure: steady-state kinetics of the forward and reverse reactions catalysed by the NAD(+)-dependent glutamate dehydrogenase of *Clostridium symbiosum*. *Biochimica et biophysica acta* **1115**: 123-130

Takada H, Yoshimura T, Ohshima T, Esaki N, Soda K (1991) Thermostable phenylalanine dehydrogenase of the *Thermoactinomyces-intermedius*, cloning, expression, and sequencing of the its gene. *Journal of Biochemistry* **109**: 371-376

- Teh JS, Yano T, Rubin H (2007) Type II NADH: menaquinone oxidoreductase of *Mycobacterium tuberculosis*. *Infectious disorders drug targets* **7**: 169-181
- Teller JK, Smith RJ, McPherson MJ, Engel PC, Guest JR (1992) The glutamate dehydrogenase gene of *Clostridium symbiosum*. Cloning by polymerase chain reaction, sequence analysis and over-expression in *Escherichia coli*. *European journal of biochemistry / FEBS* **206**: 151-159
- Turnbull AP, Ashford SR, Baker PJ, Rice DW, Rodgers FH, Stillman TJ, Hanson RL (1994) Crystallization and quaternary structure analysis of the NAD(+)-dependent leucine dehydrogenase from *Bacillus sphaericus*. *Journal of molecular biology* **236**: 663-665
- Vanhooke JL, Thoden JB, Brunhuber NM, Blanchard JS, Holden HM (1999) Phenylalanine dehydrogenase from *Rhodococcus* sp. M4: high-resolution X-ray analyses of inhibitory ternary complexes reveal key features in the oxidative deamination mechanism. *Biochemistry* **38**: 2326-2339
- Varghese JN (1999) Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Development Research* **46**: 176-196
- Veronese FM, Nyc JF, Degani Y, Brown DM, Smith EL (1974) Nicotinamide adenine dinucleotide-specific glutamate dehydrogenase of *Neurospora*. I. Purification and molecular properties. *The Journal of biological chemistry* **249**: 7922-7928
- Vilcheze C, Jacobs WR (2007) The mechanism of isoniazid killing: Clarity through the scope of genetics. *Annual Review of Microbiology* **61**: 35-50
- Vilcheze C, Weisbrod TR, Chen B, Kremer L, Hazbon MH, Wang F, Alland D, Sacchettini JC, Jacobs WR, Jr. (2005) Altered NADH/NAD⁺ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. *Antimicrobial agents and chemotherapy* **49**: 708-720
- Warman AJ, Rito TS, Fisher NE, Moss DM, Berry NG, O'Neill PM, Ward SA, Biagini GA (2013) Antitubercular pharmacodynamics of phenothiazines. *Journal of Antimicrobial Chemotherapy* **68**: 869-880
- Weinstein EA, Yano T, Li LS, Avarbock D, Avarbock A, Helm D, McColm AA, Duncan K, Lonsdale JT, Rubin H (2005) Inhibitors of type IINADH :-menaquinone oxidoreductase represent a class of antitubercular drugs. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 4548-4553
- Werner C, Stubbs MT, Krauth-Siegel RL, Klebe G (2005) The crystal structure of *Plasmodium falciparum* glutamate dehydrogenase, a putative target for novel antimalarial drugs. *Journal of molecular biology* **349**: 597-607
- Wilson BA, Salyers AA, Whitt DD, Winkler ME (2011) *Bacterial pathogenesis: a molecular approach*: American Society for Microbiology (ASM).
- Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**: 186-190
- World-Health-Organization (2013) Global tuberculosis report 2013.

Yano T, Li LS, Weinstein E, Teh JS, Rubin H (2006) Steady-state kinetics and inhibitory action of antitubercular phenothiazines on Mycobacterium tuberculosis type-II NADH-menaquinone oxidoreductase (NDH-2). *Journal of Biological Chemistry* **281**: 11456-11463

Yip KS, Stillman TJ, Britton KL, Artymiuk PJ, Baker PJ, Sedelnikova SE, Engel PC, Pasquo A, Chiaraluce R, Consalvi V, et al. (1995) The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**: 1147-1158

You KS, Arnold LJ, Allison WS, Kaplan NO (1978) Enzyme Stereospecificities for Nicotinamide Nucleotides. *Trends in Biochemical Sciences* **3**: 265-268

Zocher K, Fritz-Wolf K, Kehr S, Fischer M, Rahlfs S, Becker K (2012) Biochemical and structural characterization of Plasmodium falciparum glutamate dehydrogenase 2. *Molecular and biochemical parasitology* **183**: 52-62

AppendixI: X-ray crystallography Theory

1.0 X-Rays diffraction

X-rays radiation can be considered as electromagnetic waves with an oscillating electrical field. When X-rays interact with an atom at particular frequency the orbiting electrons start to oscillate emitting a radiation with the same identical frequency as the incident X-ray, but are 180° out of phase with it. This is known as coherent scattering. However, X-rays can also collide with atoms in a different manner, giving rise to incoherent scattering. In this case, X-rays interact with loosely bound electrons leading to the deviation of the X-ray from its initial direction. This causes the X-ray to lose energy and therefore its wavelength is altered. In X-ray crystallography incoherent scattering is generally ignored because of the fact that the sum of cooperative coherent scattering becomes bigger than the incoherent contribution.

1.1 Bragg's Law

In 1913 Sir W.L.Bragg derived a simple equation to describe why the set of lattice planes in a crystal appear to produce a reflection at a particular angle of incidence. This is Bragg's Law and can be expressed as:

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

d is the distance in angstroms that separate planes with indices h,k,l ; θ is the angle between incident and diffracted rays; n is an integer and λ is the wavelength of the incident X-ray beam. Constructive interference happens when the distance $2d\sin\theta$ is equal to an integral number of wavelengths because the emitted waves will be in phase and will contribute to a diffraction spots. If the X-rays strike at an angle that does not meet the Bragg conditions, they will be out of phase and interfere destructively, and so not produce a reflection. The intensity of each diffracted beam depends on how much electron density is sampled by the set of planes from which the diffracted X-rays originate.

1.2 Scattering by an atom

Clouds of electrons density surrounding the atoms interact with the incident X-ray beam. Scattering depends on the number of electrons and their distribution in the cloud. The atomic scattering factor f is the total wave scattered by an atom

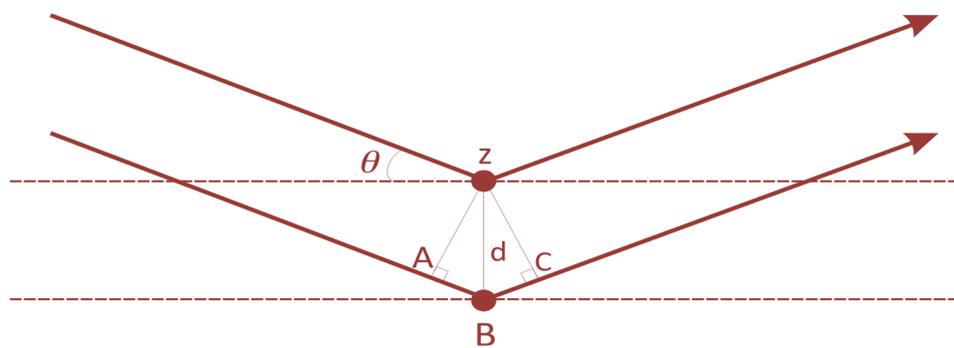


Figure 1.0. Diagram illustrating of Bragg's Law. $AB = d \sin\theta = BC$, the path length difference = $2d \sin\theta$, which must an integral number of wavelength ($n\lambda$) for constructive interference.

integrated over the volume of the atom:

$$f = \int_r \rho(\mathbf{r}) \exp(2\pi \mathbf{r} \cdot \mathbf{S}) d\mathbf{r}$$

where \mathbf{r} is the position of the electron in the cloud from the centre of the atom, $\rho(\mathbf{r})$ is the electron density at that position and vector \mathbf{S} is the direction of the scattered wave.

1.3 Scattering by a unit cell

The total scattering by n atoms at position \mathbf{r}_j ($j=1, 2, 3, 4, \dots, n$) in an isolated unit cell is the sum of the scattering from single atoms in the unit cell:

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

where $F(\mathbf{S})$ is the atomic structure factor, and is a function of the electron density distribution within the unit cell that can be stated as continuous.

1.4 Scattering by a crystal

The total wave scattered by a crystal $K(\mathbf{S})$ is the sum of overall the unit cells scattering within the crystal.

$$K(S) = F(S) \times \sum_{t=0}^{n_1} \exp(2\pi i t a \cdot S) \cdot \sum_{u=0}^{n_2} \exp(2\pi i u b \cdot S) \cdot \sum_{v=0}^{n_3} \exp(2\pi i v c \cdot S)$$

Where a, b and c are the axial directions in the crystal and n_1, n_2 and n_3 represent the number of unit cells along those axial directions. The integers t, u and v correspond to the translation steps relating the origins of the unit cells along these vectors. Due to the large number of the unit cells and because their scattering vectors are pointing in different directions, a crystal only scatters X-rays in phase if the Laue conditions $a \cdot S = h$, $b \cdot S = k$ and $c \cdot S = l$ are met where h, k and l are integers.

Diffraction from a crystal consists of reflections with reciprocal lattice indices of h, k , and l , each receiving a contribution from every atom in the unit cell.

1.5 The Ewald Sphere

The diffraction pattern of a crystal lattice can itself be considered as a lattice- the reciprocal lattice. The dimensions of the reciprocal lattice are inversely proportional to these of the crystal lattice. A graphical representation called the Ewald sphere can be used to explain the direction of particular diffracted X-rays

that fulfill Bragg's Law. The Ewald sphere has a radius $1/\lambda$, with the crystal at the center. The perimeter of the sphere represents reciprocal space where a diffraction spot will be generated when Bragg's Law is satisfied. By rotation the crystal in the X-ray beam, various lattice points $(h\ k\ l)$ contact the surface of sphere and the reflection can be measured.

2.0 Calculation of Electron Density

If x , y and z are fractional coordinates in the unit cell in which h , k and l are the indices for a reflection in the reciprocal lattice and V in the volume of the cell, the structure factor $\mathbf{F}(\mathbf{S})$ can be written as $\mathbf{F}(hkl)$:

$$\mathbf{F}(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp(2\pi i(hx + ky + lz)) dx dy dz$$

The electron density for every position x , y , z in the unit cell $\rho(xyz)$, is given by the Fourier transform of $\mathbf{F}(hkl)$ and can be written:

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

$\mathbf{F}(hkl)$ is a wave with amplitude of $|\mathbf{F}(hkl)|$ and the phase of $\alpha(hkl)$ so that:

$$\mathbf{F}(hkl) = |F(hkl)| \exp(i\alpha(hkl))$$

All of the $|\mathbf{F}(hkl)|$ can be measured experimentally because the intensity of any given reflection within the diffraction pattern is proportional to the square of its structure factor amplitude. The phase information should be measured because it is not possible to determine it directly from the experimental data. This is what is known in x-ray crystallography as the phase problem. The phase problem can be solved using several techniques involving isomorphous replacement, multiple-wavelength anomalous dispersion and molecular replacement.

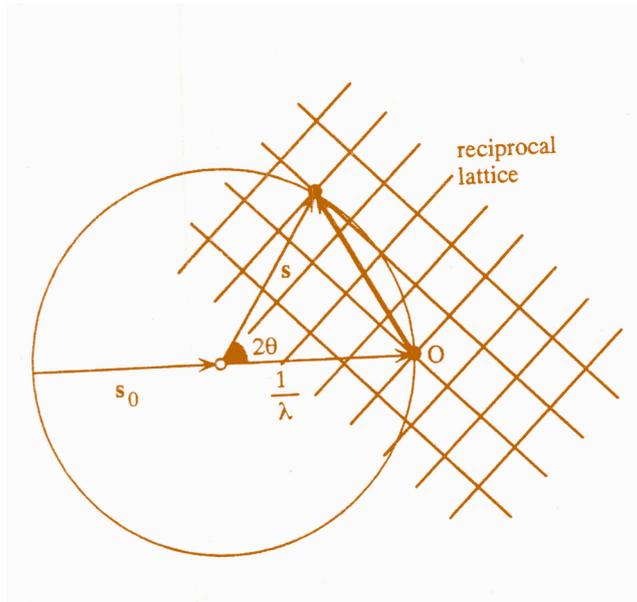


Figure 1.1. Schematic represents of the Ewald sphere. The radius is $1/\lambda$, origin of the reciprocal lattice is O, s_0 indicates the direction of the incoming beam and s indicates the direction of the scattered beam. This scheme has been adapted from (Drenth, 2007).

