

University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

**STRUCTURES AND FUNCTION
OF
5'-FLUORODEOXYADENOSINE SYNTHASE**

Andrew Robert McEwan

**A thesis submitted to
The School of Chemistry of
The University of St. Andrews
for The Degree of Doctor of Philosophy**

June 2006



CONTENTS	PAGES
I Front Cover	i
II List of contents	ii
III Abstract	vii
IV Declarations	viii
V Copyright	ix
VI Acknowledgements	x
VII Dedication	xi
VIII Abbreviations	xii
IX List of Figures	xvii
X List of Tables	xxvii
 Chapter 1	 1
Introduction	
 1.1 Fluorine in Nature	 2
1.2 Importance of fluorinated compounds	6
1.3 Engineered fluorinating enzymes	10
1.4 Fluorinated products in bacteria	12
1.4.1 Actinomycetes	12
1.4.2 <i>Streptomyces cattleya</i>	12
1.4.3 An overview of the fluorinase pathway	12
1.4.3.1 <i>The fluorinase enzyme and pathway</i>	12
1.4.3.2 <i>The PNPase enzyme</i>	14
1.4.3.3 <i>The isomerase enzyme</i>	15
1.4.3.4 <i>An aldolase enzyme</i>	16
1.4.3.5 <i>The aldehyde dehydrogenase</i>	16
1.4.3.6 <i>The pyridoxal phosphate dependent transaldolase</i>	16
1.4.4 The flA gene cluster	17
1.5 The fluorinase (FDAS) enzyme	19
1.5.1 Identification of FDAS	19
1.5.2 Gene sequence and protein sequence	19

1.5.3	Protein structure	20
1.5.4	Structure of substrate in active site	22
1.5.5	Structure of product in active site	24
1.5.6	Reaction mechanism	25
1.6	Enzymatic chlorination	28
1.6.1	Haloperoxidases and perhydrolases	28
1.6.2	Perhydrolases	30
1.6.3	Regioselective chlorination	31
1.6.4	Non-haem FeII α -ketoglutarate and O ₂ dependent halogenase	33
1.7	Aldolases and enzymatic aldol condensation	35
1.7.1	The aldol reaction	35
1.7.2	Aldolases	36
1.7.2.1	<i>DHAP aldolase</i>	37
1.7.2.2	<i>Pyruvate and PEP dependent aldolases</i>	38
1.7.2.3	<i>Glycine dependent aldolases</i>	39
1.7.2.4	<i>DERA</i>	39
1.7.2.5	<i>Class II aldolases</i>	40
Chapter 2		41
Structural Studies into 5'Fluorodeoxyadenosine Synthase; A Novel Naturally Occurring Fluorinating Enzyme		
2.1	SUMMARY	42
2.2	INTRODUCTION	43
2.2.1	FDAS as a chlorinase	43
2.2.2	Substrate specificity	46
2.3	EXPERIMENTAL AIMS	48
2.4	MATERIALS AND METHODS	49
2.4.1	Cloning FIA into pEHISTEV vector	49
2.4.2	Overexpression of recombinant 6HIS-TEV-FDAS	50
2.4.3	Typical purification of recombinant FDAS	50
2.4.4	Removal of adenosine from FDAS	51
2.4.5	Inhibitors, substrates and analogues	52
2.4.6	Crystallization	54
2.4.6.1	<i>Co-crystallization</i>	54
2.4.6.2	<i>Sparse matrix screening of apo-enzyme</i>	55
2.4.7	Data collection and processing, structure solution, refinement and validation	56
2.4.7.1	<i>Data collection of C222₁ crystals</i>	56
2.4.7.2	<i>Data processing</i>	57
2.4.7.3	<i>Structure solution</i>	58
2.4.7.4	<i>Structure refinement</i>	59
2.3.8	High Performance Liquid Chromatography (HPLC)	60
2.4.9	Circular Dichroism Spectroscopy	61
2.4.10	Isothermal Calorimetry (ITC)	61

2.4.11	Site Directed Mutagenesis	61
2.5	RESULTS AND DISCUSSION	64
2.5.1	Initial co-crystallization of 5'-chloro-5'-deoxyadenosine with FDAS	64
2.5.2	HPLC analysis of the co-crystals and identification adenosine contaminant in purified recombinant FDAS	67
2.5.3	Removal of adenosine contamination	68
2.5.4	Co-crystallization and X-ray crystal structure of FDAS complexed with 5'-chloro-5'-deoxyadenosine (5'-CIDA)	74
2.5.5	Co-crystallization of FDAS with 5'-deoxy-5'-[N-methyl-N-(γ -amino- γ -carboxypropyl)]-adenosine and chloride ion	75
2.5.6	Co-crystallization of FDAS complexed with 5'-deoxyadenosine	81
2.5.7	Co-crystallization of FDAS complexed with β -D-erythrofuranosyl adenosine and chloride ion	84
2.5.8	Co-crystallization of FDAS complexed with 5'-fluoro-2',5'-dideoxyadenosine	87
2.5.9	Co-crystallization and X-ray crystal structure of FDAS complexed with 2'-deoxyadenosine	90
2.5.10	Crystallization and structure solution of apo-FDAS	93
2.5.11	Overall structure comparisons	96
2.5.12	Isothermal Calorimetry	100
2.6	CONCLUSIONS	106
2.7	FUTURE WORK	109
Chapter 3		110
An Investigation into Sso6206; A Highly Conserved Protein in Archaea with Unknown Function.		
3.1	SUMMARY	111
3.2	INTRODUCTION	112
3.2.1	Sso 6206	112
3.3	MATERIALS AND METHODS	115
3.3.1	Cloning of 6206 gene	115
3.3.2	Typical overexpression and purification of Sso6206	116
		117
3.3.3	Circular dichroism spectroscopy	117
3.3.4	Crystallization of native Sso6206	118
3.3.5	Typical overexpression and purification of SeMet 6206	118
		119
3.3.6	Crystallization of SeMet derivative of Sso6206	119
3.3.7	Heavy atom replacement	121
3.3.8	Native crystal data collection	121
3.3.9	SeMet derivative crystal data collection	122
3.3.10	Heavy atom derivative data collection	123
3.3.11	Sulfur data collection	123

3.3.12	Data processing	123
3.4	RESULTS AND DISCUSSION	124
3.4.1	Initial purification and crystallisation	124
3.4.2	Native Sso6206 data	126
3.4.3	Phasing using heavy atom soaks	128
3.4.4	Overexpression and crystallization of SeMet Sso6206	132
3.4.5	X-ray data collection of SeMet Sso6206	134
3.4.6	Sulfur phasing	137
3.5	CONCLUSIONS	139
3.6	FUTURE WORK	140
Chapter 4		141
An Investigation into KDPPGal Aldolase; A Stereospecific Pyruvate Dependent Class I Aldolase		
4.1	SUMMARY	142
4.2	INTRODUCTION	142
4.3	EXPERIMENTAL AIMS	149
4.4	MATERIALS AND METHODS	149
4.4.1	Transformation and purification	149
4.4.2	Crystallization	151
4.4.3	Compound library	151
4.4.4	Co-crystallization	152
4.4.6	Ligand soaking	153
4.4.7	Data collection and processing	153
4.5	RESULTS AND DISCUSSION	154
4.5.1	Purification and crystallization	154
4.5.2	Crystal structure of KDPPGal aldolase and pyruvate	155
4.5.4	Co-crystal structure of KDPPGal aldolase and glyceraldehyde-3-phosphate	161
4.5.5	Co-crystal structure of KDPPGal aldolase and KDPPGal	164
4.6	CONCLUSIONS	170
4.7	FUTURE WORK	170

Bibliography 171**Appendices****AI** Crystal screens 186

a/ Hampton 1 + 2

b/ Wizard I +II

c/ The PEG's Suite

d/ Premier screen

AII Protparam 194

a/ FDAS

b/ Sso6206

c/ KDPGal aldolase

AIII Publications 197-206a/ **The fluorinase from *Streptomyces cattleya* is also a chlorinase.**

Deng H, Cobb SL, McEwan AR, McGlinchey RP, Naismith JH, O'Hagan D, Robinson DA, Spencer JB.

Angew Chem Int Ed Engl. 2006 Jan 23;45(5):759-62.

b/ **Overexpression, purification, crystallization and data collection of *Sulfolobus solfataricus* Sso6206, a novel highly conserved protein.**

McEwan AR, Liu H, Oke M, Carter L, Powers H, Dorward M, McMahon SA, White MF, Naismith JH.

Acta Crystallograph Sect F Struct Biol Cryst Commun. 2006 Mar 1;62(Pt 3):228-30.

c/ **Substrate specificity in enzymatic fluorination. The fluorinase from *Streptomyces cattleya* accepts 2'-deoxyadenosine substrates.**

Cobb SL, Deng H, McEwan AR, Naismith JH, O'Hagan D, Robinson DA.

Org Biomol Chem. 2006 Apr 21;4(8):1458-60.

ABSTRACT

Halogenation in nature is becoming better understood and in recent studies (Harper *et al.*, 2003; Murphy, 2003) there appears to be several mechanisms available to nature for halogenation. 5'-Fluorodeoxyadenosine synthase (FDAS) is the only enzyme identified in nature that is capable of formation of a C-F bond. FDAS does this by catalysing the transformation of *S*-adenosyl-L-methionine (SAM) and inorganic fluoride to 5'-fluoro-5'-deoxyadenosine (FDA). It is the essential enzyme in the fluorinase pathway in *Streptomyces cattleya* which results in the production of 4-fluorothreonine and fluoroacetate. This thesis describes a structural investigation into FDAS. Adenosine was identified as a contaminant and removed, permitting the X-ray crystal structures of FDAS in apo-form and in complex with several important substrates and substrate analogues. FDAS was identified to catalyse incorporation of chlorine by a nucleophilic mechanism. FDAS is also capable of utilising 2'-deoxy substrate analogues but not 3'-deoxy ones. The structural data allow both substrate specificity and chlorine incorporation to be rationalised. Thermodynamic studies were carried out on FDAS using isothermal calorimetry and contribute to improving our understanding of the enzyme.

Sso6206 is a small acidic protein with unknown function from *Sulfolobus solfataricus* (Sso). The protein was overexpressed, purified and crystallized in native and selenomethionine derivative forms. Data were measured to 2.4 Å indicating that the asymmetric unit contains between 15 and 30 monomers. This could be a result of a large protomer complex formation, suggesting an unusual function. Attempts to obtain phases using several methods are described.

2-Keto-3-deoxy-6-phosphogalactonate (KDPGal) aldolase was identified in *E. coli* as responsible for the reversible conversion of KDPGal to glyceraldehyde-3-phosphate (G3P) and pyruvate. The protein is homologous to the 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase which performs a similar reaction with KDPG. These protein are important biosynthetic tools. The apo-structure was solved in a previous study, however the reasons for the observed stereo-specificity were not clear. This thesis reports several complex structures and permits a rationalisation of the different stereochemistry between the enzymes.

I, Andrew Robert McEwan, hereby certify that this thesis, which is approximately 33,000 words in length, has been written by me, that this is the record of work carried out by myself and that it has not been submitted in any previous application for a higher degree.

Date: 8-9-06

Signed

Andrew Robert McEwan

I was admitted as a research student in October 2002 and as a candidate for Doctor of Philosophy in October 2003; the higher study for which this is a record was carried out at the University of St. Andrews between 2002 and 2006.

Date: 8-9-06

Signed:

Andrew Robert McEwan

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy at the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 11 SEPT 06

Signed

James H. Naismith

In submitting this thesis to the University of St. Andrews I understand that I am giving the permission for it to be made available for use in accordance with the regulations of the University Library for the time being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date: 8-9-06

Signed:

Andrew Robert McEwan

ACKNOWLEDGEMENTS

My utmost gratitude to Professor James H. Naismith (Jim) for giving me the opportunity to carry out this work and also for his guidance through the research and writing contained within this thesis and finding the money from BBSRC for funding.

I would also like to thank Professor David O'Hagan and Professor Malcolm White, who provided collaborations for the projects and for their time and experience and the colleagues I had the pleasure working with including Dr. Hai Deng, Dr. Huanting Lui, Dr. David Robinson, Dr. Xiaofeng Zhu, Dr. Steve McMahon, Dr. Muse Oke, Dr Srikanathan Velupillai (better known as Kannan) and last but not least Dr. Changjiang Dong (a true master of our art). Special thanks to Dr. Lester Carter for proof reading, Mark Dorward, and Helen Powers for excellent technical support. The rest of Lab3.10 and Sport and also Catherine, Sally and Alex for the Mass Spec service.

I extend my gratitude to my friends and family who helped through the last three years including Ryan McGlinchey (a great housemate for over 2 years), Walter de Laurentis (unquestionably the most inquisitive scientist I have ever met), Gareth Williams (a loyal friend and housemate). Thanks Mum and Dad and my 'wee' Sister (Yvonne) for being there. Finally Lynn Dannfald, you mean the World to me – LFAX.

Thanks all,

Andy.

(To those I forgot let me know and I will make it up with beverages.)

DEDICATION

To my family...

ABBREVIATIONS

2'-d-FDA	5'-fluoro-2',5'-dideoxyadenosine (also 3'd as for 3' deoxy)
4-FT	4-fluorothreonine
5'-CIDA	5'-chloro-5'-deoxyadenosine
5'-d-Ad	5'-deoxyadenosine
5'-FDA	5'-fluoro-5'-deoxyadenosine
Abg	<i>Agrobacterium</i> β -glucosidase
ACP	acyl carrier protein
Ala (A)	alanine
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
ASU	Asymmetric Unit
Aza-AdoMet	5'-deoxy-5'-[N-methyl-N-(γ -amino- γ -carboxypropyl)]-adenosine
<i>Bam</i> HI	<i>Bacillus amyloliquefaciens</i> H restriction endonuclease I (NEB)
Bar	barbamide
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
CCD	Charge Coupled Device
CD	Circular Dichroism (Spectroscopy)
CHLB	<i>Chlorobium phaeobateroides</i>
CHLI	<i>Chlorobium limicola</i>
CMA	cyclopropyl amino acid
CoA	Coenzyme A
CV	Column Volume
Cys (C)	Cysteine
Da	Dalton
DAHP	dihydroxyacetone phosphate
DERA	2-deoxy-D-ribose-5-phosphate aldolase
<i>d</i> H ₂ O	deionised water
DNA	deoxyribonucleic acid
DNase I	deoxyribonuclease I
DNP	dinitrophenyl

<i>DpnI</i>	<i>Diplococcus pneumoniae</i> G41 restriction endonuclease I (NEB)
<i>Dps</i>	<i>Desulfotalea psychrophila</i>
DTT	1,4-dithio-DL-threitol
EBI	European Bioinformatics Institute
ECOLI (<i>E.coli</i>)	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
Enz-SH	Enzyme – thiol group
ESI-MS	Electro Spray Ionisation– Mass Spectroscopy
ESRF	European Synchrotron Radiation Facility
ExPASy	Expert Protein Analysis System
F.C.	Final Concentration
FAc	fluoroacetate
FAD	flavin adenine dinucleotide (oxidized)
FADH / FADH ₂	flavin adenine dinucleotide (reduced forms)
FBP	fructose-1,6-bisphosphate
FMN	flavin mononucleotide
FruA	D-fructose-1,6-diphosphate aldolase
G3P	glyceraldehyde-3-phosphate
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
GRAVY	Grand Average Of Hydrophathy
His (H)	histidine
HPLC	High Performance Liquid Chromatography
I	Intensity (as in $I/\sigma(I)$)
Ile (I)	isoleucine
IP	Image Plate
IPTG	isopropyl- β -D-thiogalactopyranoside
IR	Infra Red (radiation)
ITC	Isothermal Titration Calorimetry
K	Kelvin
KDO8P	3-deoxy-D-manno-2-octulosonate-8-phosphate
KDPG	2-keto-3-deoxy-6-phosphogluconate

KDPGal	2-keto-3-deoxy-6-phosphogalactonate
lac	lactose
LB	Luria Bertani
Leu (L)	Leucine
L-FcuA	L-fuculose-1-phosphate aldolase
L-Met	L-methionine
LN ₂	Liquid Nitrogen
L-RhuA	L-rhamnulose-1-phosphate aldolase
L-SeMet	L-selenomethionine
Lys (K)	lysine
MAD	Multiple Anomalous Dispersion
MALDI-TOF	Matrix Assisted Laser Desorption – Time Of Flight (Mass Spec.)
Man2A	<i>Cellulomonas fimi</i> β -mannosidase
MES	2-morpholinoethanesulfonic acid (buffer)
Met (M)	methionine
MFS	Major Facilitator Superfamily
MIR	Multiple Isomorphous Replacement
MIRAS	(the combination of) MIR with Anomalous Scattering
<i>Mka</i>	<i>Methanopyrus kandleri</i>
<i>Mma</i>	<i>Methanococcus maripaludis</i>
MME	mono-methyl ether
MR	Molecular Replacement
Mr	Relative Molecular Mass (Also MW)
MS	Mass Spectrometry
<i>Mth</i>	<i>Methanothermobacter thermautotrophicum</i>
NAD / NADH	nicotinamide adenine dinucleotide (oxidized / reduced)
<i>NcoI</i>	<i>Nocardia coralline</i> restriction endonuclease I (NEB)
NEB	New England Biolabs [®]
NeuAc	N-acetyl neuraminic acid
NTA	nitrilotriacetic acid
<i>Pae</i>	<i>Pyrobaculum aerophilum</i>
PBS	phosphate buffer saline
PCR	Polymerase Chain Reaction

PCT	Pre-crystallization Test
PDB	Protein Data Bank (database also refers to file type)
PEG	poly ethylene glycol
PEP	phosphoenolpyruvate
PES	polypropylene serum (filter membrane)
PET / CT	Positron Emission Tomography / Computed Tomography
pfam	Protein Families (database)
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
Phe (F)	phenylalanine
pI	Isoelectric Point
PLP	pyridoxal-5'-phosphate
PNP	purine nucleoside phosphate
PNPase	purine nucleoside phosphorylase
PPUT	<i>Pseudomonas putida</i>
Prn	pyrrolnitrin
Pro	Promoter
Pro (P)	proline
PSI	Pounds per square inch (units used on cell disrupter)
<i>Pto</i>	<i>Picrophilus torridus</i>
QM/MM	Quantum Mechanics / Molecular Mechanics
RAMA	Rabbit muscle aldolase
rbs	Ribosome Binding Site
RMSD	Root-Mean-Square Deviation
RNA	ribonucleic acid
RONN	Regional Order Neuronal Network
RSC	Royal Society of Chemistry
SAD	Single Wavelength Anomalous Diffraction
SAM	<i>S</i> -adenosyl-L-methionine
SCAT	<i>Streptomyces cattleya</i>
SDS – PAGE	sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SeMet	selenomethionine
Ser (S)	serine

<i>Sfr</i>	<i>Shewanella frigidimarina</i>
SI	Système International d'unités
SIR	Single Isomorphous Replacement
SRS	Synchrotron Radiation Source, Daresbury Laboratory
Sso (SSO)	<i>Sulfolobus solfataricus</i>
SSPF	Scottish Structural Proteomic Facility
Syr	syringomycin
TagA	D-tagatose 1, 6-diphosphate aldolase
Ter	Terminator
TEV	Tobacco Etch Virus
THMA	<i>Thermotoga maritima</i>
Thr (T)	threonine
TLS	Translation, libration and screw rotation
TOF	Time of Flight
Tris	<i>tris</i> -(hydroxymethyl) aminomethane (buffer)
Trp (W)	tryptophan
Tyr (Y)	tyrosine
UV	Ultra Violet (radiation)
Val (V)	valine
VDW	van der Waals
V _m	volume of the unit cell ($V_m = V/(n \cdot m)$) (where n = the number of ASU's, m = Mr of contents of the ASU)

Units outside the SI: units used with the SI

Physical Quantity	Unit	Symbol for the Unit	Value (SI Unit)	SI Unit
time	minute	min	60	s
time	hour	h	3600	s
time	day	d	86400	s
plane angle	degree	°	($\pi/180$)	rad
volume	litre	L	10^{-3}	m ³
energy	electronvolt	eV	1.60218	$\times 10^{-19}$ J
length	angstrom	Å	10^{-10}	m

LIST OF FIGURES

Chapter 1

- Figure 1.1 Examples of natural organo-halogens (Murphy, 2003)
- Figure 1.2 Structures of several naturally occurring organofluorine molecules
- Scheme 1.1 Stereospecific condensation of fluoroacetyl-CoA with oxaloacetate to form fluorocitrate (Harper et al., 2003) and metabolic fate of fluorocitrate. This is converted to 4-hydroxy-trans-aconitate, a potent inhibitor of aconitase (O'Hagan and Harper, 1999).
- Scheme 1.2 The mechanism of fluorinated fatty acid biosynthesis via fluoroacetyl-CoA in *D. toxicarium* (Harper et al., 2003)
- Figure 1.3 Examples of organofluorine compounds used in industry and medicine
- Figure 1.4 Example of an ^{18}F fluoroacetate used in PET/CT (Matthies et al., 2004). Upper row: CT image slices; middle row: combined ^{18}F -fluoroacetate PET/CT; lower row: ^{18}F -fluoroacetate PET.
- Scheme 1.3 Generation of a fluorinated intermediate, by nucleophilic halogenation, in the E358S mutant of Abg, a β -glucosidase (Zechel et al., 2001)
- Scheme 1.4 Generation of β -D-mannosyl fluoride by the E429A mutant of Man2A, a β -mannosidase
- Scheme 1.5 A suggested schematic representation of the known metabolites and enzymes of the fluorinase pathway
- Scheme 1.6 Omitted
- Scheme 1.7 The predicted mechanism involved in conversion of 5'-FDA to 5-FDRP by the fluorinase related PNP
- Scheme 1.8 Mechanism for the conversion of 5-FDRP to fluoroacetaldehyde by the putative isomerase from *S. cattleya*.
- Scheme 1.9 A mechanistic proposal for the conversion of L-threonine and fluoroacetaldehyde to 4FT and acetaldehyde by the PLP dependent threonine transaldolase (Murphy et al. 2001b)

-
- Figure 1.5 Sequence alignment of FDAS and homologues identified using BLAST search. The trans-domain loop in *S. cattleya* (SCAT.) is highlighted (red). CHLI: *Chlorobium limicola*, THMA: *Thermotoga maritima*, CHLB: *Chlorobium phaeobateroides*, SSO: *Sulfolobus solfataricus*.
- Figure 1.6 Cartoon representations of FDAS monomer (rotated 90°). In the left hand image, the red circle highlights the trans-domain loop (mentioned previously)
- Figure 1.7 Cartoon representation of the FDAS trimers shown in two orientations.
- Figure 1.8 Cartoon representation of the hexamer model, as constructed from crystallographic symmetry using 1rqp (each monomer is coloured separately).
- Figure 1.9 S-adenosyl-L-methionine (SAM) in the active site of FDAS (1rqp).
- Figure 1.10 5'-fluoro-5'-deoxyadenosine (5'-FDA) in the active site of FDAS (1rqp)
- Figure 1.11 Superimposition of substrate and product in the active site.
- Figure 1.12 Representation of 5'-FDA and L-met bound to FDAS. The graphic highlights two hydrogen-bonds to the fluoro-methyl group from Ser-158. There are hydrogen bonding interactions also between Asp-16 and the 2'- and 3'-OH groups of the ribose ring.
- Figure 1.13 The S_N2 catalysed reaction of F⁻ (coordinated by Ser-158) and SAM at the C5' position to produce 5'-FDA and L-methionine via FDAS
- Scheme 1.10 Nucleophilic attack of F⁻ on SAM at the three possible α-carbons adjoined to the sulfonium leaving group.
- Scheme 1.11 Reaction mechanism of haem containing haloperoxidase showing the formation of hypohalous, the porphyrin ring system is omitted.
- Scheme 1.12 Reaction mechanism of vanadium containing haloperoxidase
- Scheme 1.13 Perhydrolase reaction mechanism
- Scheme 1.14 Pyrrolnitrin biosynthetic pathway.
- Scheme 1.15 HOCl generation via by FADH₂ in PrnA
-

- Scheme 1.16 Reaction mechanism of PrnA in 7-chlorotryptophan synthesis.
- Scheme 1.17 Mechanism of chlorination by a non haem iron halogenase
- Scheme 1.18 Summary of CMA biosynthesis, a precursor of coronatine
- Scheme 1.19 Base-catalysed aldol reaction
- Figure 1.14 Examples of aldolase donor groups.
- Figure 1.15 Complementary DHAP dependent aldolase reactions
- Scheme 1.20 NeuAc (lyase) aldolase reaction
- Scheme 1.21 Reaction mechanism for the class II FBP aldolase from *E. coli*.

Chapter 2

- Scheme 2.1 Coupled enzyme assays with 5'-CIDA driving organo-chlorine synthesis from SAM and inorganic chloride
- Figure 2.1 HPLC profiles after running the FDAS reaction (FDAS, SAM and Cl-) with and without the oxidase. (a) The result of FDAS and amino acid oxidase combination, showing 5'-CIDA production (after 12 h, RT). (b) A control without FDAS. (c) As for chromatograph (a) spiked with synthetic 5'-CIDA (Deng et al., 2006).
- Scheme 2.2 Transhalogenation reaction of 5'-CIDA to 5'-FDA using L-SeMet as the co-catalyst
- Figure 2.2 HPLC profile monitoring the FDAS mediated transhalogenation reaction of 5'-CIDA (yellow) to 5'-FDA (green) with time and using L-SeMet as a co-catalyst. L-Se-SAM is observed as a minor product.
- Scheme 2.3 The FDAS can catalyse the conversion of 2'-deoxy-FDA and 2'-deoxy-CIDA substrates with L-SeMet to generate 2'-deoxy-Se-SAM
- Figure 2.3 HPLC profile of the transhalogenation from 2'-deoxy-CIDA to 2'-deoxy-FDA via 2'-deoxy-Se-SAM, catalysed by FDAS.
- Figure 2.4 Primers for *flA* and pEHISTEV plasmid map.

-
- Figure 2.5 Orientation of C222₁ FDAS crystals for optimal data collection. TOP: The unfavourable orientation, with long cell edge (c) perpendicular to rotation axis (Ω). BOTTOM: Favourable orientation where long cell edge is parallel to rotation axis.
- Figure 2.6 Possible C222₁ unit cells in a pseudo hexagonal lattice, Hexagonal axes: Black, Three alternative C-centered orthorhombic lattices (coloured).
- Figure 2.7 PCR site directed mutagenesis (Stratagene, 2004)
- Figure 2.8 Typical crystals obtained by co-crystallization of substrate and FDAS.
- Figure 2.9 Initial density maps of 5'-CIDA modelled into the unbiased Fo-Fc (green mesh and +3.0 σ) electron density map, after molecular replacement, in the active site of FDAS (C = white and yellow; O = red; N = blue; Cl = green).
- Figure 2.10 Density after refinement of 5'-CIDA in the active site of FDAS in initial co-crystallization experiments. Fo-Fc electron density map (green mesh and +3.0 σ , red mesh and -3.0 σ), in the active site of FDAS (C = white and yellow; O = red; N = blue; Cl = green). The 2Fo-Fc (steel blue mesh and 1.0 σ) density map is also shown.
- Figure 2.11 Density observed following refinement of 5'-deoxyadenosine into the active site of FDAS in the initial experiment. Fo-Fc (green mesh and +3.0 σ , red mesh -3.0 σ) electron density map, in the active site of FDAS (C = white and yellow; O = red; N = blue). The 2Fo-Fc (steel blue mesh and 1.0 σ) density map is also shown.
- Figure 2.12 Chromatograms of the HPLC analysis of a) 5'-CIDA co-crystals, b) 5'-CIDA control.
- Figure 2.13 CD spectra and native gel of FDAS; Native - untreated (blue), Native – after treatment with deaminase (green) and N215R (red).
- Figure 2.14 a) Chromatograph showing synthetic adenosine; b) chromatograph of supernatant after recombinant FDAS is denatured; c) chromatograph of supernatant after recombinant FDAS is denatured from a sample treated with deaminase enzyme.
- Figure 2.15 Adenosine refined into the active site of FDAS. Fo-Fc (green mesh and +3.0 σ , red mesh and -3.0 σ) electron density map, in the active site of FDAS (C = white and yellow; O = red; N = blue). The 2Fo-Fc (steel-blue mesh and 1.0 σ) density map is also shown.
-

-
- Figure 2.16 Superimposition of adenosine and 5'-FDA (F= Purple)
- Figure 2.17 Ramachandran plot of the FDAS model (with bound adenosine)
- Figure 2.18 The two conformations of 5'-chloro-5'-deoxyadenosine (5'-CIDA) modelled into the unbiased Fo-Fc (green mesh and +3.0 σ) electron density map, after molecular replacement, in the active site of FDAS (C = white and yellow; O = red; N = blue; Cl = green). A possible chloride ion binding site is also indicated (Green sphere).
- Figure 2.19 Structure of the 5'-CIDA FDAS co-complex with contacts between the halogens and Ser-158 N and O atoms shown. 5'-CIDA (C = white / yellow, N = blue, O = red and Cl = green.) Shown superimposed is 5'-FDA (C = cyan, F = purple, N = blue). The chlorine atom lies further away from the backbone amide of Ser-158; hydrogen bonds are shown here Cl to N: 3.32 Å and Cl to O: 3.58 Å.
- Figure 2.20 Ramachandran plot of the model 2c2w
- Figure 2.21 Aza-AdoMet fully refined into the active site of FDAS, showing the final 2Fo-Fc electron density map (steel blue mesh and 1.0 σ) for the ligand here (C = white and yellow; O = red; N = blue).
- Figure 2.22 Superimposition of the structures of Aza-AdoMet (2c4t) and S-adenosyl-L-methionine (1rqp)
- Figure 2.23 Ramachandran plot of the model 2c4t
- Figure 2.24 5'-deoxyadenosine fully refined into the active site of FDAS, showing the final 2Fo-Fc electron density map (steel blue mesh and 1.0 σ) for the ligand here (C = white and yellow; O = red; N = blue).
- Figure 2.25 Superimposition of 5'-deoxyadenosine (PDB 2cc2) and 5'-fluoro-deoxyadenosine (PDB 1rqr). (C = white, yellow, cyan; O = red; N = blue, F = purple).
- Figure 2.26 Ramachandran plot of the model 2cc2
- Figure 2.27 β -D-eythrofuransyl adenosine fully refined into the active site of FDAS, showing the final 2Fo-Fc electron density map (steel blue mesh and 1.0 σ) for the ligand here (C = yellow and white; O = red; N = blue).
-

-
- Figure 2.28 Superimposition of CC5 (PDB 2cbx) and 5'-FDA (PDB 1rqf) (C = yellow, cyan, white; O = red; N = blue; F = purple).
- Figure 2.29 Ramachandran plot of the model 2cbx
- Figure 2.30 5'-fluoro-2',5'-dideoxyadenosine (2'd-FDA) fully refined into the active site of FDAS (C = yellow and white; O = red; N = blue; F = purple; S = orange).
- Figure 2.31 Superimposition of 5'-fluoro-2',5'-dideoxy adenosine (5F1, 2c5b), (C = white and yellow; O = red; N = blue; F = purple) and of 5'-fluoro-5'-deoxy adenosine (5FD, 1rqf) (C = cyan). Interactions of the ligand to the protein are indicated; 03' to Ser-158 (O) = 3.2 Å; 03' to Asp-16 (O1) = 2.5 Å; 03' to Asp-16 (O2) = 3.2 Å; 03' to Tyr-77 (N) = 3.3 Å
- Figure 2.32 Ramachandran plot of the model 2c5b
- Figure 2.33 2'-deoxyadenosine (2DA) and methionine modeled into the unbiased Fo-Fc (green mesh and +3.0 σ) electron density map, after molecular replacement with 1rqf (without ligand), in the active site of FDAS (C = white and yellow; O = red; N = blue).
- Figure 2.34 Superimposition of 2'-dFDA (C = cyan; O = red; N = blue and F = purple) with 2'-deoxyadenosine in the active site of FDAS. (C = white and yellow; O = red; N = blue)
- Figure 2.35 Ramachandran plot of the model 2c5h
- Figure 2.36 Crystals obtained of the apo-FDAS.
- Figure 2.37 Diagram of the unbiased Fo-Fc density (green mesh and +3.0 σ ; red mesh and -3.0 σ) in the active site of the apo crystals. The initial 2Fo-Fc density obtained is shown here also (steel blue mesh and 1.0 σ). (C = white, O = red, N = blue).
- Figure 2.38 Ramachandran plot of the model 2c4u
- Figure 2.39 Superimpositions of the carbon α trace of 1rqf (SAM co-complex: blue), 1rqf (FDA co-complex: green) and 2c4u (apo-model: red), as the hexamer.
- Figure 2.40 Dissection of the surface potential reveals the buried active site of FDAS, SAM has been modelled into each. a) Relates to the co-crystal structure of substrate (PDB: 1rqf); b) Relates to the apo-structure (PDB: 2c4u) SAM was modelled into the binding site.
-

Figure 2.41	ITC data of FDAS against Adenosine in Tris buffer
Figure 2.42	ITC data of FDAS against FDA in Tris buffer
Figure 2.43	ITC data of FDAS against SAH in Tris buffer
Figure 2.44	ITC data of FDAS against SAM in Tris buffer
Figure 2.45	ITC data of FDAS against 5'-CIDA in Tris buffer
Figure 2.46	ITC data of FDAS against SAH (Phosphate buffer)
Figure 2.47	ITC data of FDAS against SAM (Phosphate buffer)
Figure 2.48	ITC data of FDAS against 5'-CIDA (Phosphate buffer)
Figure 2.49	ITC data of FDAS against Adenosine in Tris buffer
Figure 2.50	ITC data of FDAS against FDA in Tris buffer
Figure 2.51	ITC data of FDAS against SAH (Phosphate buffer)
Figure 2.52	ITC data of FDAS against SAM (Phosphate buffer)
Figure 2.53	ITC data of FDAS against 5'-CIDA (Phosphate buffer)

Chapter 3

Figure 3.1	Sequence alignment of Sso6206 and archaeal and bacterial homologues. The four conserved acidic residues are highlighted in red, and the conserved histidine in blue. Sso6206 from <i>Sulfolobus solfataricus</i> (Sso); Pto, <i>Picrophilus torridus</i> ; Mka, <i>Methanopyrus kandleri</i> ; Pae, <i>Pyrobaculum aerophilum</i> ; Pfu, <i>Pyrococcus furiosus</i> ; Mth, <i>Methanothermobacter thermautotrophicum</i> ; Mma, <i>Methanococcus maripaludis</i> ; Dps, <i>Desulfotalea psychrophila</i> ; Sfr, <i>Shewanella frigidimarina</i> .
Figure 3.2	RONN plot of probability of disorder against residue number. The threshold value (0.5) for disorder is indicated. Residues with a probability above 0.5 are predicted to be disordered, while those below are predicted to be ordered.
Figure 3.3	Cluster analysis of Sso6206 indicating that the protein lies in cluster A since predicted pI = 4.632 and GRAVY = 0.194.

- Figure 3.4 Vector map of pEHISTEV and Sso6206 primers
- Figure 3.5 Soft ionization mass spectroscopy (Sso6206) indicating presence of monomer, dimer and hexameric peaks.
- Figure 3.6 A) Initial native crystals obtained in Wizard™ I # 36, and B) the same crystals dyed with Izit™ solution (Hampton research).
- Figure 3.7 Picture of a typical Sso6206 crystal grown in optimized conditions. The crystals appear coloured under a polarizing microscope filter (used for measuring birefringence).
- Figure 3.8 Diffraction pattern of Sso6206 collected at the ESRF (Grenoble). The resolution at the edge of the detector is 2.3 Å.
- Figure 3.9 Mass spectra of the SeMet derivative of Sso6206
- Figure 3.10 SeMet crystals grown in 0.2 M sodium citrate and 0.1 M imidazole pH 6.0
- Figure 3.11 Circular dichroism and SDS gel of the SeMet derivative of Sso6206 (red) compared with native Sso6206 (blue).
- Figure 3.12 Diffraction pattern of SeMet Sso6206 collected in-house
- Figure 3.13 Diffraction pattern of SeMet Sso6206 collected at the ESRF Grenoble
- Figure 3.14 Fluorescence scan (Se) performed on SeMet Sso6206 crystals

Chapter 4

- Scheme 4.1 – KDPG aldolase reaction versus the KDPGal aldolase. The enzymes are responsible for the breakdown of the sugars in the cell.
- Figure 4.1 Sequence alignment (ClustalW) KDPGal aldolase from ECOLI; *E. coli*, KDPG aldolase from *E. coli*, PPUT; *Pseudomonas putida*, THMA; *Thermotoga maritima*. A conserved active site lysine (red arrow) and glutamic acid (blue arrow) are indicated. The asterisk indicate the position of the active site arginines.
- Figure 4.2 KDPGal aldolase trimer.

-
- Scheme 4.2 The mechanism of KDPG Aldolase (Fullerton et al., 2006).
- Figure 4.3 A representation of the active site of the KDPGal aldolase, a slice of observed density are shown (blue mesh = 1.0 σ). Residues located at the active site have been labelled, several waters were located (red spheres).
- Figure 4.4 Cartoon representation of the KDPGal aldolase monomer at different orientations.
- Figure 4.5 Superimposition of KDPGal aldolase and superimposition of the *E. coli* KDPG (Yellow, PDB: 2c0a), *Thermotoga maritima* (Cyan, PDB: 1wa3) and *E. coli* KDPGal aldolase (Grey) monomer models.
- Figure 4.6 Superimposition of the active site of KDPG aldolase (PDB 1EUN – yellow cartoon, red labels) and KDPGal aldolase (grey cartoon, blue labels).
- Figure 4.7 KDPGal aldolase crystals grown from precipitant in conditions containing 30 % w/v PEG 200, 0.1 M MES ph 6.2 and 5 % PEG 3000.
- Figure 4.8 A diffraction image of the KDPGal aldolase crystals after soaking.
- Figure 4.9a Initial density in the active site of crystals soaked with pyruvate solution.
- Figure 4.9b Fully refined active site showing the ligands as refined into the 2Fo-Fc maps (blue mesh and 1.0 σ).
- Figure 4.10 Key interactions (black dashes) in the active site between pyruvate (covalently bound to Lys-126) and the protein. The waters (labelled W1 and W2) are those expected to be involved in proton shuttling (Fullerton et al., 2006).
- Figure 4.11 LIGPLOT representation of interactions between the pyruvate modified lysine (Lyp-126) and KDPGal aldolase.
- Figure 4.12 Active site of KDPGal aldolase (C = white; N = blue; O = red; HOH = red spheres) superimposed with the active site of KDPG aldolase of *Thermotoga maritima* (C = cyan; HOH = orange spheres).
- Figure 4.13 Ramachandran plot of the KDPGal aldolase with pyruvate model.
-

-
- Figure 4.14 Fully refined active site of KDPGal aldolase crystals soaked with G3P. The 2Fo-Fc density for obtained (steel blue mesh and 1.0 σ) for the ligands is shown. There was no extra density obtained in the final Fo-Fc maps (3.0 or -3.0 σ) which would suggest problems with the ligands. (C = white & yellow, N = blue, O = red, P = orange).
- Figure 4.15 Interactions between the G3P and phosphate to key residues in the active site of KDPGal aldolase (black dashes).
- Figure 4.16 Ramachandran plot of the G3P soaked KDPGal aldolase model
- Figure 4.17 KDPGal modelled into the unbiased Fo-Fc (green mesh and +3.0 σ) electron density map after molecular replacement with the existing protein model in the active site of KDPGal aldolase soaked with KDPGal. The 2Fo-Fc is shown also (steel blue mesh and 1.0 σ).
- Figure 4.18 Pyruvate and G3P fully refined into the active site of KDPGal aldolase, showing the final 2Fo-Fc electron density map (steel blue mesh and 1.0 σ).
- Figure 4.19 Interactions between the G3P and phosphate to key residues in the active site of KDPGal aldolase (black dashes) in the model from crystals soaked with KDPGal. (Carbons are numbered according to their final numbering scheme in KDPGal).
- Figure 4.20 LIGPLOT diagram of the interactions with respect to G3P and the pyruvate modified lysine (LYP).
- Figure 4.21 Model superimposition of G3P soak (yellow), pyruvate (blue) and KDPGal soak (white). The possible attack trajectories of the pyruvate at C4 are shown (black dash).
- Figure 4.22 The electron surface potential of A) KDPGal aldolase and B) KDPG aldolase showing G3P soak model (yellow), KDPGal soak model (white) and *Thermotoga maritima* KDPG aldolase (green) with the covalent product precursor modelled into the active site.
- Figure 4.23 Ramachandran plot of the KDPGal aldolase soaked with KDPGal model
-

LIST OF TABLES

Chapter 1

- Table 1.1 Properties of H, F and Cl as extracted from “Fluorine in organic chemistry” by R. D. Chambers. pp. 12-13 (a $X^+ + e^- \rightarrow X$, b $X + e^- \rightarrow X^-$, c Covalent radii in CX_4)
- Table 1.2 FDAS (*flA*) gene cluster in *S. cattleya* and gene locus map *functions are predicted by sequence, (Huang et al., 2006).

Chapter 2

- Table 2.1 Substrate specificity of ‘reverse’ reaction by FDAS.
- Table 2.2 Substrates, inhibitors and analogues.
- Table 2.3 FDAS grid screen for crystal growth, where optimization of % w/v PEG 1000 is varied against pH of 0.1 M phosphate - citrate buffer.
- Table 2.4 Output from pointless: scoring the possible indexing schemes, showing that for this case the data is indeed C222 and not hexagonal.
- Table 2.5 Retention times of selected compounds.
- Table 2.6 Summary of data collection and refinement statistics
- Table 2.7 Summary of data collection and refinement statistics
- Table 2.8 Summary of data collection and refinement statistics
- Table 2.9 Summary of data collection and refinement statistics
- Table 2.10 Summary of data collection and refinement statistics
- Table 2.11 Summary of data collection and refinement statistics
- Table 2.12 Summary of data collection and refinement statistics
- Table 2.13 Summary of data collection and refinement statistics
- Table 2.14 Summary of ITC data results (according to one site model), with unrestrained stoichiometry and experimental concentrations.

Table 2.15	Summary of ITC data results (according to one site model) recalculated so that $n \approx 1.0$
------------	--

Chapter 3

Table 3.1	Sso6206 crystallization grid screen for crystal growth, where optimization of sodium citrate concentration is varied against pH of 0.1 M imidazole
Table 3.2	SeMet Sso6206 crystallization grid screen for crystal growth, where optimization of sodium citrate concentration is varied against pH of 0.1 M imidazole
Table 3.3	Sso6206 crystal data and data collection statistics. Values in parentheses refer to the highest resolution shell. Solvent calculated for 22 monomers per ASU.
Table 3.4	Summary of heavy atom soaking experiments
Table 3.5	Sso6206 crystal data and data collection statistics for heavy atom screening. Values in parentheses refer to the highest resolution shell. Solvent calculated for 22 monomers per ASU.
Table 3.6	Table of anomalous data as output from SCALA. The CC_{anom} is the anomalous correlation coefficient which indicates a weak anomalous signal.
Table 3.7	Summary of data statistics for the SeMet data collection.
Table 3.8	Table of anomalous data as output from SCALA. The CC_{anom} is the anomalous correlation coefficient (red box) indicates an anomalous signal to $\sim 5.5 \text{ \AA}$.
Table 3.9	Summary of data statistics for the sulfur SAD experiment.
Table 3.10	Table of anomalous data (Sulfur SAD), the column containing the correlation coefficient for anomalous data is highlighted.

Chapter 4

Table 4.1	Grid screen for E. coli KDPGal aldolase crystal growth, where optimization of % w/v PEG 200 is varied against pH of 0.1 M MES and 5 % PEG 3000 is used throughout.
Table 4.2	Compounds used for co-crystallization and soaking experiments.
Table 4.3	Summary of data collection and refinement statistics
Table 4.4	Summary of data collection and refinement statistics
Table 4.5	Summary of data collection and refinement statistics

1

Introduction

1 INTRODUCTION

1.1 Fluorine in nature

Naturally occurring organo-halogens are diverse (as shown in Figure 1.1). Since the discovery of diiodotyrosine in 1886 (Dreschel, 1886) there have been about 4000 such molecules reported, from a diverse variety of organisms. The overwhelming majority contain either Cl or Br. The mechanisms developed by organisms to produce such compounds are of considerable interest because of their importance in medicinal research. The mechanisms can also serve as a template for novel methods of organic synthesis. X-ray crystallography has elucidated mechanisms of biological halogenation.

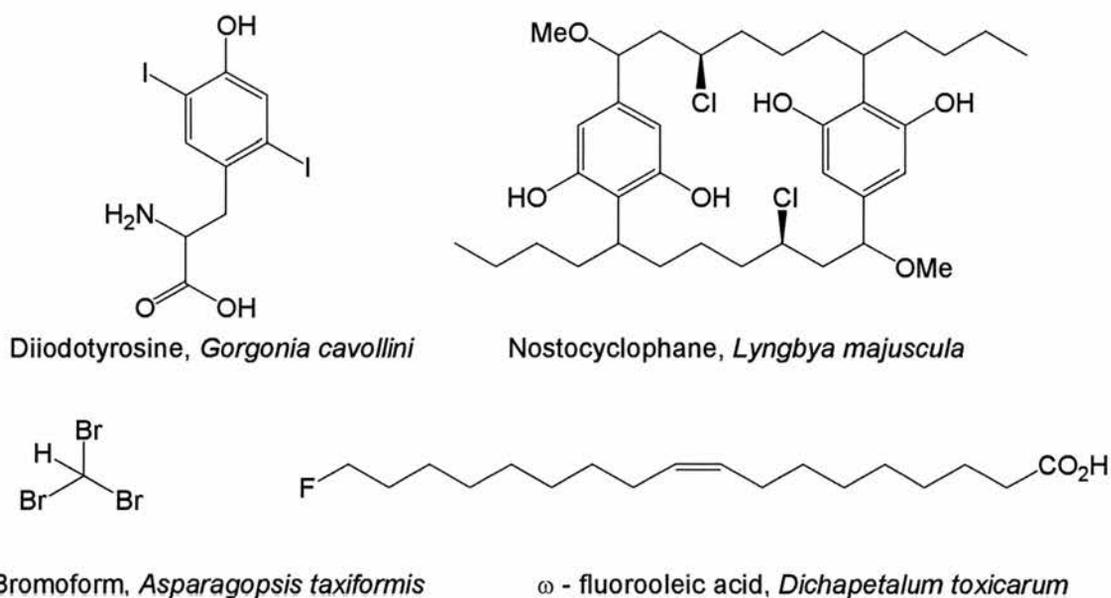


Figure 1.1 – Examples of natural organo-halogens (Murphy, 2003)

Fluorine is the 13th most abundant element in nature (Paul and Huang, 1980) however naturally occurring organofluorine products are rare (O'Hagan *et al.*, 2002; Xu

et al., 2003). To date only around fourteen naturally occurring fluorinated molecules have been reported, some of which are detailed in Figure 1.2 (Deng *et al.*, 2004).

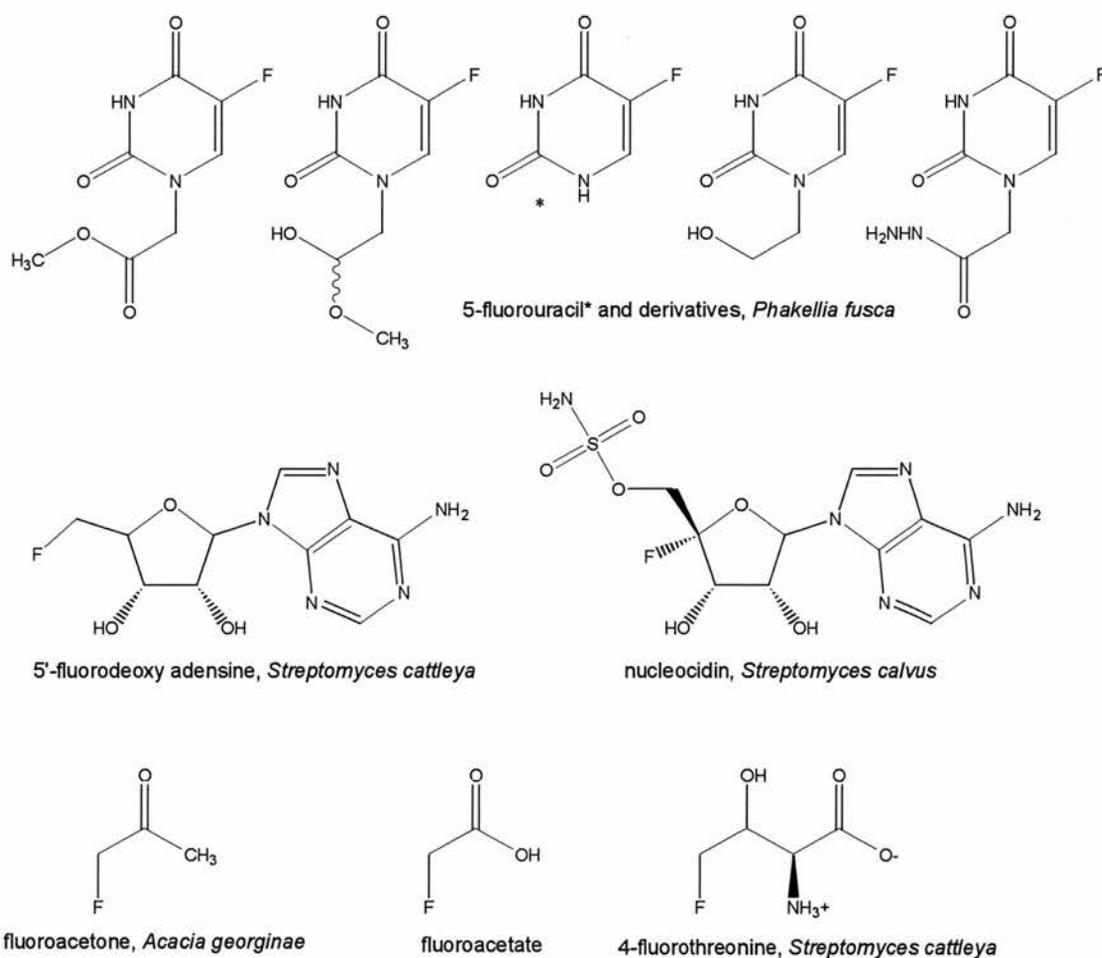
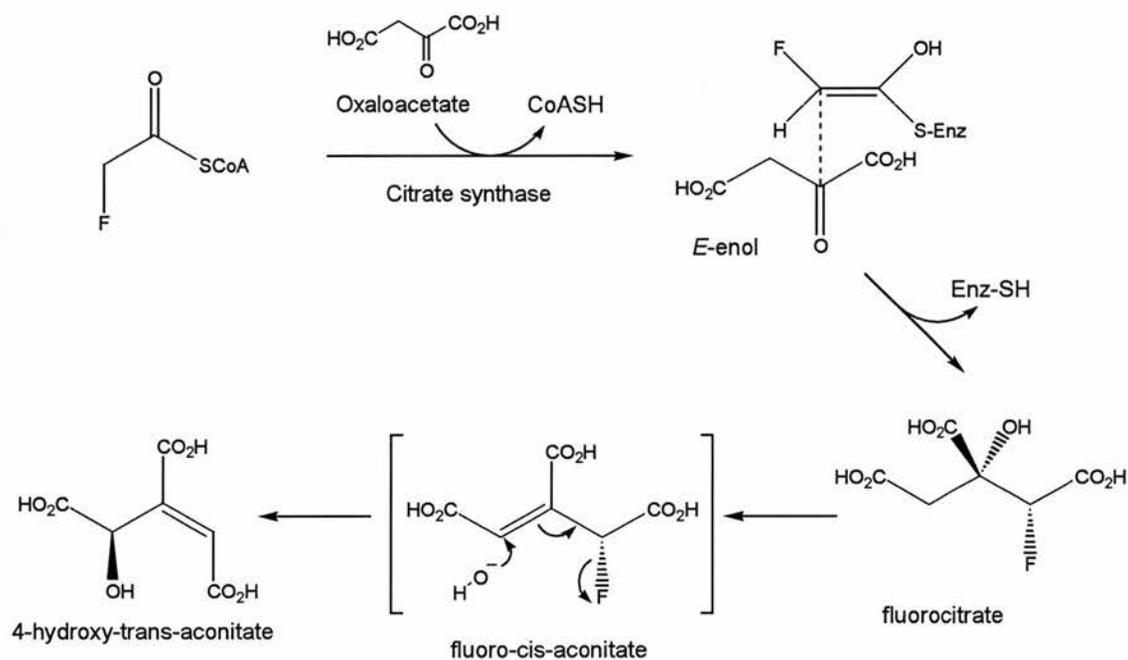


Figure 1.2 – Structures of several naturally occurring organofluorine molecules

Fluoroacetate (FAc), the first organofluorine compound reported in the literature (Marais, 1944), was identified as part of a defence mechanism in an African plant *Dichapeta cymosum* and has been identified in several other plants since (Harper *et al.*, 2003). This compound is toxic to mammals, being incorporated into the citric acid cycle to produce (2R,3R)-2-fluorocitrate (Peters *et al.*, 1959) which in turn is metabolised to 4-

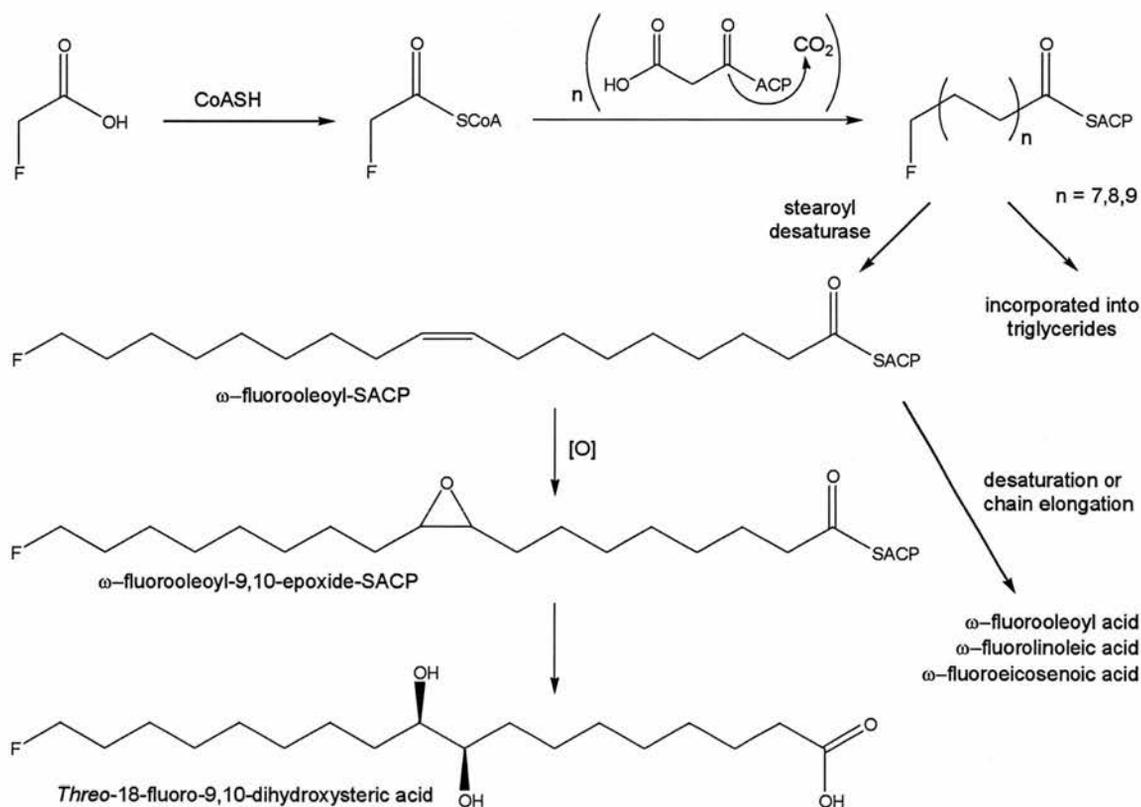
hydroxy-trans-aconitate. The product is a potent inhibitor of aconitase and therefore blocks the cycle (Harper and O'Hagan, 1994).



Scheme 1.1 – Stereospecific condensation of fluoroacetyl-CoA with oxaloacetate to form fluorocitrate (Harper *et al.*, 2003) and metabolic fate of fluorocitrate. This is converted to 4-hydroxy-trans-aconitate, a potent inhibitor of aconitase (O'Hagan and Harper, 1999).

FAC has been identified as a natural product from the bacterium *S. cattleya* (Sanada *et al.*, 1986). Another product of the fluorinase pathway in *S. cattleya* is the novel amino acid 4-fluorothreonine (4-FT) (Hamilton *et al.*, 1997; Marais, 1944). This compound has been shown to exhibit a broad range of antimicrobial activity (Harper and O'Hagan, 1994).

Fluorinated fatty acids such as; ω -fluorooleic acid, ω -fluoropalmitic acid, ω -fluoropalmitoleic acid, ω -fluorostearic acid, ω -fluorolinoleic acid and ω -fluoroicosenoic acid, are produced in *Dichapeta toxicarum* (Christie *et al.*, 1998; Hamilton and Harper, 1997; Peters and Shorthouse, 1967; Peters *et al.*, 1959). It is thought that these compounds are derived from fluoroacetone via fluoroacetyl-CoA since the fluorine is located at the (ω) terminal carbon of the lipid (Hamilton and Harper, 1997). This is shown in Scheme 1.2.



Scheme 1.2 – The mechanism of fluorinated fatty acid biosynthesis via fluoroacetyl-CoA in *D. toxicarium* (Harper *et al.*, 2003)

Nucleocidin (shown in Figure 1.2), an antibiotic, was identified as a product of the actinomycete, *S. calvus* and was shown to contain fluorine (Thomas *et al.*, 1956). Attempts to repeat the bacterial production of nucleocidin have failed (Deng *et al.*, 2004).

In 2002 (Xu *et al.*, 2003) it was reported that several derivatives of 5-fluorouracil were isolated from the marine sponge, *Phakallia fusca* (Figure 1.2). There is some speculation concerning the basis of fluorine incorporation in these compounds (some of which are recognised as potent anti tumour drugs). It has been suggested that they result from industrial contamination rather than fluorination of a precursor by the sponge (Deng *et al.*, 2004).

1.2 Importance of fluorinated compounds

Fluorine is the most electronegative element with a very high oxidation potential ($V = -3.06$ v) and forms the strongest single bond to carbon (~ 490 kJ mol⁻¹). Fluorine when bound to a carbon has three pairs of electrons in its outer electron shell which do not form covalent bonds to other atoms. The van der Waals (VDW) radius of fluorine (in a C-F system) is similar to a hydrogen atom (in a C-H system). Fluorinated derivatives of existing compounds can be prepared often resulting in changes in chemical, physical and pharmacological properties (Bohm *et al.*, 2004). Typically fluorinated compounds are more chemically stable compounds and a rational approach in drug discovery is to produce fluorinated derivatives of lead compounds, to alter their properties. A good example of this is shown in Prozac[®] where the CF₃ group enhances transferability across the blood brain barrier, where it elicits its effect.

	H	F	Cl
Electronic configuration	$1s^1$	$\dots 2s^2 2p^5$	$\dots 3s^2 3p^5 3d^0$
Electronegativity (Pauling)	2.20	3.98	3.16
Ionisation energy (kJ mol^{-1}) ^a	1312	1681	1251
Electron affinity (kJ mol^{-1}) ^b	74.0	332.6	348.5
Bond energies C-X in C-X_4 (kJ mol^{-1})	446.4	546.0	305.0
Bond lengths of C-X ^c (Å)	1.091	1.319	1.767
van der Waals radius (Å)	1.20	1.47	1.75
Preference as leaving group	H^+	F^-	Cl^-

Table 1.1 – Properties of H, F and Cl as extracted from “Fluorine in organic chemistry” by R. D. Chambers. pp. 12-13 (^a $\text{X}^+ + e^- \rightarrow \text{X}$, ^b $\text{X} + e^- \rightarrow \text{X}^-$, ^c Covalent radii in CX_4)

Commercially important organofluorine molecules include; the antidepressant Prozac® (Eli Lilly), the antibiotic Cipro® (Bayer), and the herbicides Treflan® (Eli Lilly) and Fusilade® (ICI), these are shown in Figure 1.3.

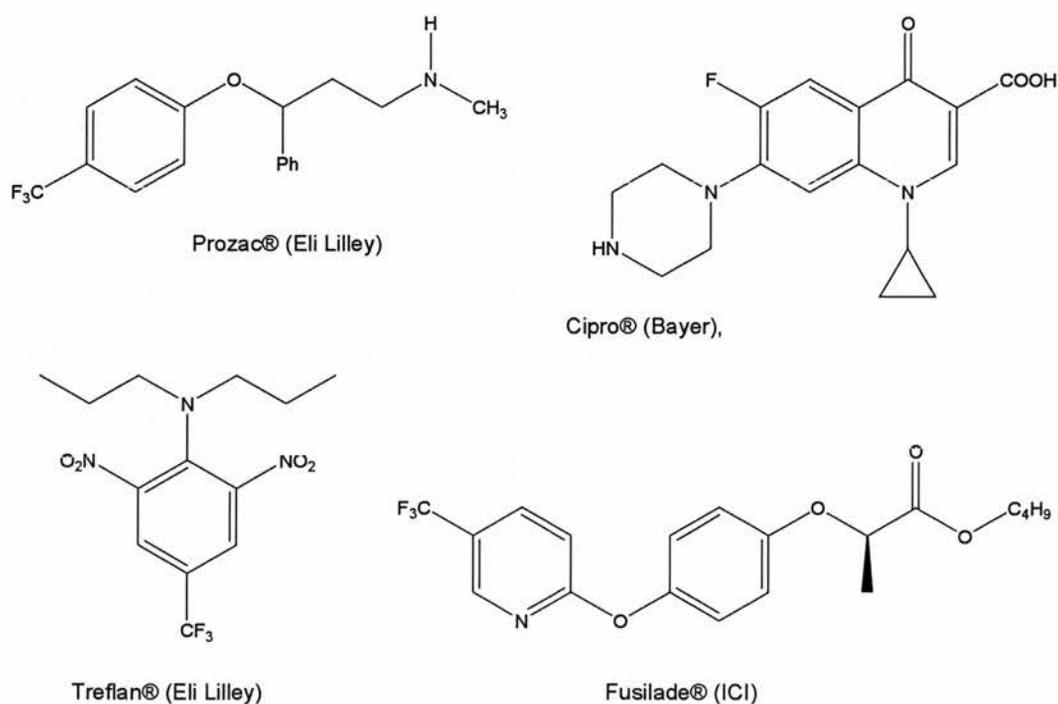


Figure 1.3 – Examples of organofluorine compounds used in industry and medicine

In addition to pharmaceuticals, a further use of organofluorine compounds is in positron emission tomography (PET), an example of which is shown in Figure 3.4. In PET the ^{18}F isotope is used with its half life of 109 min. (Volkow *et al.*, 1988), which is more useful than ^{11}C (20 min.), ^{15}O (2 min.) and ^{13}N (10 min.). In oncology, ^{18}F PET is of interest as it can be used to locate tumours in tissues and organs inaccessible by other techniques. For example ^{18}F -fluoroacetate is used for detection of the recurrence and metastases from prostate cancer (Matthies *et al.*, 2004).

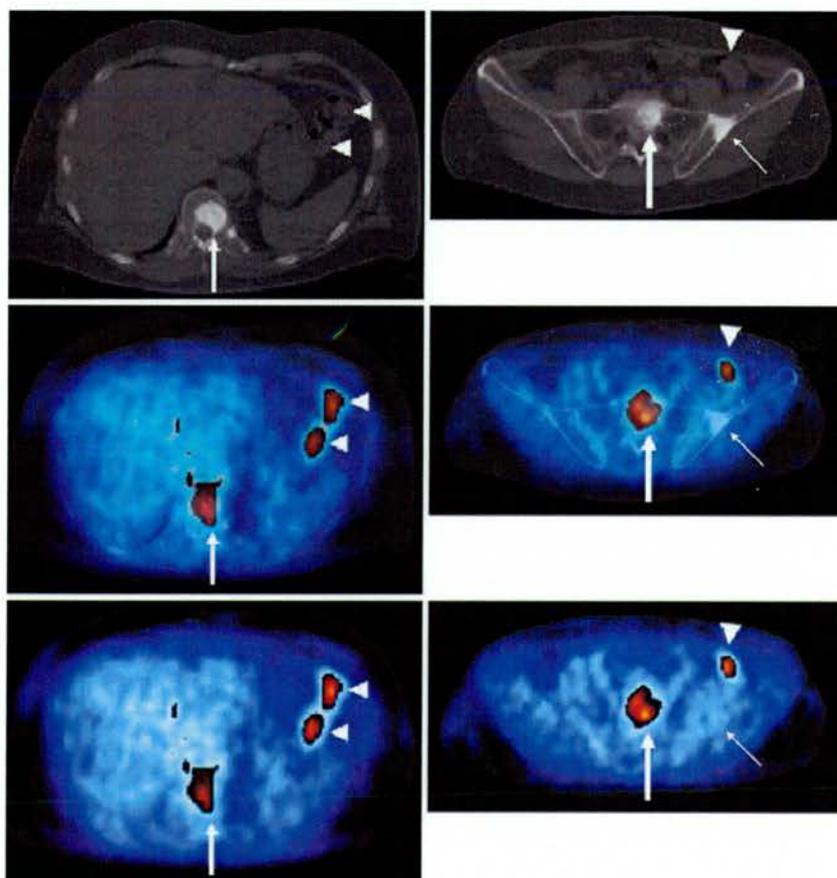


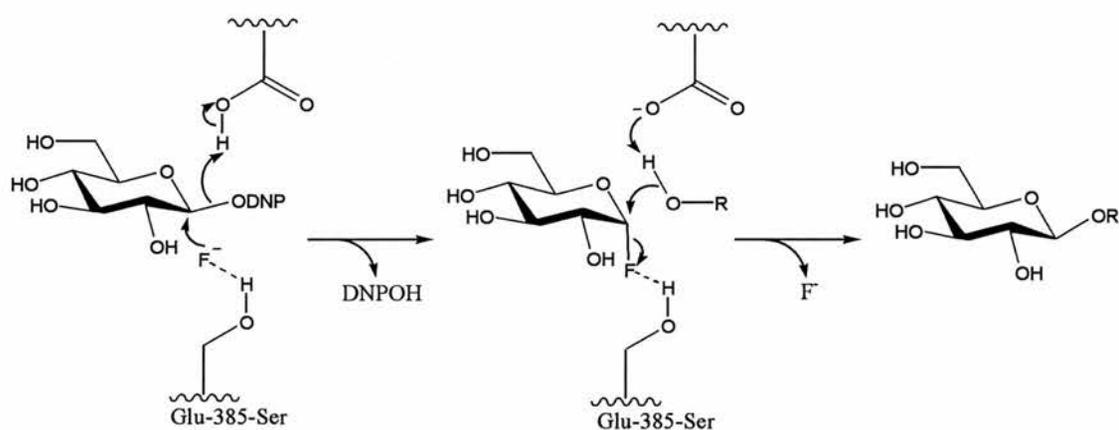
Figure 1.4 – Example of ^{18}F fluoroacetate used in PET/CT (Matthies *et al.*, 2004). Upper row: CT image slices; middle row: combined ^{18}F -fluoroacetate PET/CT; lower row: ^{18}F -fluoroacetate PET.

Although organofluorine chemistry is a vibrant field of research, chemical synthesis is time consuming and difficult. Understanding the biosynthesis of natural organofluorine compounds may yield better strategies and approaches for fluorine chemistry. There are relatively few organofluorine molecules. This may reflect the lower bioavailability of the aqueous ion. Most of the world's fluorine is found as fluorospar. In addition to low availability there are also chemical reasons which present barriers to its incorporation. The high oxidation potential of fluorine means that many of the enzymatic processes available for non-fluorine halogenation are implausible; for example peroxide oxidation of the anion in a non-fluorinating halogenase can not be translated to fluorine. In the gas phase the fluoride anion can be a good nucleophile (Glukhovtsev *et al.*, 1996), however the energy required to desolvate $F^-_{(aq)}$ ion ($\Delta G^\circ > 439 \text{ kJ mol}^{-1}$) presents a formidable obstacle to nucleophilic chemistry (Vincent and Hillier, 2005).

The scarcity of organo-fluorine compounds makes identification of fluorinating enzymes a difficult task. Plant sources, where most organo-fluorine compounds have been discovered, are arguably one of the most demanding models to investigate due to the lack of plant material, variability in expression and lack of genetic data. The observation of fluorinated metabolites in bacteria has presented a more tractable target system.

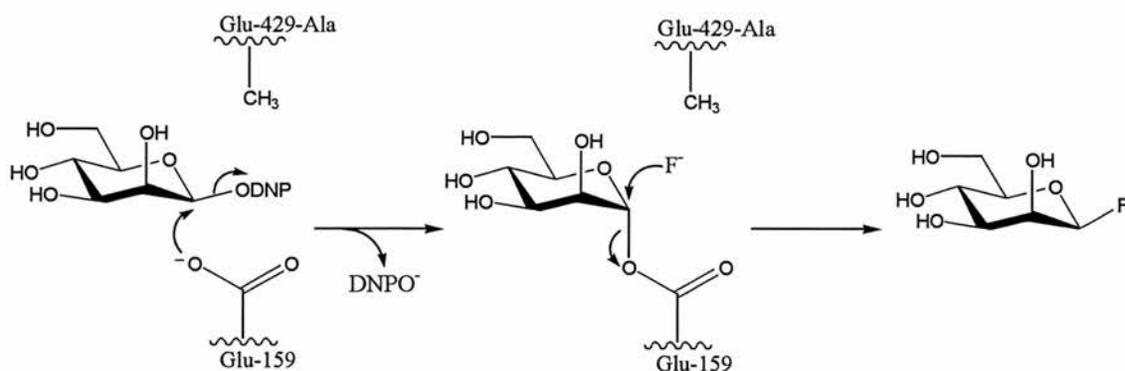
1.3 Engineered fluorinating enzymes

Two mutant glycosidase enzymes have been shown to catalyse carbon fluorine bond formation (Zechel *et al.*, 2001). The first mutant was formed by mutation of the catalytic glutamate, of *Agrobacterium* sp. β -glucosidase (Abg), to an alanine, serine or glycine. The mutants show no glycosidase activity under normal conditions; however this can be restored by addition of fluoride. This is thought to be due to fluoride acting as the catalytic nucleophile producing an α -glucosyl fluoride which in turn can be converted to the expected product (as described in Scheme 1.3).



Scheme 1.3 – Generation of a fluorinated intermediate, by nucleophilic halogenation, in the Glu-358-Ser mutant of Abg, a β -glucosidase (Zechel *et al.*, 2001)

The second example was shown whereupon an Glu-429-Ser acid-base mutant of *Cellulomonas fimi* β -mannosidase (Man2A) turns over 2,5-dinitrophenyl β -mannoside in the presence of fluoride to produce β -D-mannosyl fluoride (Zechel *et al.*, 2001). The predicted reaction mechanism is shown in Scheme 1.4.



Scheme 1.4 – Generation of β -D-mannosyl fluoride by the Glu-429-Ala mutant of *Man2A*, a β -mannosidase

The reaction mechanism proposed involves desolvation of the $F^-_{(aq)}$ ion. It is also noted that chloride and bromide will also act as nucleophiles, where the order of reactivity is $F^- > Cl^- > Br^-$ (Zechel and Withers, 2001). These observations are consistent with what is expected in the gas phase but not what is expected in an aqueous environment suggesting that desolvation is likely to occur in the active site of the enzymes. Desolvation is suggested to occur by formation of protein to fluorine hydrogen bonds, which reduces the energetic penalty for desolvation. The authors predicted that this occurs in nature.

1.4 Fluorinated products in bacteria

1.4.1 Actinomycetes

Actinomycetes are organisms belonging to the Actinomycetales (major subdivision of Prokaryotes). Actinomycetales are on occasion referred to as ‘higher bacteria’. Actinomycetes are gram-positive and tend to grow in soil as filamentous mycelial colonies. Streptomyceteae belong to the Actinomycetes and are atypically pathogenic but are responsible for production of over 50 % of all known natural antibiotics. Two strains (*Streptomyces calvus* and *Streptomyces cattleya*) have been shown to produce fluorinated compounds which show antibiotic effects.

1.4.2 *Streptomyces cattleya*

Fluorinated products, fluoroacetate (FAc) and 4-fluorothreonine (4-FT), were identified in the bacteria *S. cattleya* (Sanada *et al.*, 1986). The production of fluorinated compounds was observed during cultures of the bacterium in the presence of inorganic fluoride. The conditions reported in the literature occurred during optimization of the growth media for production of the antibiotic thienamycin. The reasons for production of these fluorinated metabolites are speculative. It has been suggested they are produced for defence as antibiotics, or for incorporation into protein (Harper and O'Hagan, 1994).

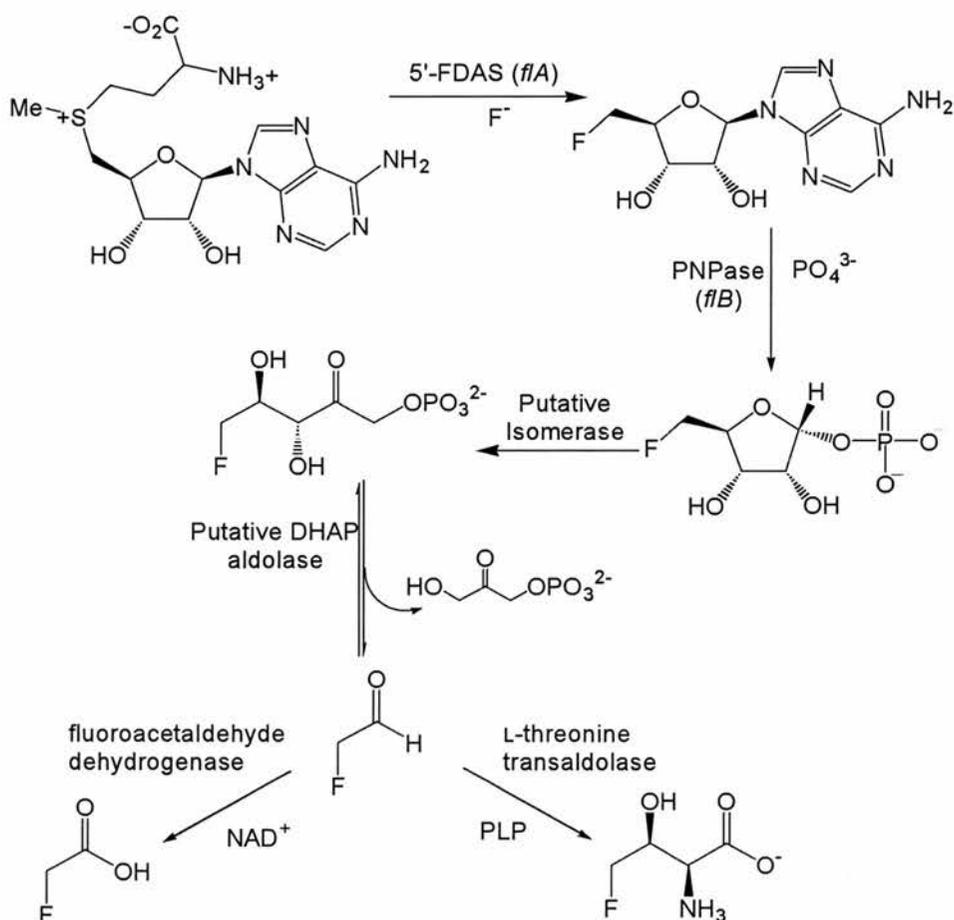
1.4.3 An overview of the fluorinase pathway

1.4.3.1 *The fluorinase enzyme and pathway*

Investigations into the production of FAc and 4-FT in *S. cattleya* led to the assumption of a ‘fluorinase pathway’ (Hamilton *et al.*, 1997). Investigations into the

pathway identified an enzyme capable of catalysing the formation of a carbon fluorine bond (O'Hagan *et al.*, 2002; Schaffrath *et al.*, 2003).

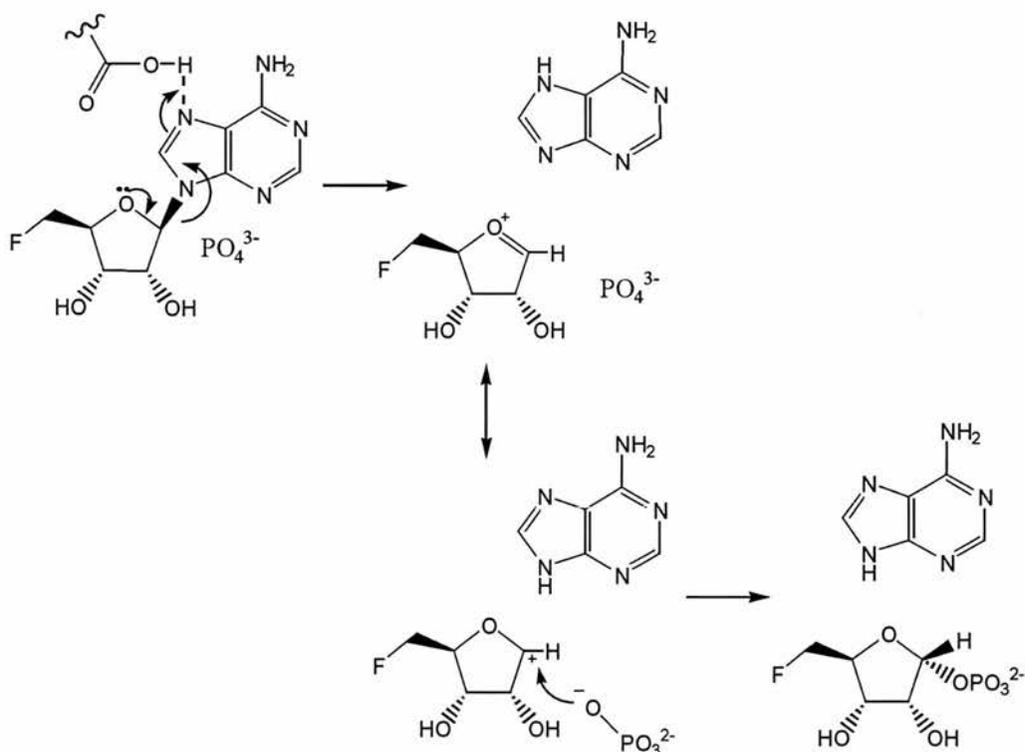
The enzyme was dubbed, 'fluorinase' but its proper name is 5'-fluorodeoxyadenosine synthase (FDAS) (EC 2.5.1.63) and is the first enzyme in the 'fluorinase pathway'. The 'fluorinase pathway' is a part of secondary metabolism active only under certain conditions. The remainder of the pathway from 5'-FDA to FAc and 4FT has been under intense investigation (Cobb *et al.*, 2004; Murphy *et al.*, 2001a; Murphy *et al.*, 2001b; Murphy *et al.*, 2003). The current state of research is illustrated in Scheme 1.5.



Scheme 1.5 – A suggested schematic representation of the known metabolites and enzymes of the fluorinase pathway

1.43.2 The PNPase enzyme

The purine nucleoside phosphorylase (PNPase) reaction which is thought to decompose 5'-FDA has been well studied (Taylor Ringia and Schramm, 2005). The mechanism for the 5'-FDA decomposition is likely to be identical to that already known for PNPase's. The gene predicted to be responsible for the PNPase enzyme (*flB*) in *S. cattleya* has been identified (Deng *et al.*, 2006). This has been successfully cloned into expression vectors and is currently being investigated for structural studies. Attempts to overexpress this enzyme have been limited by its low solubility.



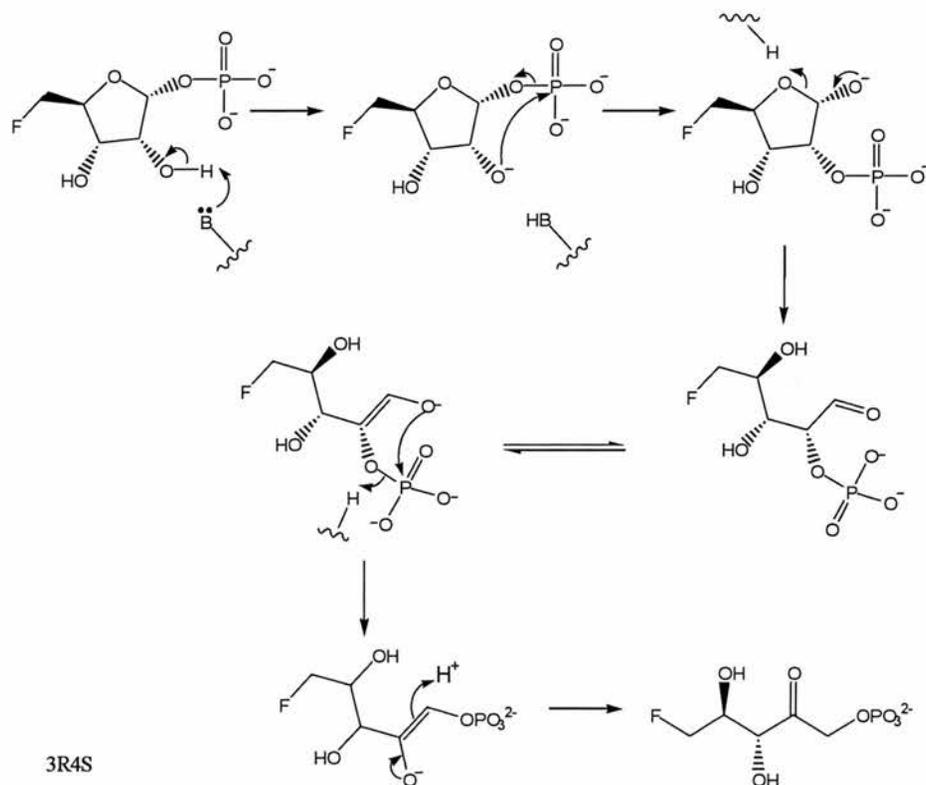
Scheme 1.7 – The predicted mechanism involved in conversion of 5'-FDA to 5-FDRP by the fluorinase related PNP

It was shown that this enzyme has a preference for 5'-FDA over adenosine and it is hoped that subsequent structural analysis may identify the reasons for this

(McGlinchey, 2006). The protein *f1B* has 53% sequence identity to a 5'-methyl-thiol-adenosine phosphorylase (*Geobacter metallireducens*), an enzyme involved in the methionine salvage pathway.

1.4.3.3 The isomerase enzyme

The transformation of 5-FDRP to 5-fluororibulose phosphate requires an isomerase. This enzyme has been partially purified and has been shown to require the phosphate group of 5-FDRP for activity and may be related to a 5'-methyl-thiol-isomerase (McGlinchey, 2006). A mechanism has been proposed which involves attack of a basic residue (—B:) on the 2-hydroxyl position of 5-FDRP before rearrangement to the product (Scheme 1.8).



Scheme 1.8 – Mechanism for the conversion of 5-FDRP to 5-fluororibulose phosphate by the putative isomerase from *S. cattleya*.

1.4.3.4 An aldolase enzyme

The carbon-carbon bond breakage is likely to involve an aldolase. This part of the pathway is the least well characterised. The final steps in the pathway are relatively well understood and the enzymes responsible for the conversion of fluoroacetaldehyde to fluoroacetate and 4-FT have been identified (Murphy *et al.*, 2001a; Murphy *et al.*, 2001b).

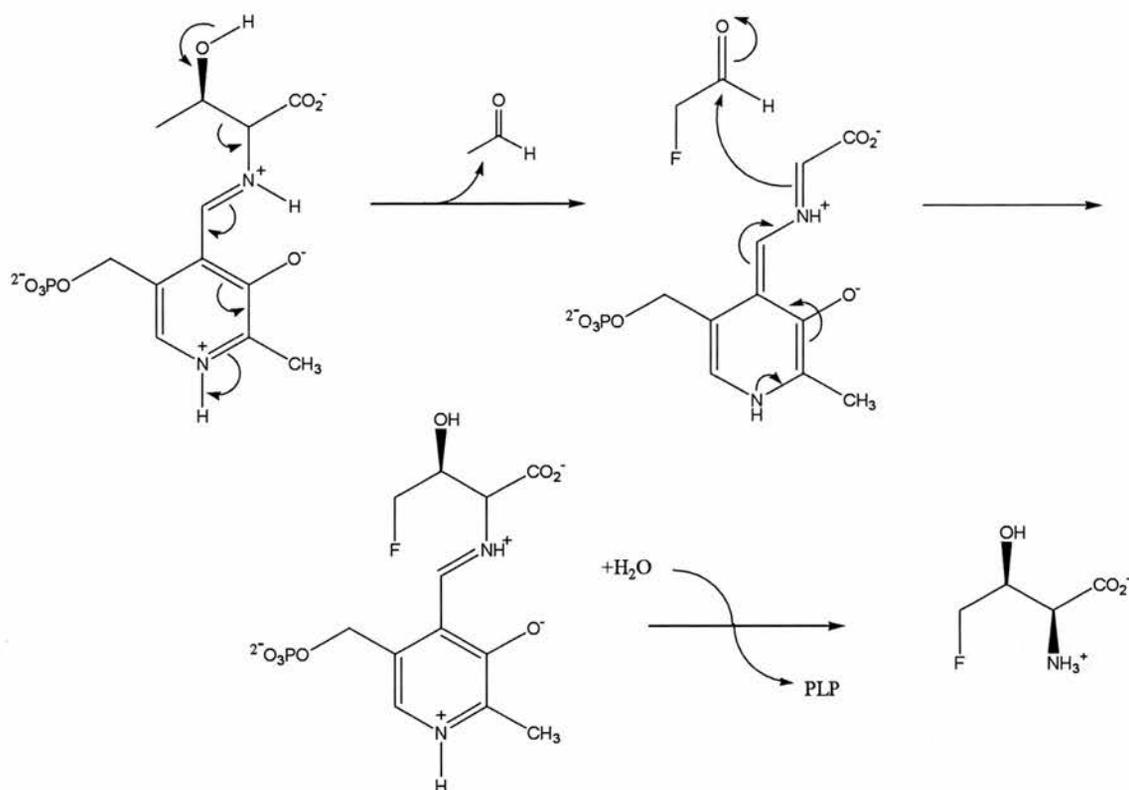
1.4.3.5 The aldehyde dehydrogenase

An aldehyde dehydrogenase permits conversion of fluoroacetaldehyde to fluoroacetate in the presence of nicotinamide adenine dinucleotide (NAD⁺) (Murphy *et al.*, 2001a). The dehydrogenase is thought to be a primary metabolic enzyme exploited by the fluorinase pathway (O'Hagan *et al.*, 2002).

The protein was subsequently purified from *S. cattleya* and revealed to be homologous to known bacterial acetaldehyde dehydrogenase enzymes. The enzyme exhibits a preference for fluoroacetaldehyde over acetaldehyde suggesting a specialised aldehyde dehydrogenase may be part of the pathway. This enzyme will accept chloroacetaldehyde as a substrate (Murphy *et al.*, 2001a).

1.4.3.6 The pyridoxal phosphate dependent transaldolase

A pyridoxal phosphate (PLP) dependent transaldolase (EC 2.2.1.8) converts fluoroacetaldehyde and L-threonine to 4-fluorothreonine demonstrated by studies into cell-free extracts (Murphy *et al.*, 2001b). (Scheme 1.9). This enzyme has also been shown to produce 4-chlorothreonine in the presence of L-threonine and chloroacetaldehyde.



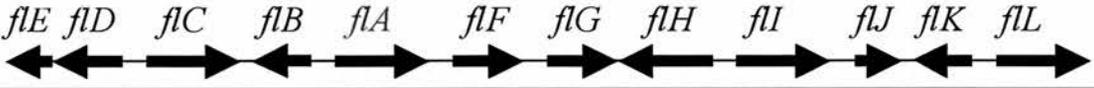
Scheme 1.9 – A mechanistic proposal for the conversion of L-threonine and fluoroacetaldehyde to 4FT and acetaldehyde by the PLP dependent threonine transaldolase (Murphy et al. 2001b)

Overall the fluorinase pathway represents an important step in the biosynthesis of fluorinated natural products. Synthesis of other fluorinated compounds may be possible through understanding of the mechanisms involved. This has implications in synthetic chemistry as it provides a means of stereospecific incorporation of fluorine.

1.4.4 The *flA* gene cluster

Although the complete genome of *S. cattleya* has not been sequenced, identification of the gene responsible for FDAS (*flA*) has permitted sequencing of genes

around it. The gene encoding the PNPase (*flB*) is adjacent to *flA*. Many other enzymes identified by chemical studies are not however adjacent to the gene. Table 1.2 shows the gene cluster surrounding *flA*.



ORF	START/STOP (bp)	Length (aa)	Function
E	130-795c	222	DNA binding regulatory protein*
D	875-1504c	216	Dehalogenase / Phosphatase*
C	1845-3036	397	MFS permease*
B	3057-3953c	299	PNPase
A	4173-5069	299	FDAS
F	5197-5751	185	DNA binding regulatory protein*
G	5951-6652	234	DNA binding regulatory protein*
H	6652-8052c	467	Na ⁺ / H ⁺ antiporter*
I	8314-9780	489	Homocysteine lyase*
J	9803-10195	131	DNA binding regulatory protein*
K	10592-10176c	139	Thioesterase / acetyltransferase*
L	10700-11374	225	DNA binding regulatory protein*

Table 1.2 – FDAS (*flA*) gene cluster in *S. cattleya* and gene locus map *functions are predicted by sequence, (Huang *et al.*, 2006).

Several of these putative proteins may be connected by function to FDAS. It is thought that the presence of several of these genes is vital for the regulation of the fluorometabolite production. For example *flI* codes for a putative *S*-adenosyl-L-homocysteine hydrolase (homocysteine lyase) will regulate homocysteine (a recognised inhibitor of FDAS) within the cell (Huang *et al.*, 2006).

1.5 The fluorinase (FDAS) enzyme

1.5.1 Identification of FDAS

The most novel enzyme in the pathway is FDAS and recent investigation has revealed much about its function. The presence of a fluorinase enzyme in *S. cattleya* had been predicted in 1998 (Hamilton *et al.*, 1998). The enzyme is expressed after several days in *S. cattleya* when grown in the presence of fluoride and was shown to have a relative mass of 32.2 kDa. The protein was shown to catalyse the transformation of *S*-adenosyl-L-methionine and inorganic fluoride to generate 5'-FDA (Schaffrath *et al.*, 2003).

1.5.2 Gene sequence and protein sequence

The FDAS gene (*flA*) was identified by Spencer and co-workers (as indicated previously). From the sequence the protein appears to belong to a conserved sequence family found in Archaea and Bacteria (Dong *et al.*, 2004).

A long N-terminal loop [-AAKGGARGOWAGSGAGFERAEGS-] which transverses domains (shown in the crystal structure below) and is heavily involved in FDAS trimer contacts, is missing in all homologues (Figure 1.5). This loop has been proposed as functionally important as it contributes to the active site architecture (Dong *et al.*, 2004). It is stated that the loop is responsible for 25 % of the buried surface area of the protein and several of the hydrogen bonding interactions of the trimer interface. The recombinant FDAS enzyme was overexpressed, purified and was shown to retain full activity (Dong *et al.*, 2003).

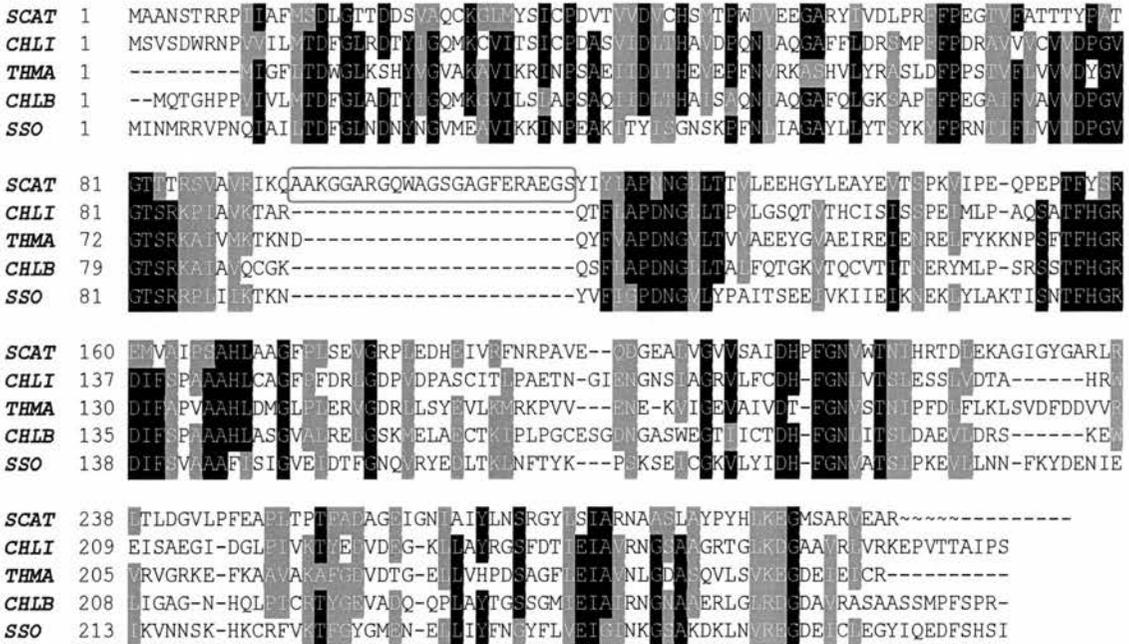


Figure 1.5 – Sequence alignment of FDAS and homologues identified using BLAST search. The trans-domain loop in *S. cattleya* (SCAT) is highlighted (red). **CHLI:** *Chlorobium limicola*, **THMA:** *Thermotoga maritima*, **CHLB:** *Chlorobium phaeobateroides*, **SSO:** *Sulfolobus solfataricus*. The red box highlights a loop region only found in FDAS, black shading are identical and grey shading are similar.

1.5.3 Protein structure

X-ray crystal structures of FDAS with substrate (SAM) and product (5'-FDA) have been reported (Dong *et al.*, 2003). The SAM complex was obtained from the protein purified from *S. cattleya*. The product complex used the recombinant material incubated with substrates.

The structure revealed that the protein is a dimer of trimers, consistent with the hexamer predicted from size exclusion chromatography (Dong *et al.*, 2003). The structure is composed of a novel fold (Dong *et al.*, 2004). The monomer consists of two

domains, a large α/β fold domain at the N-terminal and a small β -barrel domain at the C-terminus (Kozbial and Mushegian, 2005).

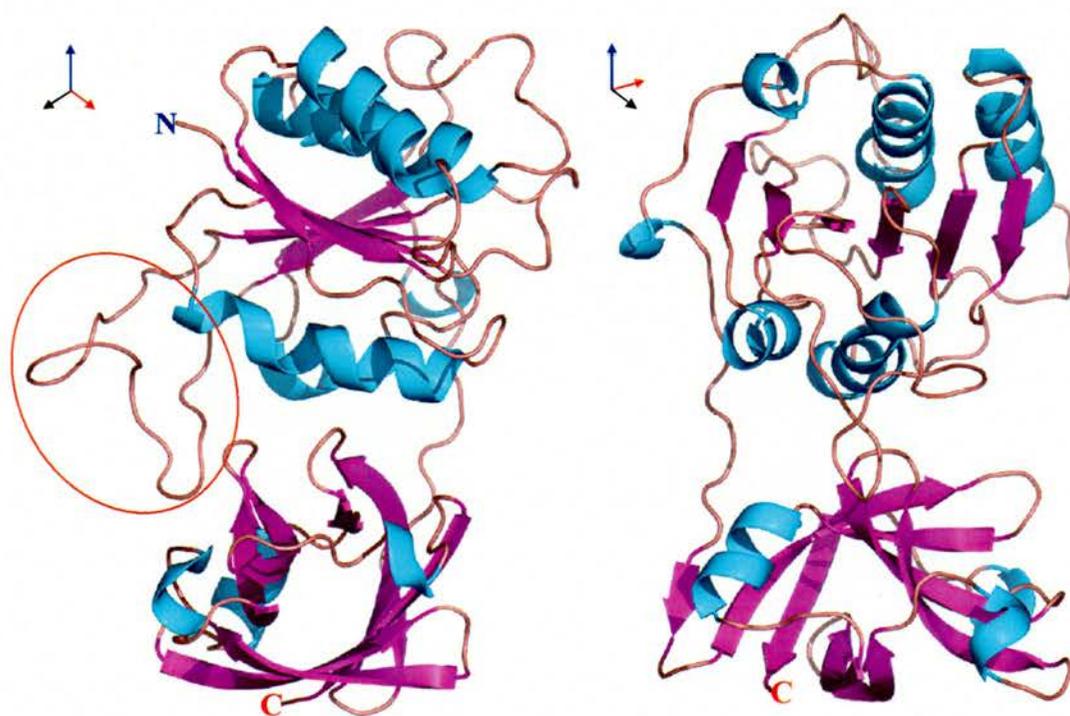


Figure 1.6 – Cartoon representations of FDAS monomer (rotated 90°). In the left hand image, the red circle highlights the trans-domain loop (mentioned previously)

The trimer (Figure 1.7) has a central three-fold axis, in which the C-terminal domain of one monomer forms a complex with the N-terminal domain of a second monomer. The active site is created at the interface between monomers (three per hexamer).

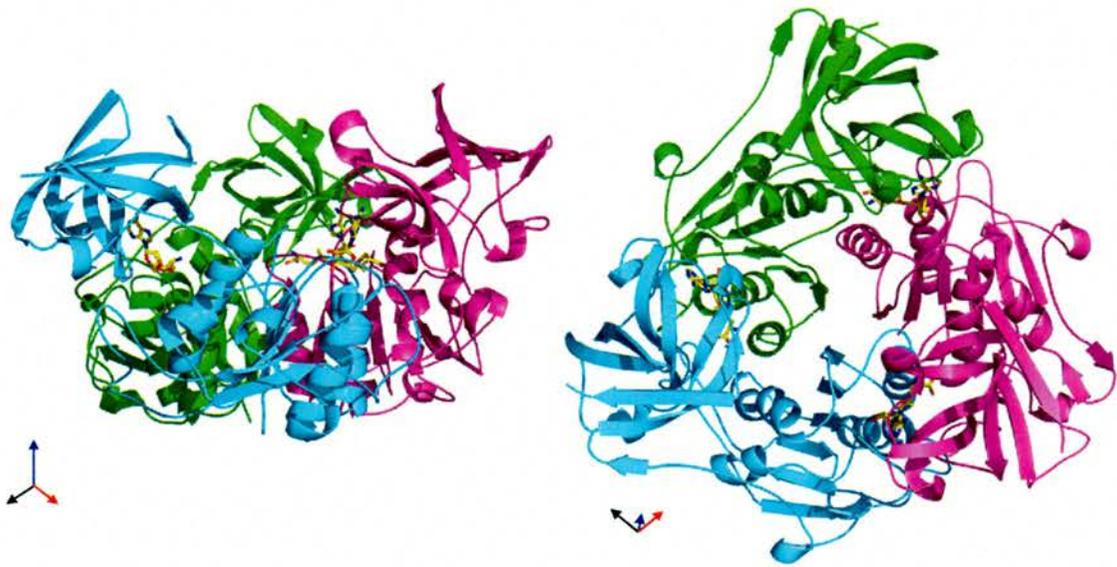


Figure 1.7 – Cartoon representation of the FDAS trimers shown in two orientations.



Figure 1.8 – Cartoon representation of the hexamer model, as constructed from crystallographic symmetry using 1rqp (each monomer is coloured separately).

The hexamer is constructed so that the small C-terminal domains of one trimer interacts with the same domains of the second trimer (Figure 1.8).

1.5.4 Structure of substrate in active site

In the SAM complex model the ribose ring conformation of the adenosine portion is constrained in an atypically planar conformation. This was suggested to result from hydrogen bonding interactions between the Asp-16 and the 2', 3' hydroxyls of the ligand and strain of binding (Dong *et al.*, 2004). Further hydrogen bonds to the ribose hydroxyls are made by Tyr-77, Trp-50 and Ser-158.

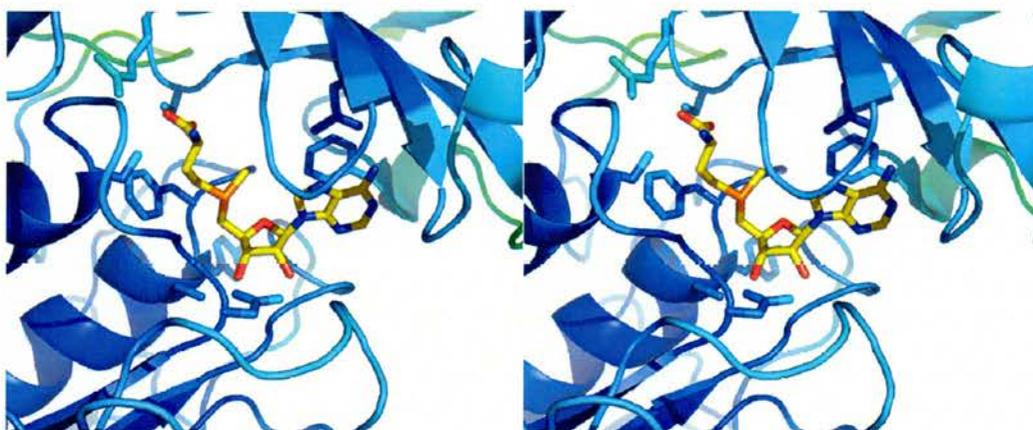


Figure 1.9 – *S-adenosyl-L-methionine (SAM) in the active site of FDAS (1rqp).*

The adenine moiety is bound by interactions with Asn-215, Phe-254, Ala-279 and Trp-50. The methionine moiety forms interactions with Asp-21, Ser-23, Ser-269, Arg-270 and Leu-17.

1.5.5 Structure of product in active site

The model of products (5'-FDA and L-Met) of the FDAS reaction in the active site of FDAS was obtained (to 2.68 Å) by co-crystallisation of SAM in the presence of fluoride ion (Figure 3.9).

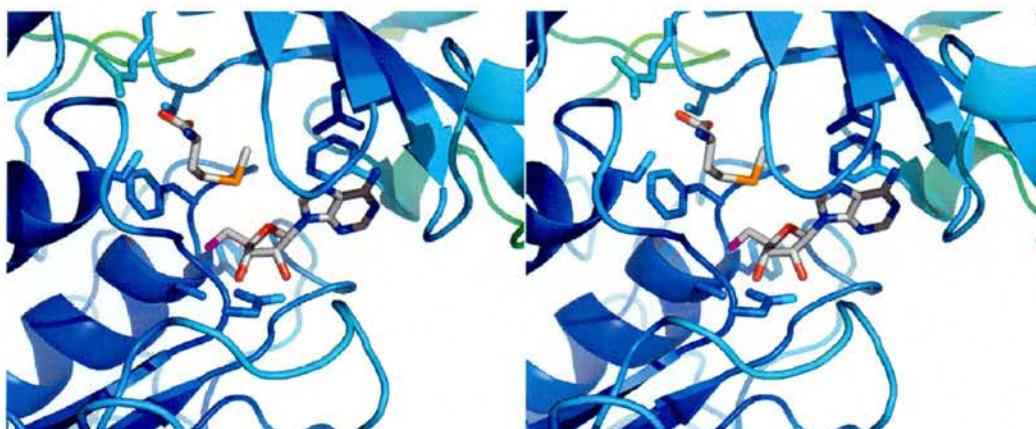


Figure 1.10 – 5'-fluoro-5'-deoxyadenosine (5'-FDA) in the active site of FDAS (1rqr)

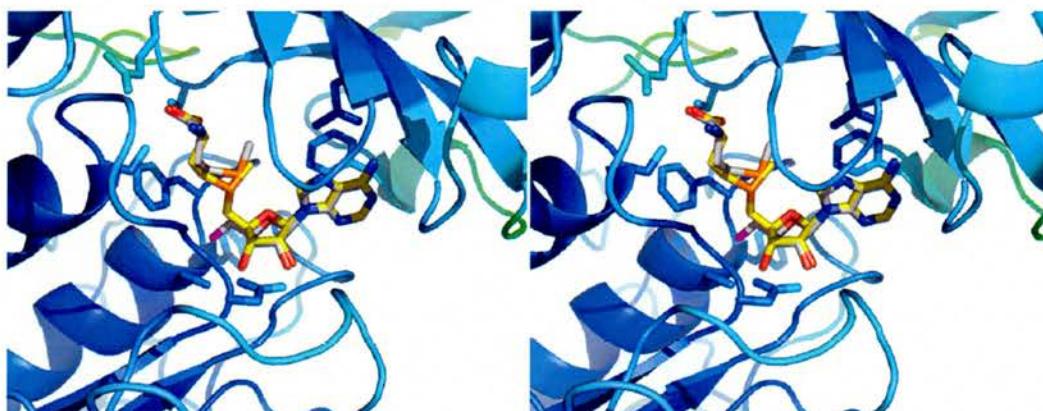


Figure 1.11 – Superimposition of substrate and product in the active site.

Comparison of these models (1rqp and 1rqr) revealed few differences in the overall positions of substrate and products (shown in Figure 1.11). The L-Met sulfur was displaced (~ 1.6 Å) relative to the C5' of the substrate.

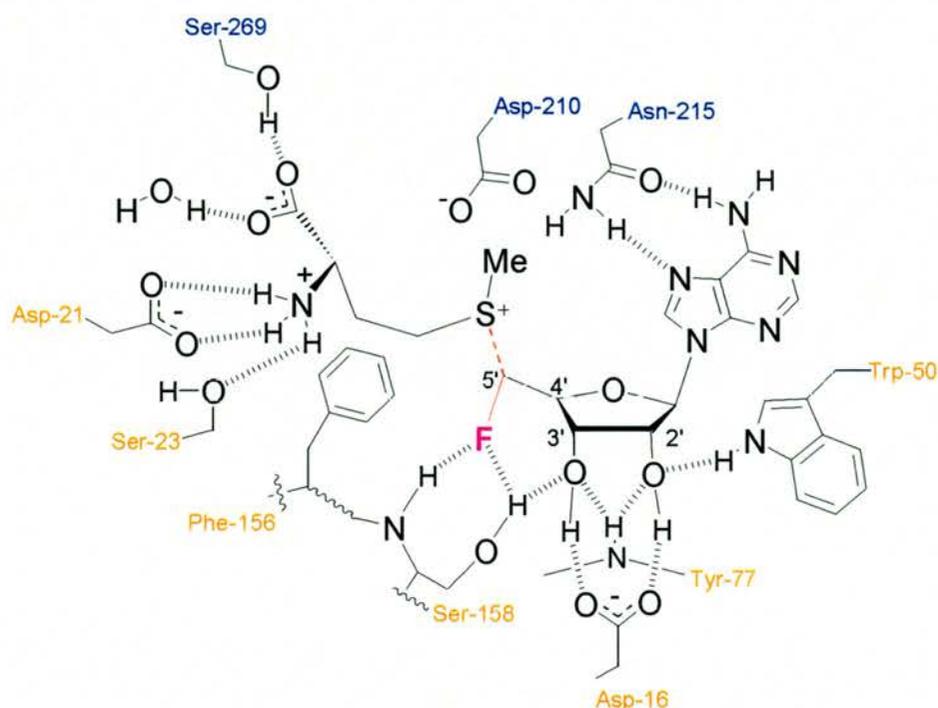


Figure 1.12 – Representation of 5'-FDA and L-Met bound to FDAS. The graphic highlights two hydrogen-bonds to the fluoro-methyl group from Ser-158. There are hydrogen bonding interactions also between Asp-16 and the 2'- and 3'-OH groups of the ribose ring.

1.5.6 Reaction mechanism

The X-ray crystal structure (Dong *et al.*, 2004) and stereochemical analysis (Cadicamo *et al.*, 2004) indicated that the conversion of SAM to 5'-FDA in FDAS is consistent with an S_N2 type reaction mechanism (Figure 1.13). This was corroborated by QM and QM/MM calculation studies on the structure (Senn *et al.*, 2005).

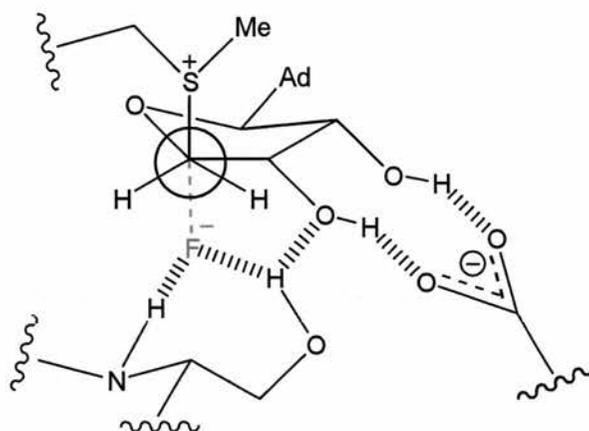
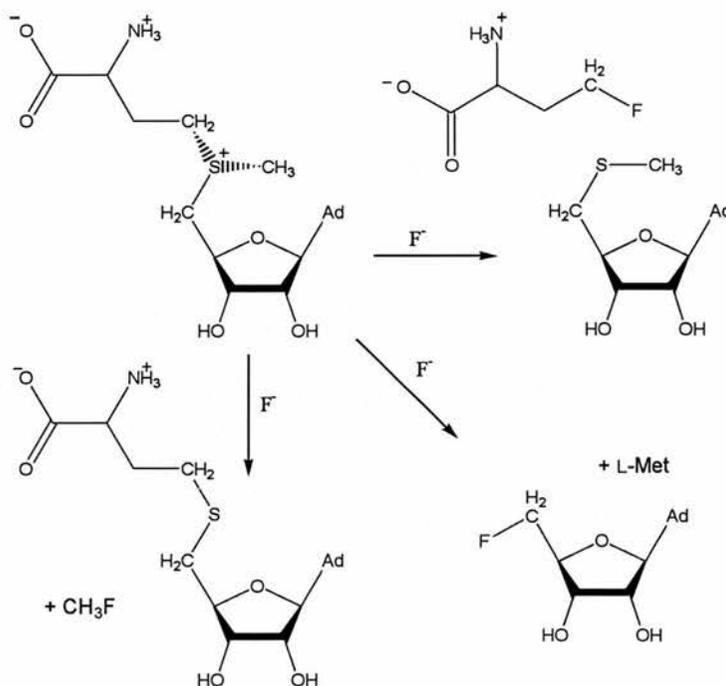


Figure 1.13 – The S_N2 catalysed reaction of F^- (coordinated by Ser-158) and SAM at the C5' position to produce 5'-FDA and L-methionine via FDAS

The intrinsic reactivity of SAM towards fluoride was investigated. There are three equally favourable potential reactions (Senn *et al.*, 2005) (Scheme 1.10).



Scheme 1.10 – Nucleophilic attack of F^- on SAM at the three possible α -carbons joined to the sulfonium leaving group.

It follows since there is no thermodynamic selectivity, the fluorination is kinetically controlled, which means the stereochemistry at the enzyme is crucial. Theoretical calculations agree with the hypothesis from structural studies that F^- is desolvated and stabilised as an active nucleophile. A theoretical study (Senn *et al.*, 2005) based on the existing FDAS models identifies Thr-80 and Ser-158 as crucial in the process during calculated minimisation of a reactant model (where F^- is inserted manually). In this model Thr-80 and Ser-158 appear to rearrange so the respective hydroxyl groups become oriented towards the anion. The authors stated that major role of the enzyme is in the preparation and positioning of the substrates. Overall the protein is calculated to lower the barrier for nucleophilic attack by $\sim 40 \text{ kJ mol}^{-1}$, since the S_N2 activation energy is predicted to be 92 kJ mol^{-1} in solution but only 50 kJ mol^{-1} in the enzyme.

1.6 Enzymatic chlorination

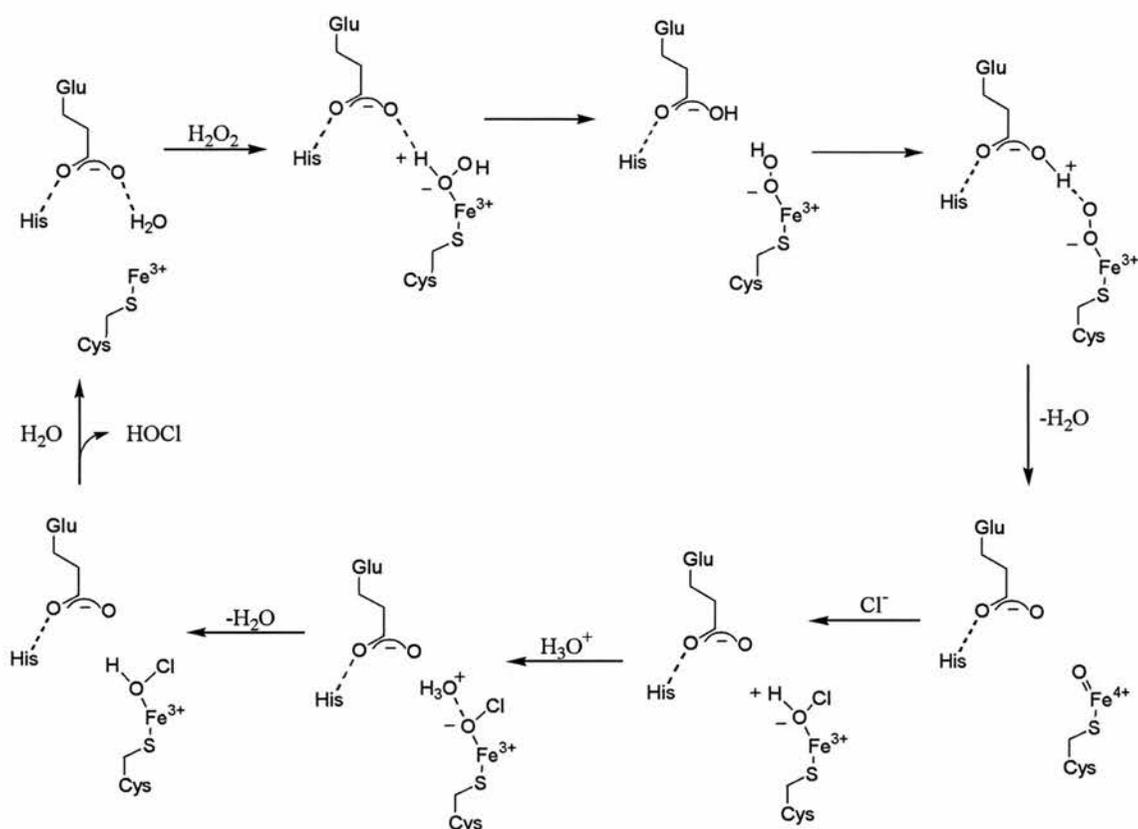
There are many enzymes capable of non-fluorine halogenation. This mirrors the relative abundance of chlorinated or brominated natural compounds. Until recently haloperoxidases were thought to be responsible for the formation of organo-chlorides (Littlechild, 1999). Renewed interest in this field over the last decade has revealed several different mechanisms are utilized.

1.6.1 Haloperoxidases and perhydrolases

The first halogenase was identified during investigation of caldariomycin biosynthesis in *Caldariomyces fumago* (Beckwith *et al.*, 1963). The enzyme was shown to chlorinate late intermediates in the biosynthetic pathway, in the presence of hydrogen peroxide and chloride ion. The enzyme can also use bromine and iodine.

Haloperoxidases are generally classified by halospecificity and are termed chloroperoxidases (capable of transferring Cl, Br and I), bromoperoxidases (Br and I) and iodinases (I).

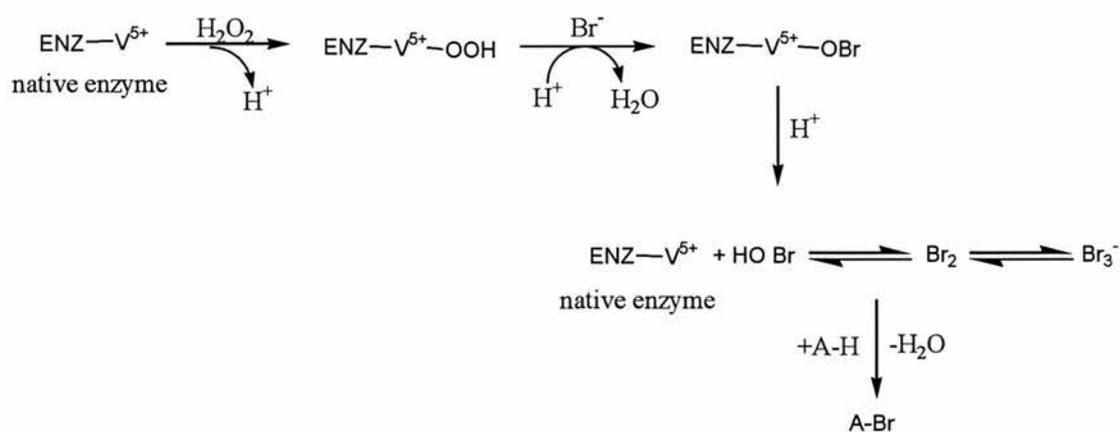
The chloroperoxidase (*Caldariomyces fumago*) was subsequently shown to be a haem dependent glycoprotein similar to iron peroxidases (Morris and Hager, 1966). A reaction mechanism for haem containing haloperoxidase was proposed based on the protein crystal structure of the *Caldariomyces fumago* enzyme (Sundaramoorthy *et al.*, 1998; Wagenknecht and Woggon, 1997). Chlorination is thought to occur after release of hypohalous acid (HOCl) generated at the haem (Scheme 1.11) (van Pee and Unversucht, 2003).



Scheme 1.11 – Reaction mechanism of haem containing haloperoxidase showing the formation of hypohalous, the porphyrin ring system is omitted.

Haem dependent haloperoxidases are not the only type of enzyme capable of chlorination, some peroxidases in fungi and algae are shown to contain vanadium (Vilter, 1995). Other halogenating enzymes isolated from bacteria apparently require hydrogen peroxide but had no metal cofactor (Burd *et al.*, 1995; Wiesner *et al.*, 1986).

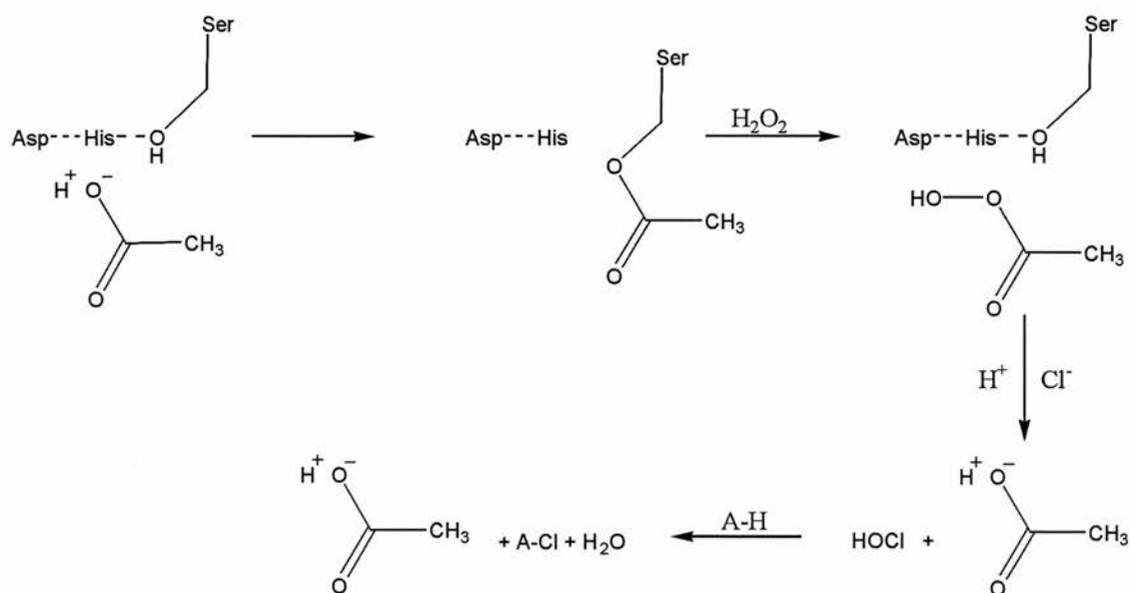
The vanadium dependent halogenase reaction is thought to follow the mechanism outlined in Scheme 1.12, and also involves production of HOBr (or HOCl) (Almeida *et al.*, 2001; van Pee and Unversucht, 2003).



Scheme 1.12 – Reaction mechanism of vanadium containing haloperoxidase

1.6.2 Perhydrolases

Perhydrolases have been shown to be a further type of enzyme shown to be capable of halogenation (van Pee, 1996). Such enzymes have been identified in strains of *Pseudomonas* and *Streptomyces*, and have been shown to contain neither a haem nor metal containing cofactor.

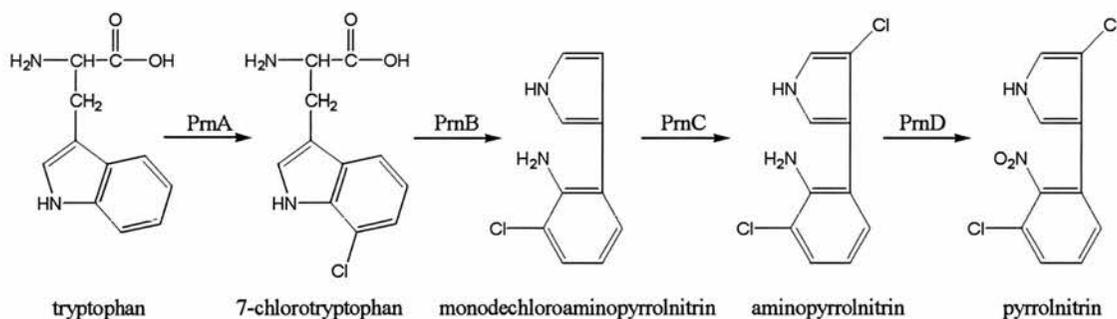


Scheme 1.13 – Perhydrolase reaction mechanism

In the presence of hydrogen peroxide (H_2O_2) a short chain carboxylic acid can be transformed, via an acyl-enzyme intermediate, to a peracid by perhydrolysis. The peracid is then able to oxidize a chloride to produce hypohalous acid (Scheme 1.13) (van Pee and Unversucht, 2003). These enzymes however are not thought to be involved in halometabolite biosynthesis (Dong *et al.*, 2005).

1.6.3 Regioselective chlorination

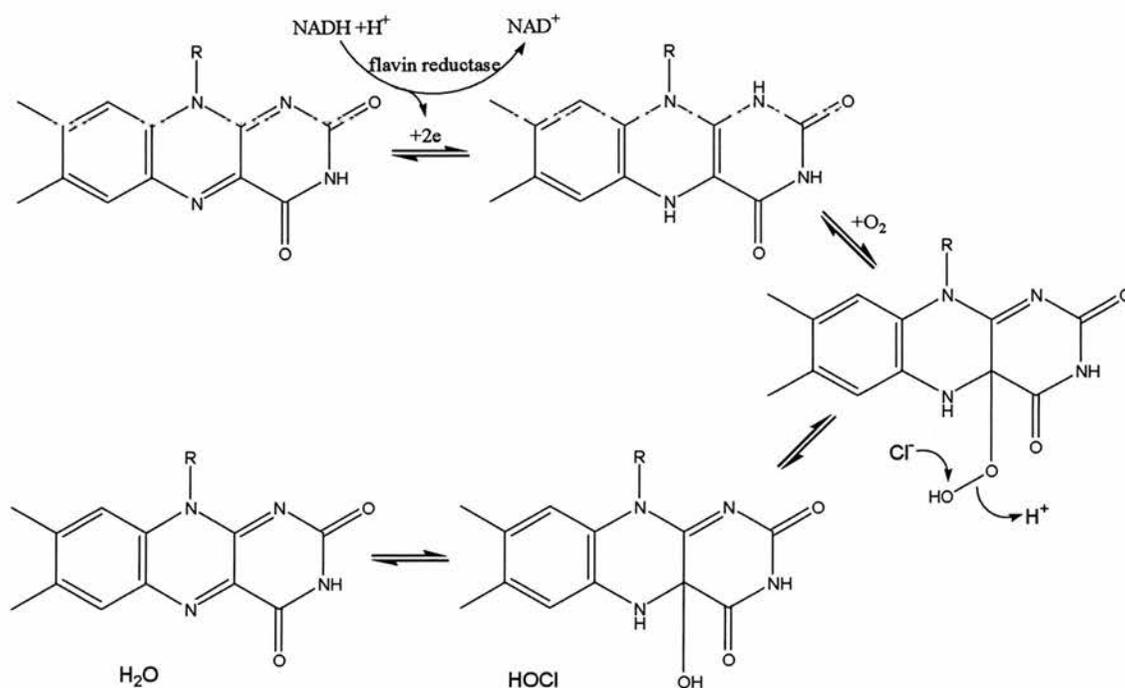
The first evidence for a regioselective chlorinase came from the elucidation of the gene cluster for 7-chlorotetracycline biosynthesis (Dairi *et al.*, 1995). In this cluster there were no enzymes for peroxidases or perhydrolases but a novel flavin protein was seen.



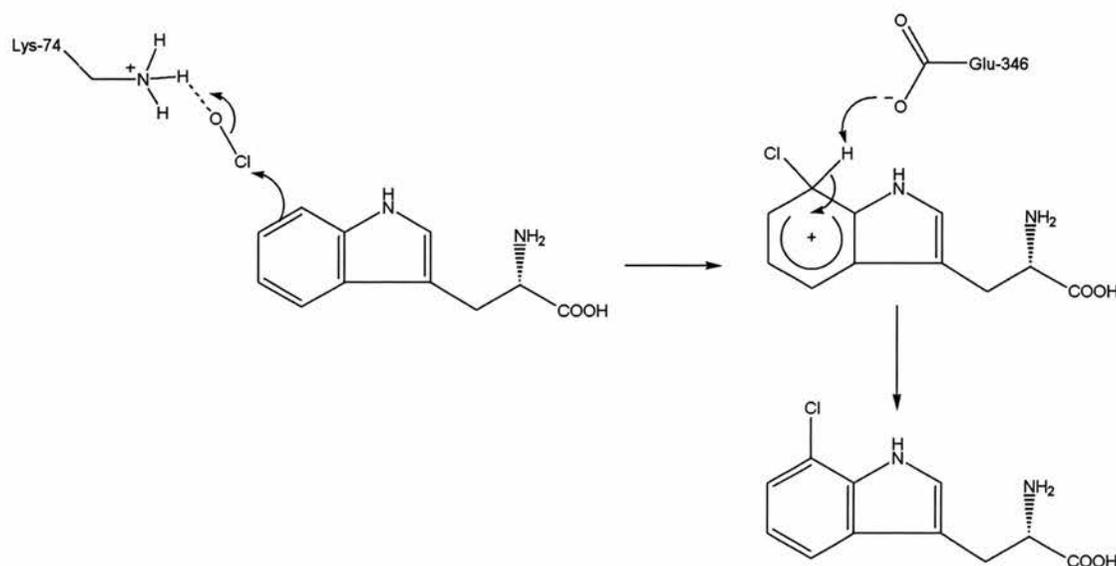
Scheme 1.14 – Pyrrolnitrin biosynthetic pathway.

Sequencing of the pyrrolnitrin gene cluster identified four enzymes (*PrnA*, *B*, *C* and *D*) of which (*A* and *C*) belong to the same flavin dependent family of enzymes (Scheme 1.12) (Hammer *et al.*, 1997; Kirner *et al.*, 1998). These are dependent on $FADH_2$ as a cofactor, molecular oxygen and halide ion (Cl^- or Br^-) for activity and have been shown to be capable of regiospecific chlorination. A reaction mechanism based on

the protein crystal structure of PrnA suggested a mechanism for regioselective chlorination (Dong *et al.*, 2005).

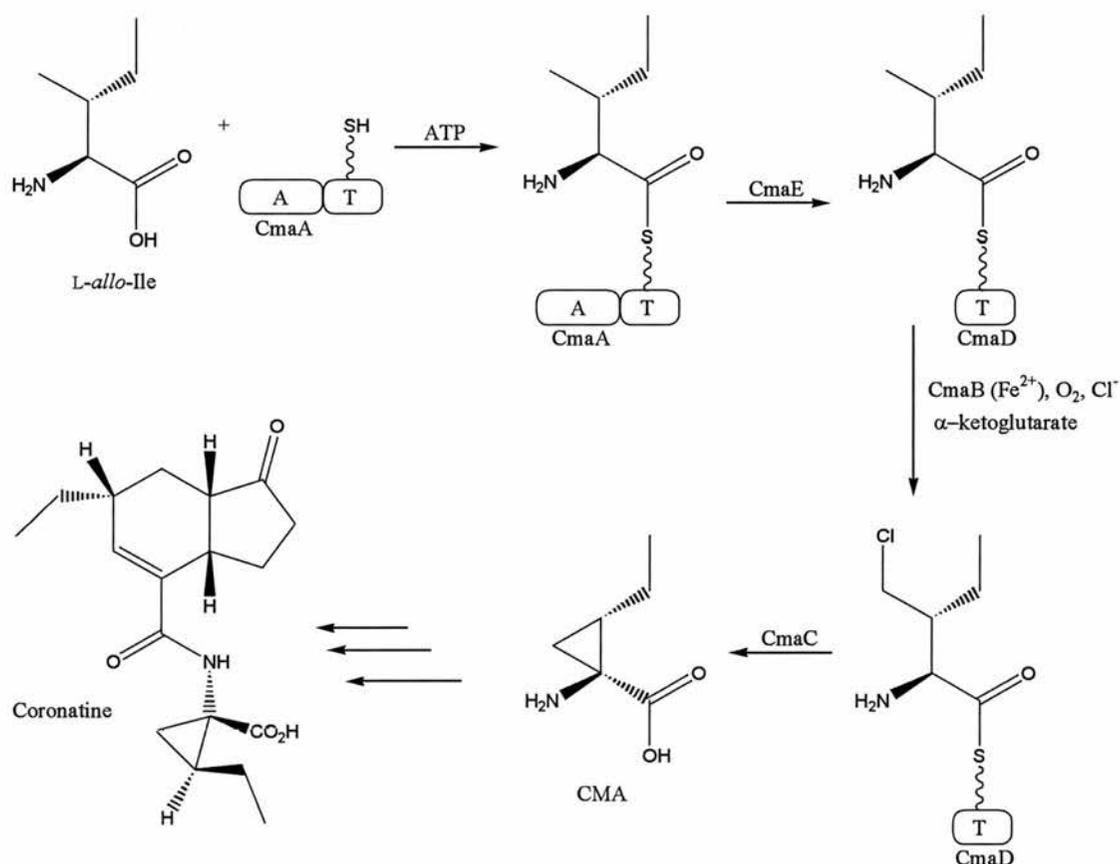


Scheme 1.15 – HOCl generation via by FADH₂ in PrnA



Scheme 1.16 – Reaction mechanism of PrnA in 7-chlorotryptophan synthesis.

An enzyme (CmaB) has been shown to perform chlorination of an intermediate involved in cyclopropyl amino acid (CMA) biosynthesis (Vaillancourt *et al.*, 2005a). The reaction mechanism is outlined in Scheme 1.18, and is thought to involve radical formation by a non haem Fe^{2+} , O_2 and α -ketoglutarate. CMA is a precursor produced during coronatine biosynthesis in *Pseudomonas syringae* (as summarised in Scheme 1.18). Two other related enzymes have been identified in syringomycin biosynthesis (SyrB2 - *Pseudomonas syringae*) and barbamide biosynthesis (BarB1/B2 - *Lyngbya majuscula*).



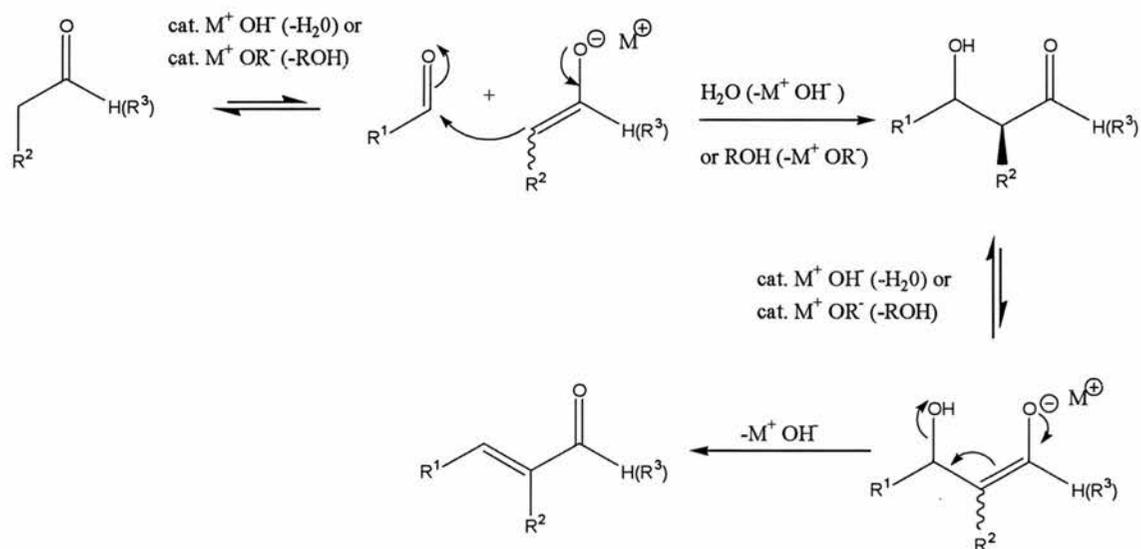
Scheme 1.18 – Summary of CMA biosynthesis, a precursor of coronatine CmaA (A) and CmaD (T) complex prior to binding Ile.

To conclude several enzymes are capable of chlorination. Two such enzymes use an oxygenation type mechanism, effectively generating Cl^+ . A third uses a radical based mechanism. A fourth class of chlorinase does appear to use a nucleophilic mechanism but is limited to methyl transfer (Attieh *et al.*, 1995; Wuosmaa and Hager, 1990); and little is known about it. A robust nucleophilic chlorinating enzyme would open up significantly the range of compounds which could be prepared by biotransformation. The ability of FDAS to use chlorine by a nucleophilic mechanism is investigated and discussed in Chapter 2.

1.7 Aldolases and enzymatic aldol condensation

1.7.1 The aldol reaction

The aldol reaction can be described as the reaction of two carbonyl compounds where one is a nucleophile and the other is the electrophile (March and Smith, 2001). The mechanism of the base catalyzed aldol reactions is shown in Scheme 1.19.



Scheme 1.19 – Base-catalysed aldol reaction

Asymmetric aldol reactions are interesting since they carry out carbon – carbon bond formation. In synthetic organic chemistry such reactions are challenging as they often require auxiliary groups to obtain enantiomerically pure product. Control can be obtained through use of enzymes which are by their nature regio-selective and stereo-selective.

1.7.2 Aldolases

Aldolases catalyse the electrophilic substitution at the α -carbon of an enol or enolate anion. Aldolases are classed as lyases and are ubiquitous (Horecker, 1972), involved specifically in the metabolism of carbohydrates, amino acids and hydroxy acids.

Two types of aldolases have been identified (Rutter, 1964). Class I aldolases are ubiquitous in nature forming a Schiff base with the donor substrate (Verlinde and Quigley, 1999). Class I aldolases have been identified in humans, containing A, B and C isozymes (Voet and Voet, 2004) expressing unique activity profiles and tissue specificity. Class II aldolases, are mainly found in prokaryotes (and lower eukaryotes) and they require a Zn^{2+} . More than 30 different aldolases have been characterised (Gefflaut *et al.*, 1995). These can be further categorized depending on the carbon donor group (Figure 1.14), the dihydroxyacetone phosphate (DHAP) dependent, pyruvate dependent, phosphoenolpyruvate (PEP) dependent, glycine dependent and acetaldehyde dependent aldolases.

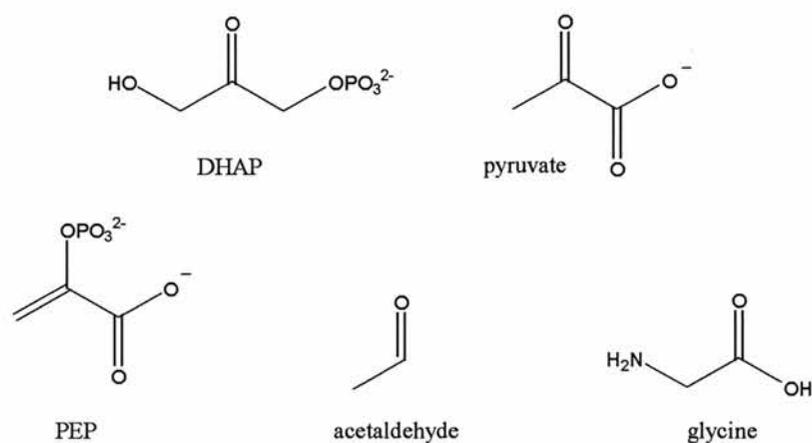


Figure 1.14 – Examples of aldolase donor groups.

1.7.2.1 DHAP aldolase

DHAP dependent aldolases are widely studied (Espelt *et al.*, 2003). These include fructose-1,6-bisphosphate aldolase which accepts glyceraldehyde-3-phosphate as a natural substrate. Other DHAP dependent aldolases include FruA (*Staphylococcus carnosus*) (Witke and Gotz, 1993), L-rhamnulose-1-phosphate aldolase (L-RhuA - *E. coli*) (Vidal *et al.*, 2005) and the L-fuculose-1-phosphate aldolase (L-FcuA - *E. coli*) (von der Osten *et al.*, 1990). Such aldolases create two new stereogenic centres and can produce four possible stereo-isomers (Schoevaart *et al.*, 2000), shown in Figure 1.15.

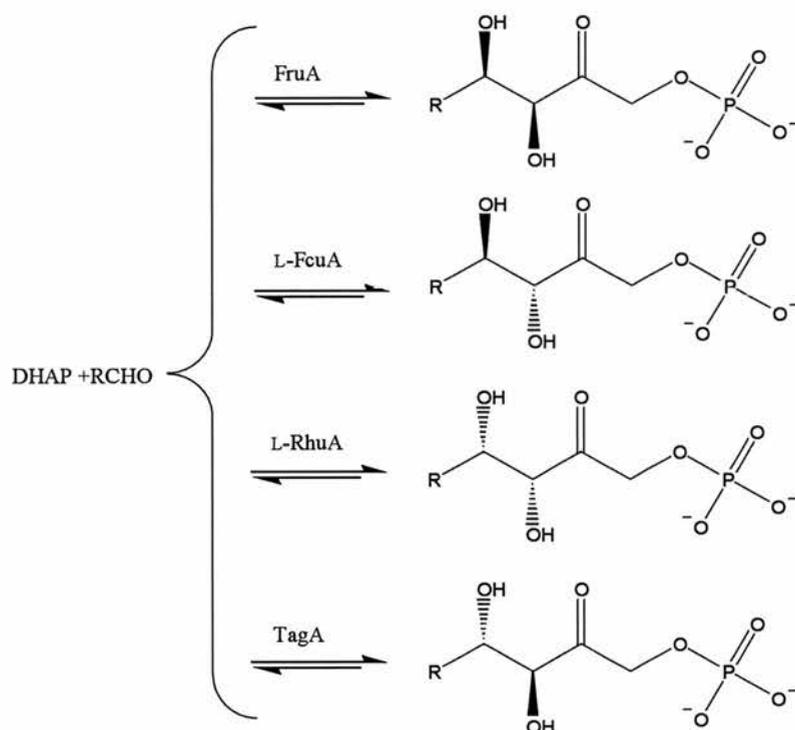
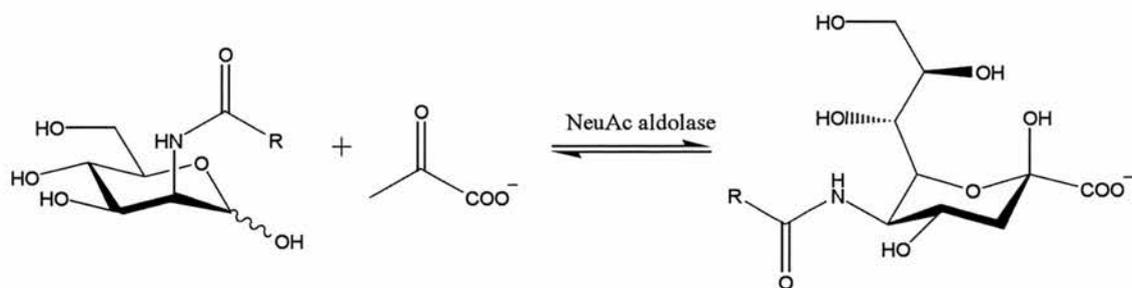


Figure 1.15 – DHAP dependent aldolase reactions

1.7.2.2 Pyruvate and PEP dependent aldolases

Pyruvate and PEP dependent aldolases typically only occur in prokaryotes. Examples of aldolases which use pyruvate (donor) are N-acetylneuraminic acid (NeuAc) aldolase, 2-keto-3-deoxy-6-phosphogluconate-6-phosphate (KDPG) aldolase, and 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) aldolase (discussed in Chapter 4). The NeuAc aldolase is well studied (Rodriguez-Aparicio *et al.*, 1995) and condenses *N*-acetyl-D-mannosamine and pyruvate (Scheme 1.20).



Scheme 1.20 – NeuAc (lyase) aldolase reaction

The PEP dependent aldolases, such as KDO8P synthase, transfer the enolpyruvyl group to various acceptors. Incorporation of PEP is often the first step in amino acid synthesis pathways and is the first committed step in the shikimate pathway (Sundaram and Woodard, 2000). Directed evolution of KDPGal aldolase has shown that pyruvate can replace PEP in systems where DHAP synthase activity is silenced (Ran *et al.*, 2004), suggesting a model for aromatic biosynthesis which may exist in nature.

1.7.2.3 Glycine dependent aldolases

Glycine (Gly) dependent aldolases form β -hydroxy- α -amino carbionic acids, using PLP as a cofactor. These include threonine aldolase and serine hydroxymethyl transferase (Liu *et al.*, 1998; Ogawa *et al.*, 2000). The PLP dependent threonine transaldolase in *Streptomyces cattleya* (Murphy *et al.*, 2001b) was described in Section 1.4.3.1.

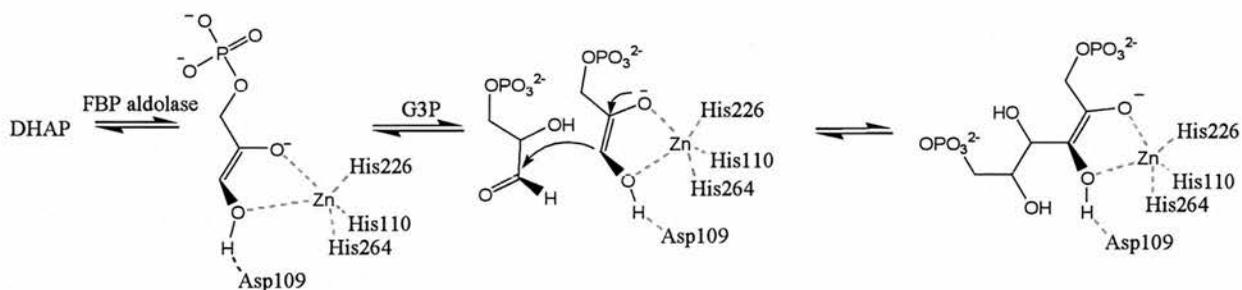
1.7.2.4 DERA

The 2-deoxy-D-ribose-5-phosphate aldolases (DERA) are acetaldehyde dependent aldolases of which one type has been identified (Machajewski and Wong, 2000). DERA catalyses the synthesis of 2-deoxyribose-5-phosphate from acetaldehyde and G3P. This

is the only example of an aldolase which combines two different aldehydes resulting in the formation of an aldehyde as a product, which can undergo a further aldol condensation acting as the acceptor substrate. These enzymes are tolerant towards diversity in the acceptor substrate and show some variety in donors (i.e. propanal, acetone and fluoroacetone) (Wong *et al.*, 1992).

1.7.2.5 Class II aldolases

Class II aldolases are related to class I aldolase enzymes by function but are separated by evolution (Marsh and Lebherz, 1992). An example of a class II aldolase is fructose-1,6-bisphosphate (FBP) aldolase from *E. coli* (Hall *et al.*, 1999). This enzyme chelates DHAP at the divalent zinc, which results in carbanion formation through deprotonation of DHAP by an active site base. The nucleophile undergoes reaction with glyceraldehyde-3-phosphate (G3P) to form the product (as shown in Scheme 1.21).



Scheme 1.21 – Reaction mechanism for the class II FBP aldolase from *E. coli*.

Aldolases have obvious applications in the commercial synthesis of enantiomerically pure bioactive molecules, such as carbohydrates. Therefore a clear understanding of novel aldolases can help in the development of novel catalysts. An understanding of protein structure at the atomic level can give great insight into function and subtle differences in mechanism.

2

Structural Studies into 5'-Fluorodeoxyadenosine Synthase;

A Novel Naturally Occurring Fluorinating Enzyme

2.1 SUMMARY

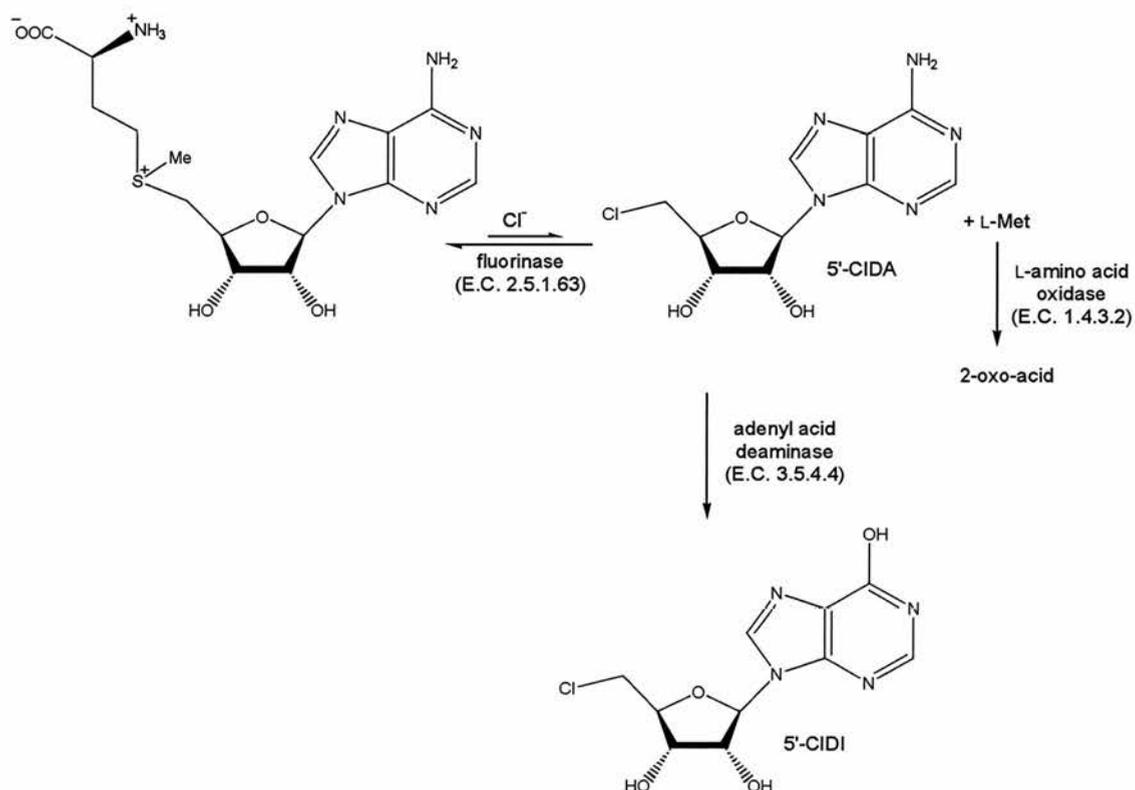
The enzyme 5'-fluorodeoxyadenosine synthase (FDAS) is shown to synthesise 5'-fluoro deoxyadenosine (5'-FDA) from *S*-adenosyl-L-methionine (SAM) and inorganic fluoride. Recent studies show FDAS also to be capable of turning over other substrates, in particular 5'-chloro-5'-deoxyadenosine (5'-CIDA) and 2'-deoxy analogues but not 3'-deoxy analogues.

Adenosine was identified as a contaminant, which was not removed in standard purification of the recombinant protein. The purification protocol was revised to remove adenosine. Structures (2.0 to 2.5 Å) of adenosine, 5'-CIDA, 5'-deoxy-5'-[N-methyl-N-(γ -amino- γ -carboxypropyl)]-adenosine, β -D-erythrofuransyl adenosine, 5'-fluoro-5',2'-dideoxyadenosine and 2'-deoxyadenosine, were obtained in the active site of FDAS. This chapter discusses several structures of substrates and analogues in the active site of FDAS and relevance to the enzyme mechanism. Crystallization conditions for the apo-form of the enzyme were identified and the apo-structure was obtained to 2.5 Å resolution. A thermodynamic analysis using isothermal titration calorimetry (ITC) was carried out to investigate binding interactions of several ligands to FDAS. The structures and thermodynamic data contribute to the overall understanding of the enzyme mechanism. These findings open up the repertoire of compounds that can be synthesised through FDAS and reveals a novel mechanism of enzymatic chlorination.

2.2 EXPERIMENTAL INTRODUCTION

2.2.1 FDAS as a chlorinase

It had been reported that FDAS did not utilise chlorine and in the structure the halogen binding site appeared to be too small to accept chlorine (Dong *et al.*, 2004). The experimental evidence has evolved and FDAS does in fact accept chloride as a substrate (Deng *et al.*, 2006). The lack of activity was due to an equilibrium of which lies almost entirely with substrate. Incorporation of chloride is only seen when a coupled assay is used which moves the equilibrium towards product. Two separate coupled enzyme assays are described in Scheme 2.1.



Scheme 2.1 – Coupled enzyme assays with driving synthesis from SAM and inorganic chloride through 5'-CIDA.

SAM and chloride in the presence of FDAS yields 5'-CIDA (Figure 2.1), in the presence of L-amino acid oxidase (*Bothrops atrox* venom, Sigma Chem. Co. Ltd). The L-amino acid oxidase converts L-methionine (L-Met) to 2-oxo-methionine, removing the substrate for the reverse reaction. Addition of an adenyl acid deaminase (*Aspergillus sp.*) to the FDAS reaction with chloride also drives the equilibrium from substrate to product, as 5'-CIDA is converted to 5'-chloro-5'-deoxyinosine (5'-CIDI).

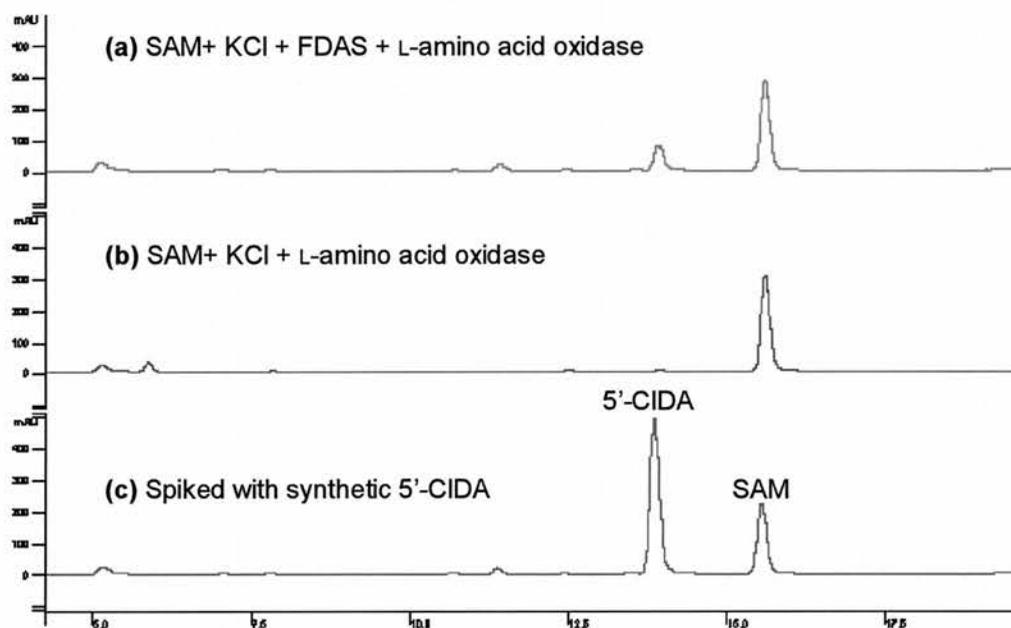
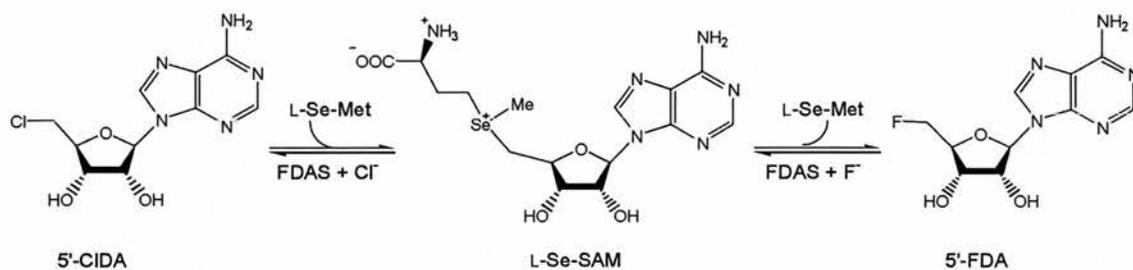


Figure 2.1 - HPLC profiles after running the FDAS reaction (FDAS, SAM and Cl) with and without the oxidase. **(a)** The result of FDAS and amino acid oxidase combination, showing 5'-CIDA production (after 12 h, RT). **(b)** A control without FDAS. **(c)** As for chromatograph (a) spiked with synthetic 5'-CIDA (Deng et al., 2006).

The reversibility of FDAS was confirmed by generation of [^{13}C -methyl]-SAM (or Se-SAM) from L-[^{13}C -methyl]-methionine (or SeMet) and 5'-FDA. This led to the

development of the transhalogenation reaction (Scheme 2.2) which demonstrates the conversion of 5'-CIDA to 5'-FDA via *Se*-adenosyl-L-selenomethionine (L-*Se*-SAM).



Scheme 2.2 – Transhalogenation reaction of 5'-CIDA to 5'-FDA using L-SeMet as the co-catalyst

Accumulation of 5'-FDA was monitored by analytical HPLC (shown in Figure 2.2). The analysis shows that this reaction favours the production of 5'-FDA over 5'-CIDA (in the presence of FDAS and L-SeMet).

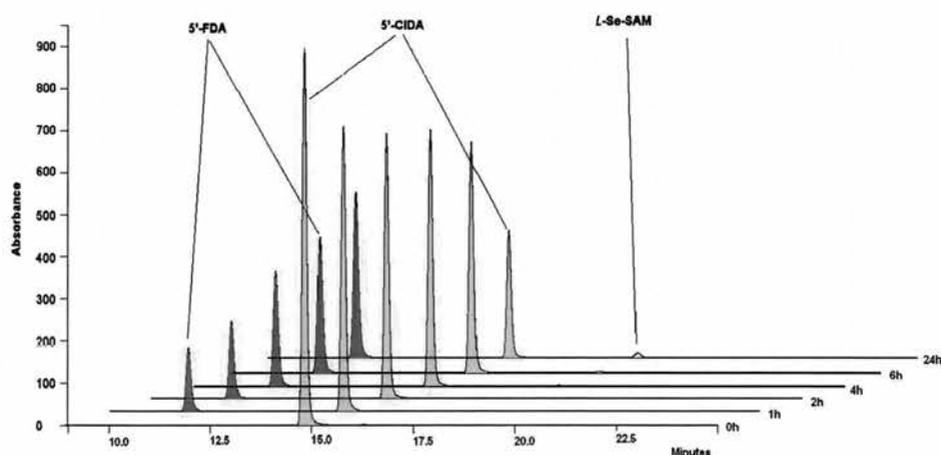
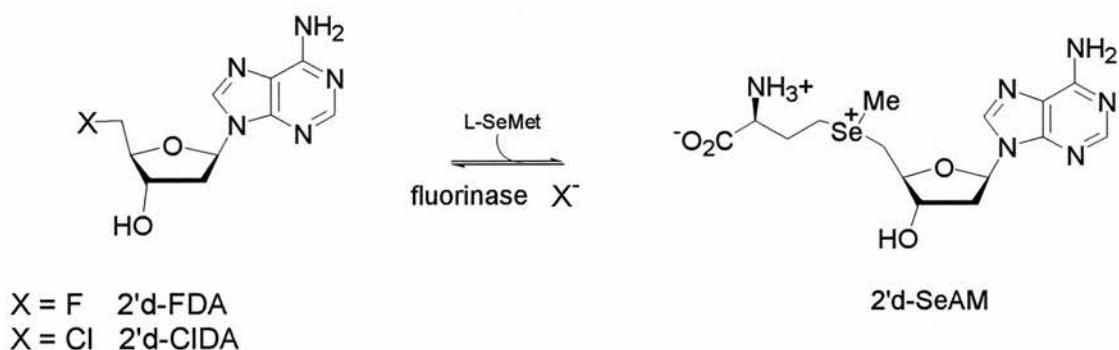


Figure 2.2 - HPLC profile monitoring the FDAS mediated transhalogenation reaction of 5'-CIDA (yellow) to 5'-FDA (green) with time and using L-SeMet as a co-catalyst. L-Se-SAM is observed as a minor product.

2.2.2 Substrate specificity

Investigations were carried out to explore substrate specificity by using 2'-deoxy and 3'-deoxy adenosine analogues. In a transhalogenation reaction (as illustrated in Scheme 2.3), the 2'-deoxy substitute appeared to be a substrate for FDAS (Figure 2.3). The rate was recorded as being 10 times less than that observed for 5'-FDA (Table 2.1) (Cobb *et al.*, 2006).



Scheme 2.3 - The FDAS can catalyse the conversion of 2'-deoxy-FDA and 2'-deoxy-CIDA substrates with L-SeMet to generate 2'-deoxy-Se-SAM

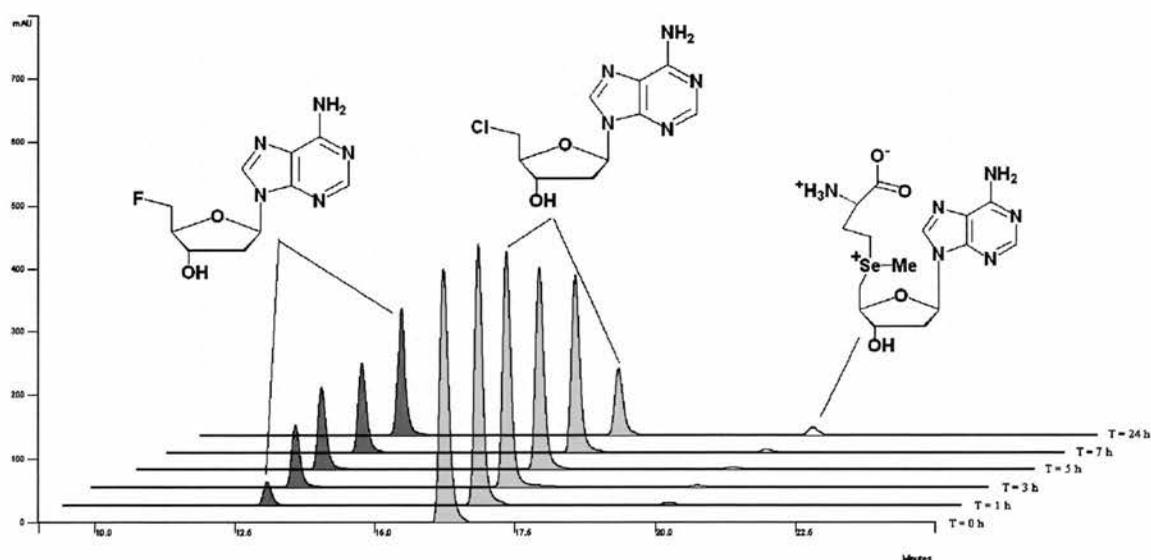


Figure 2.3 - HPLC profile of the transhalogenation from 2'-deoxy-CIDA to 2'-deoxy-FDA via 2'-deoxy-Se-SAM, catalysed by FDAS.

Incubation of both fluoro and chloro 3'-deoxy analogues with L-Met (with FDAS) did not result in the production of 3'-deoxy SAM (Shown in Table 2.1), suggesting that the 3'-hydroxyl of the ribose ring is essential for enzyme catalysis.

Substrate	R(5')	R(3')	R(2')	Specificity (%)
5'-FDA	F	OH	OH	100
2'-deoxy-5'-FDA	F	OH	H	12
3'-deoxy-5'-FDA	F	H	OH	0
5'-CIDA	Cl	OH	OH	740
2'-deoxy-5'-CIDA	Cl	OH	H	52
2'-deoxy-5'-CIDA	Cl	H	OH	0
adenosine	OH	OH	OH	0

Table 2.1 – Substrate specificity of 'reverse' reaction by FDAS.

Adenosine was also assayed by the same method and was shown not to result in the production of SAM (Table 2.1). FDAS also shows no activity towards the 5'-bromo and 5'-iodo deoxyadenosine derivatives in these experiments. FDAS therefore seems limited to Cl and F.

2.3 EXPERIMENTAL AIMS

The purpose of this project is to obtain insights into a novel naturally occurring fluorination enzyme through structural and thermodynamic studies. The enzyme has been shown also to be capable of a novel mechanism of chlorination (as previously discussed). FDAS has also been shown to catalyse the transformation of a repertoire of substrates. We aimed to describe the mechanism of chlorination by FDAS by investigation of the structure. This will be achieved by obtaining the co-crystal structure of FDAS with 5'-CIDA in the active site. We also aimed experimentally, to identify the chloride (fluoride) ion binding site by co-crystallization. This will provide a valuable insight into the reaction mechanism of FDAS. A structural investigation into substrate specificity will be carried out to reveal the reaction mechanism and account for disparity in reactivity. Finally, the thermodynamics of ligand binding by FDAS are investigated through isothermal titration calorimetry (ITC).

2.4 MATERIALS AND METHODS

2.4.1 Cloning FIA into pEHISTEV vector

To permit purification of the recombinant FDAS by automated methods the FDAS gene was amplified and re-cloned from the pET28a (+) vector, used to express the protein for the previous crystallization, into a more efficient pEHISTEV vector (Naismith and Liu, 2002) using 5' end primer containing a *NcoI* site and a 3' end primer containing *BamHI* site. The pEHISTEV construct was employed in all subsequent experiments since it had higher expression, the 6xHis-tag is readily cleavable and purification can largely be carried out automatically. The primers and plasmid map are shown Figure 2.4.

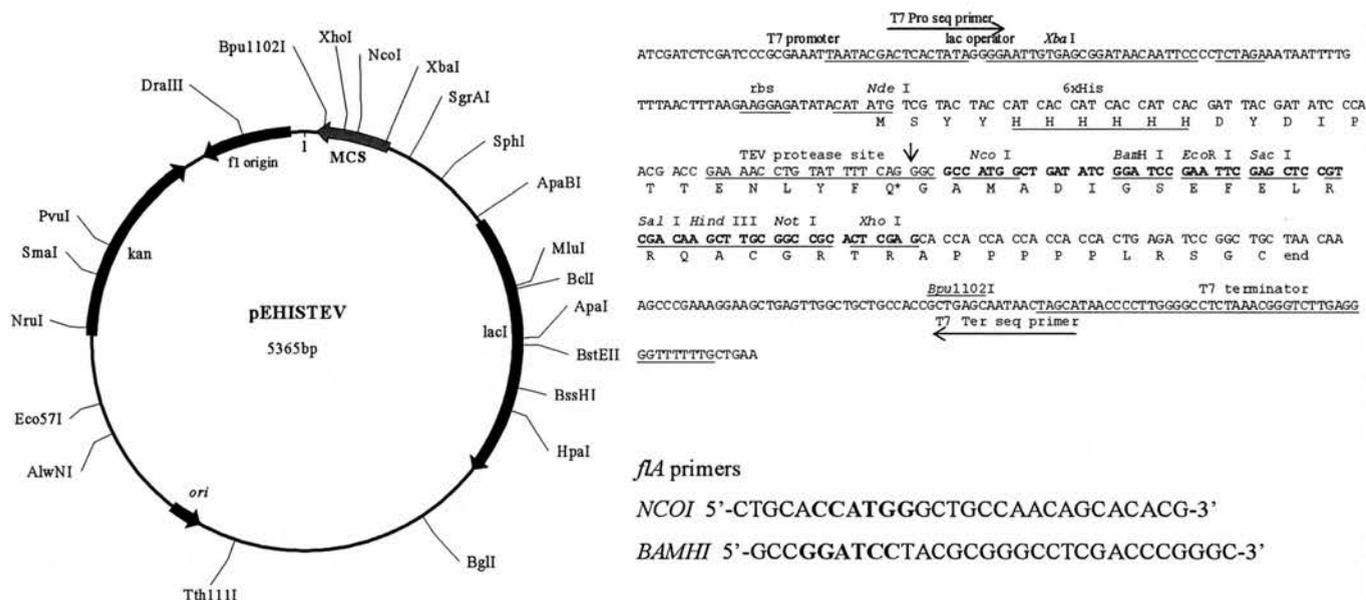


Figure 2.4 – Primers for *fIA* and pEHISTEV plasmid map.

The pEHISTEV-*fIA* construct plasmid DNA was amplified by use of RAPIDTRANS™ TAM1 chemically competent cells (Active Motif), purification of the plasmid DNA was from overnight cultures (10 mL) as directed in the protocol of the

Qiagen miniprep kit. The DNA was eluted into 30 μL $d\text{H}_2\text{O}$. Purity was determined by calculation of absorbance ratio at $\lambda = 260$ to $\lambda = 280$ nm. The concentration was determined from the absorbance of $\lambda = 260$ nm. The construct was stored at 253 K until required. DNA sequencing was carried out by The Sequencing Service (University of Dundee – www.dnaseq.ac.uk) using T7 forward and T7 reverse primers. This was carried out on an ABI 3730 capillary DNA sequencer.

2.4.2 Overexpression of recombinant 6HIS-TEV-FDAS

Plasmid DNA (pEHISTEV-*flA*) was transformed into BL21 (DE3) (Novagen) as described above (Naismith and Liu, 2002). An aliquot (1 mL) of overnight culture was used to inoculate LB (*Luria Bertani*) (500 mL), treated with kanamycin ($100 \mu\text{g mL}^{-1}$), in a baffled flask (2 L) and grown (310 K, 200 rpm) to an $\text{OD}_{600} = 0.6$. Protein expression was induced using IPTG (1 mM) with constant temperature and agitation. The cells were harvested after 4 h by centrifugation (8000 rpm, 20 min, 277 K, Beckman Avanti J20-XP, JL8.100 rotor). The cells were resuspended in PBS to remove any excess growth media, and re-spun (8000 rpm, 20 min, 277 K, Beckman Avanti J20-XP, JL8.100 rotor). The cells were stored at 253 K until required.

2.4.3 Typical purification of recombinant FDAS

The cells from four 500 mL cultures were resuspended in lysis buffer [25 mM Tris-HCl pH 7.75, 150 mM NaCl, 30 mM imidazole, 20 μM lysozyme and $20 \mu\text{g mL}^{-1}$ DNase I] and stirred for 30 min at 298 K. The cells were disrupted by Constant Systems cell disruptor (2 cycles, 20 KPSI, 298 K). The disrupted product was centrifuged (15000

rpm, 20 min, 277 K, Beckman Avanti J20-XP, JA 25.50 rotor). The supernatant was collected and passed through a 0.22 μm syringe filter membrane (Millipore) prior to application to a (15 mL) Ni^+ charged, Ni-NTA metal chelating column (Amersham Biosciences). The column was washed using 15 CV of wash buffer [25 mM Tris-HCl pH 7.75, 150 mM NaCl and 30 mM imidazole]. The protein was eluted using a single step of elution buffer [25 mM Tris-HCl pH 7.75, 150 mM NaCl and 250 mM imidazole]. The fractions containing the pure target protein (determined by SDS-PAGE) were, pooled and concentrated to 5 mg mL⁻¹ (determined by Bradford assay). The protein was diluted 10 fold with cleavage buffer [25 mM Tris-HCl pH 7.75 and 150 mM NaCl], prior to addition of TEV protease and incubated overnight (298 K). Progress of cleavage was checked by SDS gel electrophoresis, on completion the product mixture was passed through a 0.22 μm syringe filter membrane (Millipore) and a second Ni-NTA column (to remove any non-cleaved protein, tag and protease). The flow through was concentrated to 5 mg mL⁻¹ (estimated by Bradford assay) before further purification by size exclusion gel filtration. The protein integrity and identity were confirmed by mass spectrometry. The purity of the sample was determined by Coomassie™ Blue stained SDS-PAGE gel. In total 80 mg of protein is purified 2 L cell culture.

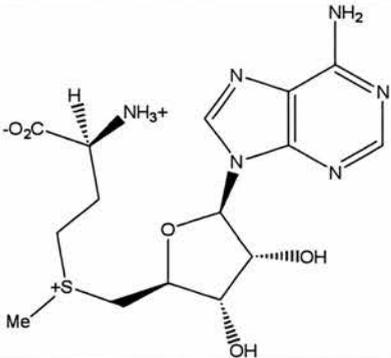
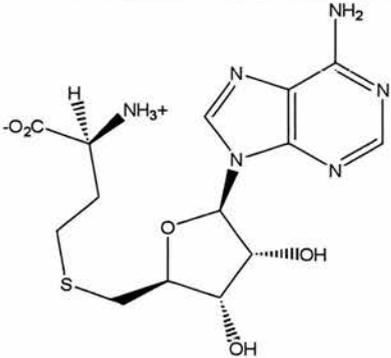
2.4.4 Removal of adenosine from FDAS

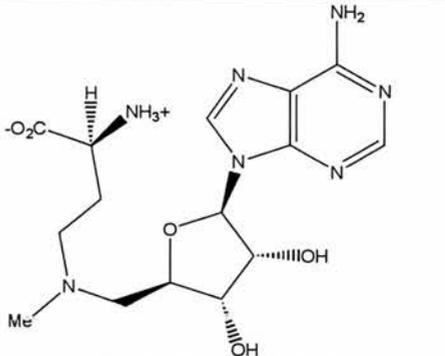
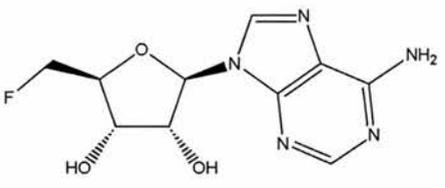
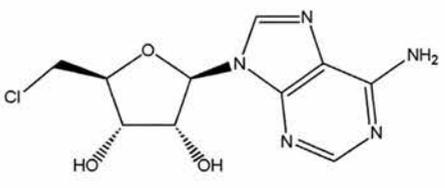
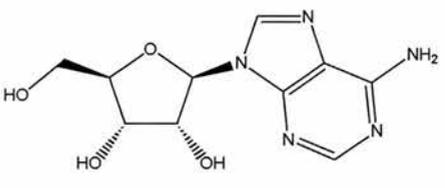
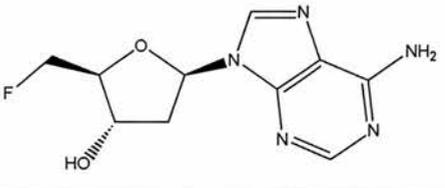
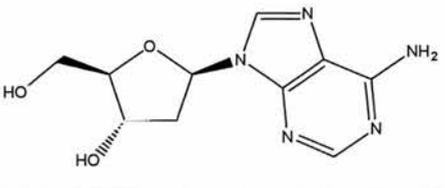
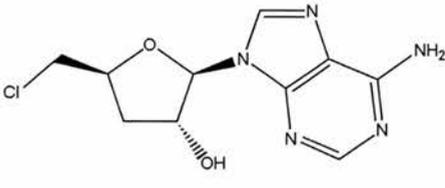
A supplementary step (after adenosine contamination was identified) was necessary to enhance the purification of the FDAS. This was achieved by incubation (298 K, with stirring) of the FDAS (10 μg per 100 mg FDAS) with 'Type X' adenosine deaminase (Sigma Chem. Co. Ltd. – *purified from bovine spleen*) prior to size exclusion

chromatography to remove adenosine contamination. The final purity of the sample was determined by HPLC (as described below).

2.4.5 Inhibitors, substrates and analogues

Substrate analogues and inhibitors were provided by our collaborators. The following compounds were available for use in both crystallization trials and ITC and are detailed in Table 2.2.

Name (Abbreviation) Source	Structure
<p>S-adenosyl -L- methionine (SAM) Sigma Chem. Co. Ltd., UK</p>	
<p>S-adenosyl -L- homocysteine (SAH) Sigma Chem. Co. Ltd., UK</p>	

<p>AZA-ADO-MET (SA8) (Thompson et al., 1999)</p>	
<p>5'-fluoro-5'-deoxyadenosine (5'-FDA) (Schaffrath et al., 2002)</p>	
<p>5'-chloro-5'-deoxyadenosine (5'-CIDA) O'Hagan</p>	
<p>Adenosine (Ad) Sigma Chem. Co. Ltd., UK</p>	
<p>5'-fluoro-2',5'-dideoxyadenosine (2dFDA) Sigma Chem. Co. Ltd., UK</p>	
<p>2'-deoxyadenosine (2DAd) Sigma Chem. Co. Ltd., UK</p>	
<p>5'-chloro-3',5'-dideoxyadenosine (3dCIDA) O'Hagan</p>	

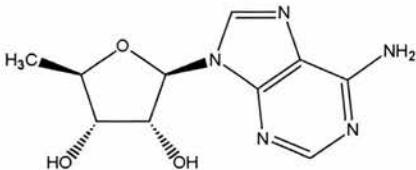
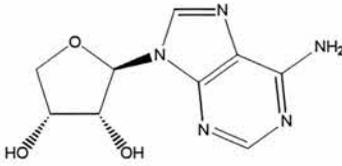
5'-deoxyadenosine (MeMet) O'Hagan	
β -D-erythrofuranosyl-adenosine (CC5) O'Hagan	

Table 2.2 – Substrates, inhibitors and analogues.

2.4.6 Crystallization

2.4.6.1 Co-crystallization

Optimal crystallization conditions were previously identified and reported as; 22 % PEG 1000, 0.1 M phosphate - citrate pH 4.2 and 0.2 M Li_2SO_4 (Dong et al., 2003). A general optimisation grid screen for co-complexes was prepared from these conditions, as shown below.

		<i>0.1 M phosphate - citrate</i>				
		4.0	4.1	4.2	4.3	4.4
PEG 1000	18 %					
	20 %					
	22 %					
	24 %					
	26 %					
	28 %					
	30 %					
	32 %					

Table 2.3 – FDAS grid screen for crystal growth, where optimization of % w/v PEG 1000 is varied against pH of 0.1 M phosphate - citrate buffer.

FDAS at a concentration of 4 mg mL^{-1} was incubated (4 h, 298 K) with 20 mM substrate (unless otherwise stated). The protein was crystallized by vapour diffusion method using 2 + 2 μL drops.

2.4.6.2 Sparse matrix screening of apo-enzyme

As the apo-protein did not to crystallize in the optimized screen (Table 2.3), sparse matrix screening was subsequently carried out on the native apo-enzyme and an active site mutant (Asn-215-Arg) at several concentrations (1, 3, 5, 7, 9 mg mL^{-1}) in several commercially available screens (as detailed in appendix A.I) including Crystal screen 1 and 2 (Hampton), Wizard™ I and II (Emerald Biosciences), the Pegs™ (Nextal) and Premier (In-house). The screening was performed at the SSPF (University of St. Andrews) using 0.1 μL of protein and 0.1 μL precipitant using a nano-drop crystallization robot (Cartesian HONEYBEE™) as part of the Hamilton-Thermo Rhombix™ system. Crystals which were obtained from the screening of the Asn-215-Arg mutant were optimised accordingly from the most promising hits using both native and mutant protein. Diffracting crystals of the native protein precipitated using 2 μL of protein solution (4 mg mL^{-1}) and 2 μL precipitant by vapour-diffusion method against 200 mM sodium thiocyanate, 20 % w/v PEG 3350.

2.4.7 Data collection and processing, structure solution, refinement and validation

2.4.7.1 Data collection of C222₁ crystals

Crystals grown from grid screens (Table 2.3) were cryo-protected by increasing PEG 1000 to 40 % or by adding glycerol (20 %) or PEG 200 (20 %). PEG 200 gave more reliable cryo-protection than glycerol (and was easier to prepare than PEG 1000), as crystals survived for longer before flash freezing. Once a single crystal had been selected and soaked in cryo-protectant it was mounted on a cryo-loop and flash cooled to 100 K in IN₂ (Garman, 1999).

Screening of crystals relied on a rotating anode X-ray generator (Rigaku), which produced Cu *K*α radiation ($\lambda = 1.5418 \text{ \AA}$) with Osmic mirrors. The detector was a Raxis IV⁺⁺ image plate (Rigaku). Data sets were collected as detailed in the discussion. Data collection was problematic even with 0.5° oscillations, when crystals were oriented such that diffraction of the long cell edge was perpendicular to the rotation axis. This resulted in numerous overlapping spots. Rather than collecting smaller oscillation angles; attention was taken when mounting the crystals such that the orientation of crystals ensured that the long cell edge (*c*) was close to parallel with respect to the rotation axis. The orientation of the long cell edge (~190 Å) could be predicted from the crystals morphology (Figure 2.5)

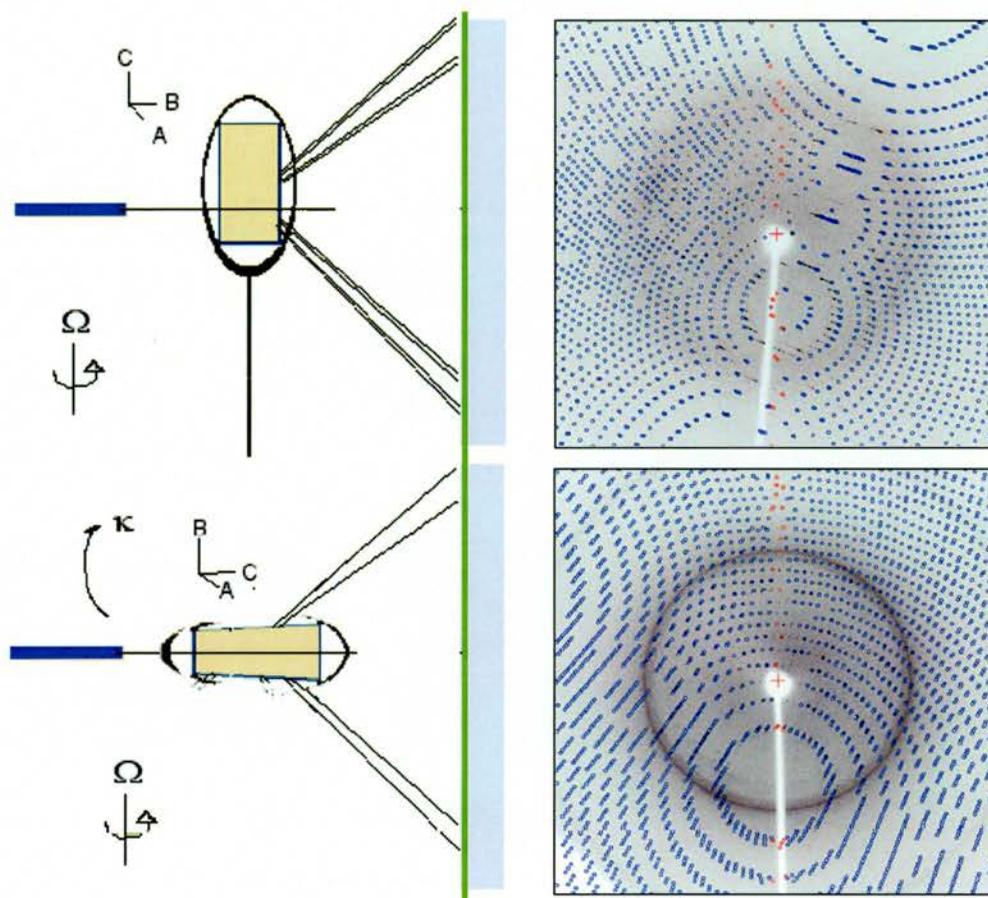


Figure 2.5 – Orientation of $C222_1$ FDAS crystals for optimal data collection. *TOP*: The unfavourable orientation, with long cell edge (c) perpendicular to rotation axis (Ω). *BOTTOM*: Favourable orientation where long cell edge is parallel to rotation axis.

2.3.7.2 Data processing

Data were indexed and integrated using MOSFLM (Leslie, 1999; Leslie, 1992) and merged using SCALA (Evans, 1997). The reflections were often indexed incorrectly in MOSFLM. This occurred from the pseudo-hexagonal nature of the unit cell where reflections can be indexed as $C222_1$ in three related transformations, two of which are incorrect (illustrated in Figure 2.6).

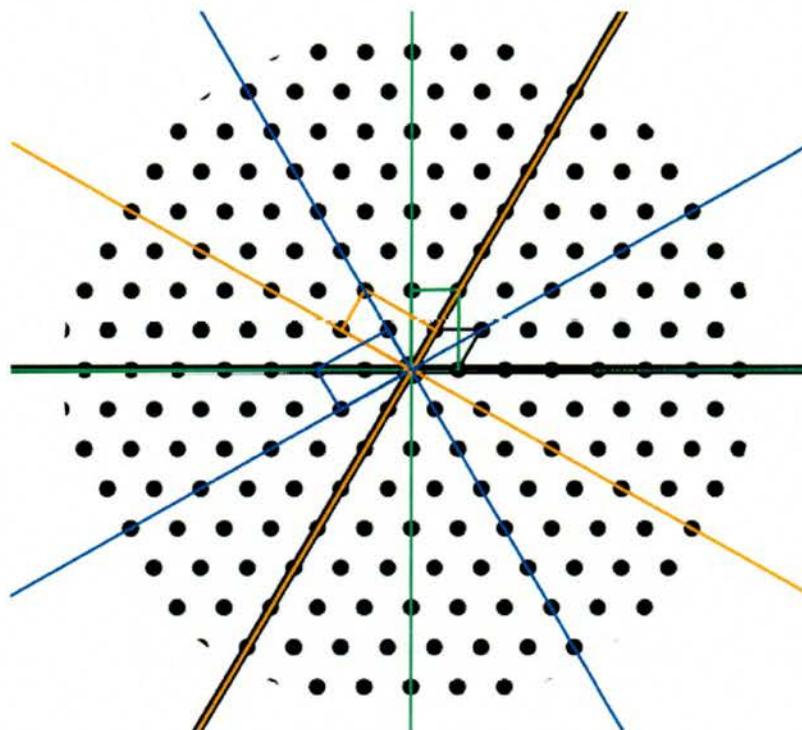


Figure 2.6 – Possible $C222_1$ unit cells in a pseudo hexagonal lattice, Hexagonal axes (black); three alternative C-centered orthorhombic lattices (orange, blue and green) (Evans, 2004).

The program 'POINTLESS' (Evans, 2004) was used to predict the correct transformation and data could be appropriately reindexed. Data may be transformed by the reindex operators highlighted. A typical output from POINTLESS is shown in Table 2.4.

Netmt	Z-cc	CC	Z-rms	N	rmsD	Rmerge	Symmetry&operator	(in Lattice Cell)
> 1	10.22	0.70	2.89	21018	-10.23	0.225 ***	2-fold l	(0 0 1) {-h,-k,+l}
2	-0.52	-0.03	1.05	26383	-19.52	0.570	2-fold	(1-1 0) {-k,-h,-l}
3	0.11	0.02	0.78	13746	-20.89	0.632	2-fold	(2-1 0) {+h,-h-k,-l}
> 4	11.37	0.78	3.66	5303	-6.32	0.164 ***	2-fold h	(1 0 0) {+h+k,-k,-l}
5	-0.83	-0.05	0.44	24881	-22.65	0.619	2-fold	(1 1 0) {+k,+h,-l}
6	0.22	0.02	0.54	20745	-22.10	0.646	2-fold k	(0 1 0) {-h,+h+k,-l}
> 7	11.60	0.79	4.02	16554	-4.50	0.151 ***	2-fold	(-1 2 0) {-h-k,+k,-l}
8	-0.03	0.01	0.18	23384	-23.96	0.626	3-fold l	(0 0 1) {-h-k,+h,+l}
9	0.70	0.06	0.30	35908	-23.33	0.644	6-fold l	(0 0 1) {-k,+h+k,+l}

Laue Group	NetZoc	Zoc+	Zoc-	CC	Rmerge	ReindexOperator
1 C m m m ***	10.94	10.97	0.03	0.75	0.19	[3/2*h+1/2*k,1/2*h-1/2*k,-1]
6 C m m m	2.48	4.68	2.21	0.33	0.47	[1/2*h+1/2*k,3/2*h-1/2*k,-1]
9 C m m m	-0.71	2.42	3.13	0.17	0.48	[h,k,l]
5 P6/m m m	2.86	2.86	0.00	0.20	0.51	[1/2*h+1/2*k,1/2*h-1/2*k,-1]

Table 2.4 – Output from POINTLESS: scoring the possible indexing schemes, showing that for this case the data is indeed C222 and not hexagonal. The data lines highlighted in blue refer to the wrong solutions, and the red line is the correct solution identified by the high correlation coefficient (CC).

2.4.7.3 Structure solution

All data sets were indexed and merged in the space group C222₁, with exception of the apo-enzyme, which indexed and merged in P2₁2₁2₁. Given correct indexing the structures were solved by molecular replacement using MOLREP (Vagin and Teplyakov, 2000) with protein coordinates (trimer) derived from existing PDB entries (1rqr & 1rqp) as the search model.

2.4.7.4 Structure refinement

Structures were refined using Refmac5 (Murshudov et al., 1997) with ‘translation, libration, screw rotation’ (TLS) (Potterton et al., 2003; Winn et al., 2003). Water molecules were added to the model automatically using ARP/wARP (Perrakis et al., 1999). The ligands were modelled into the unbiased Fo-Fc when the density was unambiguous. The final models were validated and Ramachandran plots were prepared

using PROCHECK (CCP4i suite) (Potterton et al., 2002). The details of refinement and validation are discussed in the results for each structure. Crystallographic figures were produced using PyMOL™ (Version 0.99 – DeLano Scientific) unless otherwise stated. Ligand libraries were prepared using PRODRG (Schuttelkopf and van Aalten, 2004).

2.4.8 High Performance Liquid Chromatography (HPLC)

HPLC of crystallized and dissolved protein was performed to identify substrates, inhibitors and analogues. Protein samples (4 mg mL^{-1}) were either denatured by heat treating to 363 K or by addition of HCl to denature the protein. Protein was removed by centrifugation and 20 μL of supernatant (20 μL) was injected onto a C-18 (Hypersil™ 5 μm) column attached to a Varian™ 325 UV-Vis detector, 400 autosampler through a C-18 column pre-equilibrated with mobile phase A [50 mM KH_2PO_4 , acetonitrile 95:5 v/v]. Sample was eluted by a smooth gradient from 100 % mobile phase A to 100% mobile phase B (over 30 CV) [50 mM KH_2PO_4 , acetonitrile 80:20 v/v]. The eluant was monitored at 260 nm. The column was calibrated with a known standard (Table 2.5).

Compounds	Retention Time (Min.)
inosine	5.3
adenosine	7.4
2'-deoxyadenosine	7.8
5'-fluoro-5'-deoxyadenosine	9.9
5'-fluoro-2',5'-dideoxyadenosine	10.7
5'-chloro-5'-deoxyadenosine	12.7
5'-chloro-2',5'-dideoxyadenosine	13.8
S-adenosyl -L- homocysteine	14.8
S-adenosyl -L- methionine	15.0
5'-bromo-5'-deoxyadenosine	15.2

Table 2.5 – Retention times of selected compounds.

2.4.9 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy (Kelly et al., 2005; Kelly and Price, 2000) was performed to verify protein folding. The CD scans were carried out at the University of St Andrews using 1 mg mL⁻¹ protein in water on a JASCO J-810 CD spectropolarimeter with a 1 mm cell, and processed using JASCO Spectra Manager™.

2.4.10 Isothermal Calorimetry (ITC)

ITC was carried out in both The University of St Andrews and The University of Glasgow (MicroCal™ VP-ITC). Initial buffer to protein and ligand to buffer titrations were performed as a control. Ligands were titrated against FDAS at known concentrations. Ligand and protein were both equilibrated in the same buffer [25 mM Tris-HCl, pH 7.75 or 25 mM Tris-HCl pH 7.8] to reduce noise from the heat of dilution. Experiments were carried out at 298 K. The final FDAS concentration ($\epsilon_{\text{FDAS}} = 35995 \text{ M cm}^{-1}$) in solution was determined by UV 280 nm. The concentration of ligands were determined ($A_{260} = \epsilon \cdot c \cdot l$), (using the appropriate ϵ , for example $\epsilon_{\text{adenosine}} = 14200 \text{ M cm}^{-1}$). The raw data were analysed using Origin™ ITC (Version 5.1 – Microcal™) software.

2.4.11 Site Directed Mutagenesis

The QuikChange™ site directed mutagenesis protocol (Stratagene™) was used to introduce point mutations in the *flA* gene (*Streptomyces cattleya*). The pEHISTEV-*flA* construct (section 2.4.1) was used as the DNA template and generates large quantities of mutated DNA using the Polymerase Chain Reaction (PCR). The non-mutated (parental) DNA is methylated by *E. coli* so it can be digested by the addition *DpnI* endonuclease

(McClelland and Nelson, 1992). The non-methylated DNA as a product of PCR is therefore not targeted by the enzyme for degradation. The following reaction mix was used for PCR: 20 ng *Template*, 5 μ L 10x *Pfu* DNA polymerase buffer, 4 μ M *d*NTP, 100 ng 5' primer (sense), 100 ng 3' primer (anti-sense). The mixture was made up to 49 μ L with water and 1 μ L of *Pfu* DNA polymerase was added. The PCR reaction was carried out (over 12 cycles) in a GeneAmp™ PCR thermocycler (Applied Biosystems).

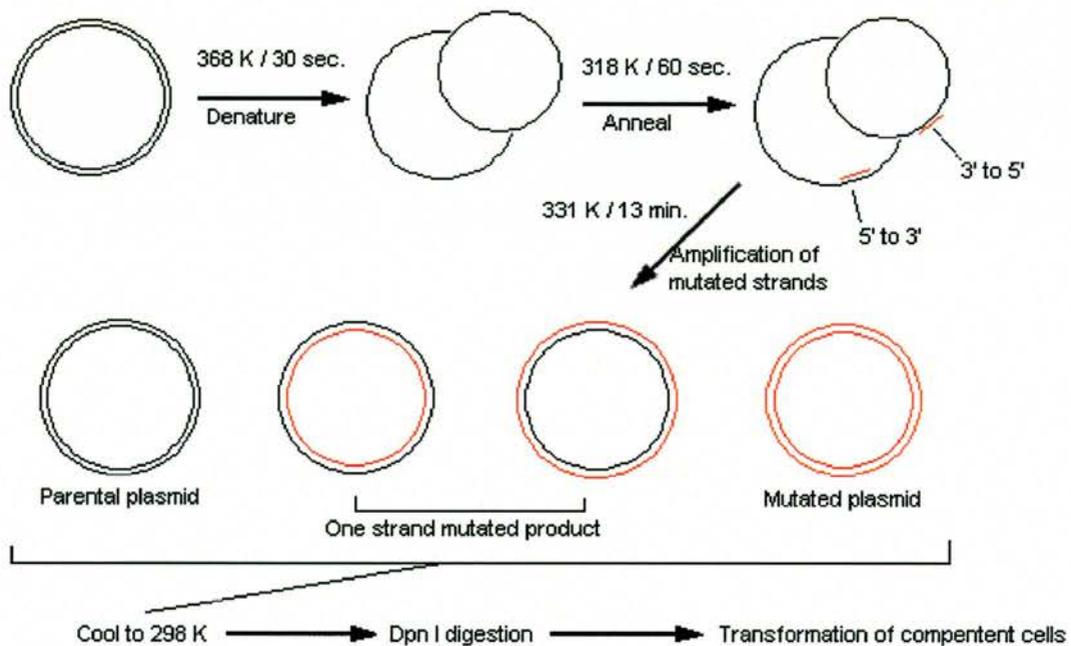


Figure 2.7 – PCR site directed mutagenesis (Stratagene, 2004)

After 12 cycles PCR was confirmed by running a sample of the reaction on a 1 % agarose gel. Dpn I was added to the reaction to remove parental DNA and another sample was run on an agarose gel, after digestion. The mutant construct plasmid DNA was directly amplified by transformation of RapidTrans™ TAM1 chemically competent *E.coli* cells (Active Motif), plated on L-agar (containing 30 μ g mL⁻¹ kanamycin). A

colony was chosen and purification of the plasmid DNA were from overnight cultures (10 mL) as directed in the protocol of the Qiagen miniprep kit. The DNA was eluted into 30 μL $d\text{H}_2\text{O}$. Purity was determined by calculation of absorbance ratio at $\lambda = 280 / \lambda = 260$ nm. The concentration was determined from the absorbance of $\lambda = 260$ nm. The construct was stored at 253 K until required. DNA sequencing was carried out to confirm successful mutagenesis.

2.5 RESULTS AND DISCUSSION

2.5.1 Initial co-crystallization of 5'-chloro-5'-deoxyadenosine with FDAS

A solution of FDAS were incubated with 5'-chloro-5'-deoxyadenosine (5'-CIDA) and set up in crystal trials (consistent with the procedure outlined in the methods), and crystals, of diffraction quality, were obtained after 3 days to 1 week. These crystals were found in similar conditions and appeared to be consistent with the morphology of crystals obtained in previous studies (pictured in Figure 2.8) (Dong et al., 2003).

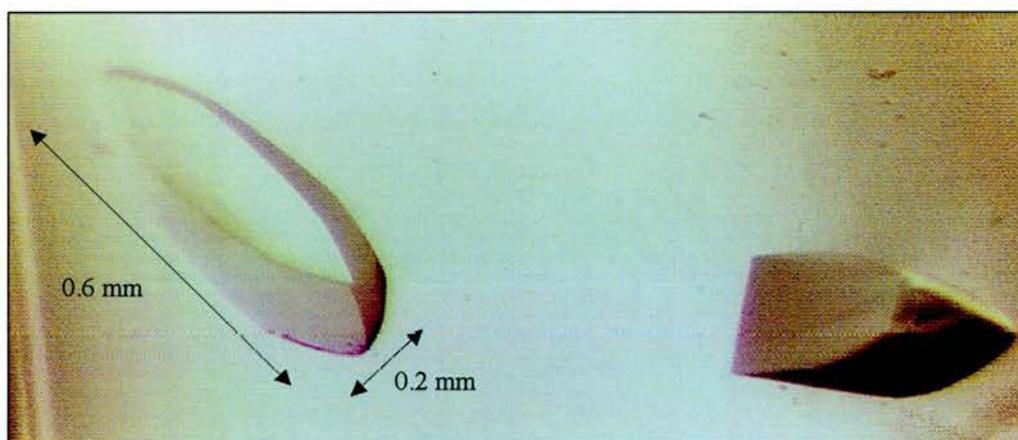


Figure 2.8 – Typical crystals obtained by co-crystallization of substrate and FDAS.

X-ray diffraction data were measured (consistent with the previously described method) on these crystals to 2.3 Å. The data was scaled and processed and the structure was determined by molecular replacement using the protein structure from 1rqp as the search model. Initial refinement omitting ligand from the model, showed density for 5'-CIDA bound in the active sites of the protein. This is shown in Figure 2.9

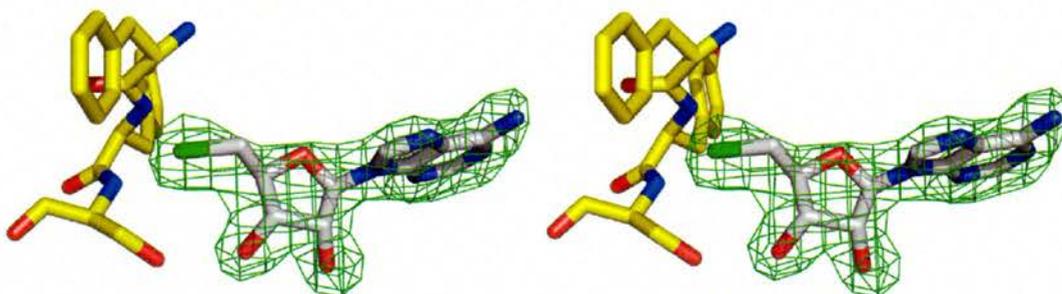


Figure 2.9 – Initial density maps of 5'-CIDA modelled into the unbiased F_o - F_c (green mesh and $+3.0 \sigma$) electron density map, after molecular replacement, in the active site of FDAS (C = white and yellow; O = red; N = blue; Cl = green). Ser-158, Tyr-157 and Phe-156.

As refinement of ligand and protein was continued it became apparent that there were problems. Negative F_o - F_c density (-3.0σ) was seen at the chlorine atom of the 5'-CIDA molecule (Figure 2.10). This was confirmed by analysis of the B-factors; the chlorine atom was $\sim 55 \text{ \AA}^2$, C5' was $\sim 38 \text{ \AA}^2$, C4' was $\sim 30 \text{ \AA}^2$. The average B-factor for the model was $\sim 25 \text{ \AA}^2$.

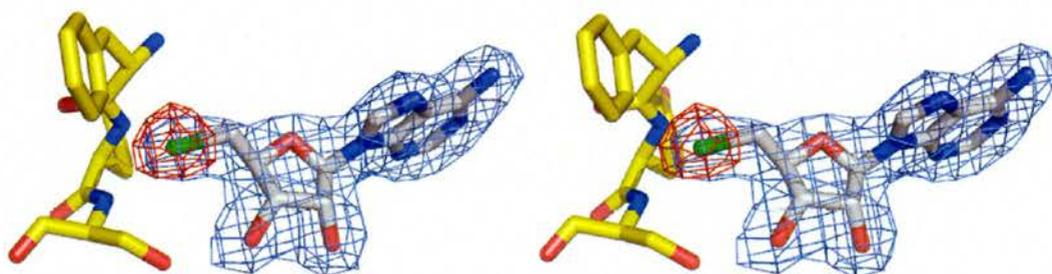


Figure 2.10 – Density after refinement of 5'-CIDA in the active site of FDAS in initial co-crystallization experiments. F_o - F_c electron density map (green mesh and $+3.0 \sigma$, red mesh and -3.0σ), in the active site of FDAS (C = white and yellow; O = red; N = blue; Cl = green). The $2F_o$ - F_c (steel blue mesh and 1.0σ) density map is also shown.

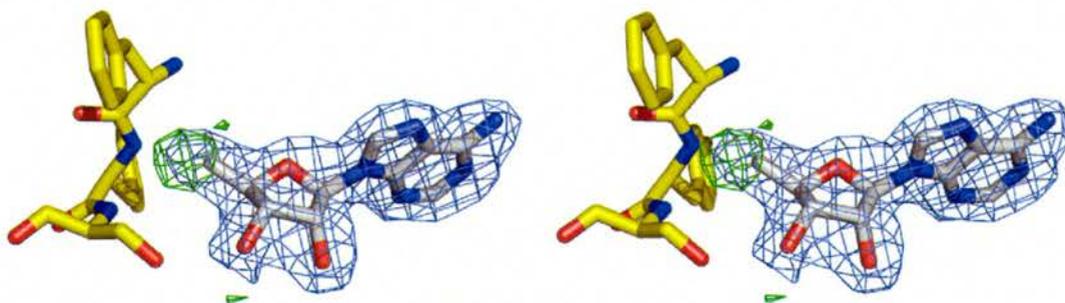


Figure 2.11 – Density observed following refinement of 5'-deoxyadenosine into the active site of FDAS in the initial experiment. F_o-F_c (green mesh and $+3.0 \sigma$, red mesh -3.0σ) electron density map, in the active site of FDAS (C = white and yellow; O = red; N = blue). The $2F_o-F_c$ (steel blue mesh and 1.0σ) density map is also shown.

In order to see if other conformations were present the ligand was changed to 5'-deoxyadenosine and refinement restored (Figure 2.11). The adenine and ribose ring moieties were exchanged, however established that the only density was at the 5' position in a single conformation and that the additional density was 4σ where as one would expect the density to be higher ($\sim 7\sigma$) for a fully ordered chlorine. The density is consistent with a hydroxyl, a fluoro, a methyl or disordered chlorine. A second data set using a different crystal grown in a different drop confirmed that our observations were not an artifact of the initial data collection. In addition, one 3'-deoxyadenosine co-crystal structure was solved. The observed density was similar to the above structure.

In summary it was clear that the density of the ligand observed in the refinement of the model was not consistent with the density expected for 5'-CIDA. Since the compound was from a reliable source and the purity and identity had been confirmed by HPLC, MS and NMR, we suspected that there may be some degradation or reaction of 5'-CIDA during the crystallization process. Since fluoride and L-methionine were absent from the crystallization conditions, the transhalogenation reaction was discounted.

2.5.2 HPLC analysis of the co-crystals and identification adenosine contaminant in purified recombinant FDAS

Co-crystals grown from the same experiments were subjected to HPLC analysis. Crystals were removed from the drop (2 + 2 μL) in which they were grown using a standard loop and transferred to a second drop (2 μL) containing stabilization buffer (as used for crystallization). As we expect the ligand bound to the protein to remain bound while crystallised this was necessary to reduce the presence of ligand as used in the co-crystallization. After a few seconds the protein crystal was transferred into a drop of water where the crystal dissolved immediately. This was repeated so that the drop contained a relatively high concentration of protein (approx. 10 mg mL⁻¹). The drop was diluted ten fold and transferred to a thin walled PCR tube. The sample was heated and centrifugated (to remove the protein) and the supernatant was injected directly onto an HPLC (Figure 2.12).

This showed a peak consistent with the retention time of 5'-CIDA (12.7 min.) and a second peak at 7.4 min. The two peaks were collected from the HPLC and analysed by ESI-MS. This confirmed the identity of the 5'-CIDA and indicated that the second peak was a compound of $m/z = 268$ (M + H), i.e. adenosine. Authentic adenosine showed the same retention time. Purified recombinant protein was denatured and analysed by HPLC to show this same impurity. In addition native protein was shown not contain adenosine but did contain SAM throughout purification.

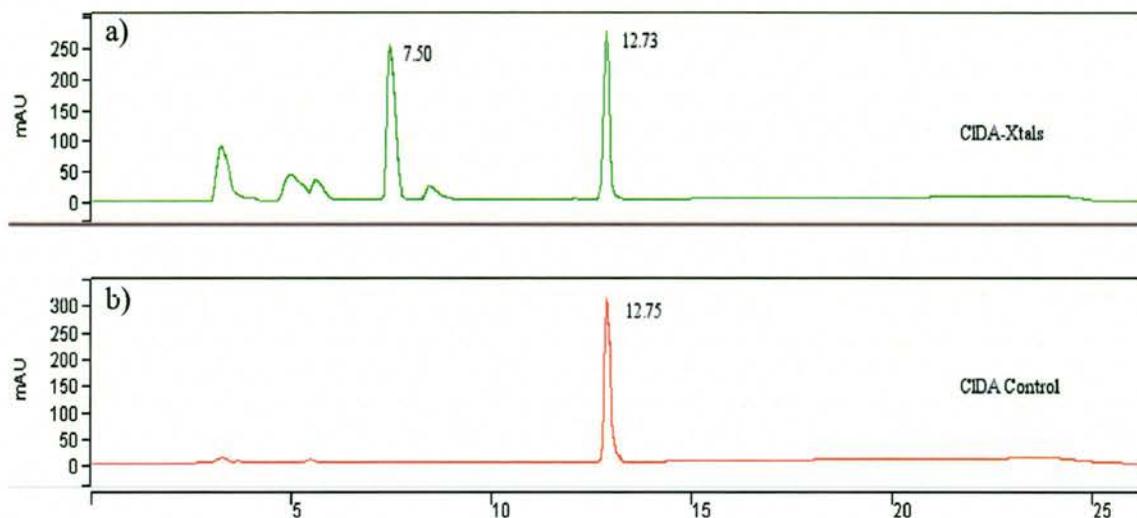


Figure 2.12 – Chromatograms of the HPLC analysis of a) 5'-CIDA co-crystals, b) 5'-CIDA control.

2.5.3 Removal of adenosine contamination

Before repeating co-crystallization experiments, adenosine must be removed from the protein. Several methods were unsuccessful, including attempts at displacement by bromide or by reaction of adenosine using an *E. coli* purine nucleoside phosphorylase (PNPase) enzyme (E.C. 2.4.2.1) (Sigma). Recombinant FDAS was incubated overnight with 2 M NaBr at 298 K. This was expected to displace adenosine by competing for the active site and could be removed by dialysis however this was not observed. An *E. coli* PNPase (0.1 mg per 10 mg FDAS) (Sigma) was incubated with FDAS at 298 K. Incubation of the PNPase destroyed adenosine as expected however also resulted in degradation of the protein suggesting the presence of a protease impurity.

To understand the importance of adenosine contamination an active site mutant (Asn-215-Arg) was prepared (as described in section 2.4.11) designed to inhibit substrate

binding as the arginine residue would point into the active site. The structure should be the same as the apo-enzyme. The (Asn-215-Arg) mutant was purified as before for the recombinant enzyme, gel filtration indicated that the protein was a hexamer. The protein integrity and identity were confirmed by mass spectrometry. The purity of the sample was determined by Coomassie™ Blue stained SDS-PAGE gel. A CD spectrum and a native gel (Tris-Glycine) indicated the mutant protein was not significantly different to the native protein (Figure 2.13). HPLC analysis confirmed no impurities of small molecules after purification. To understand if this affected protein folding attempts were made to obtain the crystal structure. The Asn-215-Arg protein did not crystallize in the previously reported conditions (Table 2.3) however hits were obtained after sparse matrix screening, the significance of which is discussed in Section 2.5.10.

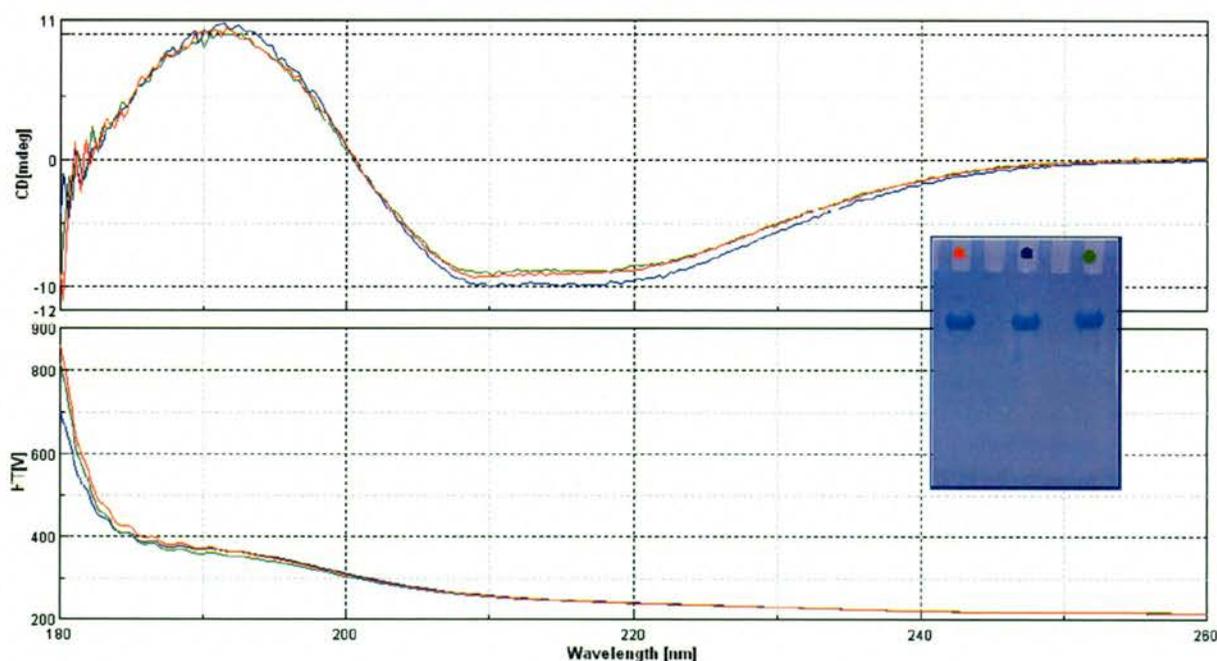


Figure 2.13 – CD spectra and native gel of FDAS; Native - untreated (blue), Native – after treatment with deaminase (green) and Asn-215-Arg (red).

Successful removal of adenosine was eventually achieved by addition of (10 μg per 100 mg FDAS) 'Type X' adenosine deaminase (E.C. 3.5.4.4) (Sigma – Purified from bovine spleen), which fully metabolised the adenosine to inosine. The protein was further purified using size-exclusion chromatography to remove the adenosine deaminase and small molecules. The protein was eluted into the buffer used for crystallization and concentrated as before (Section 2.5.1). It is noteworthy that the recombinant FDAS runs at the same elution volume by gel filtration suggesting that the protein remains as a hexamer, consistent with native gel (Tris-Glycine) electrophoresis.

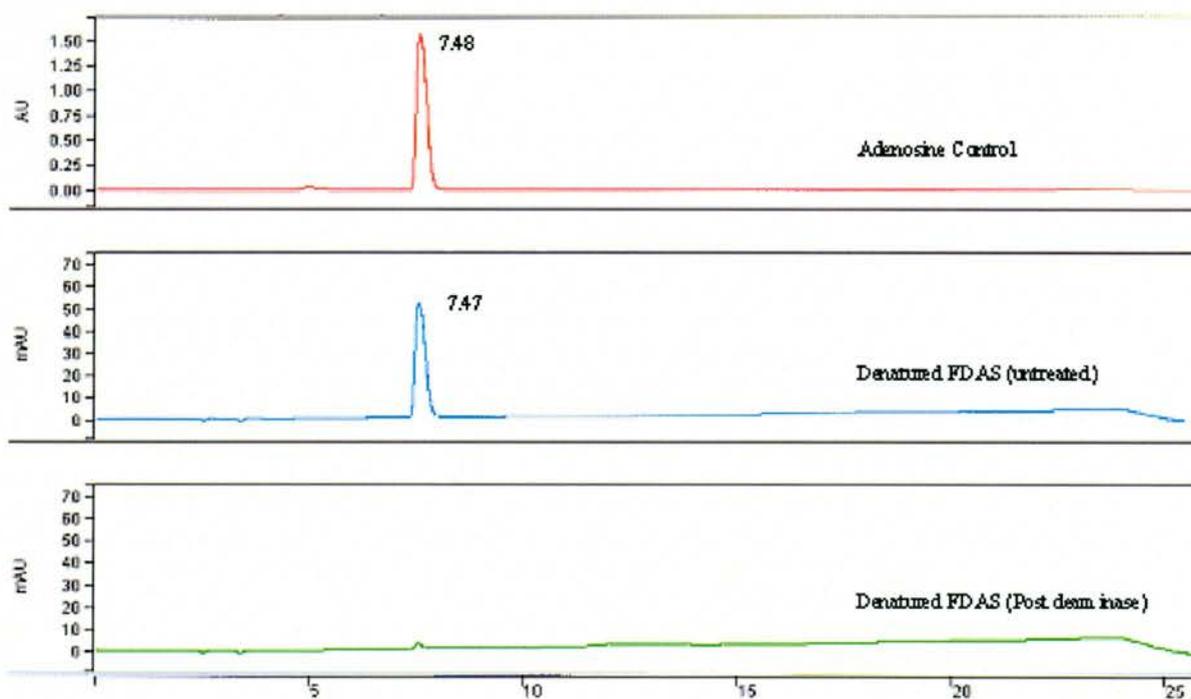


Figure 2.14 – a) Chromatograph showing synthetic adenosine; b) chromatograph of supernatant after recombinant FDAS is denatured; c) chromatograph of supernatant after recombinant FDAS is denatured from a sample treated with deaminase enzyme.

The reaction does not stop with the conversion to inosine as expected, however further metabolism is observed (from the chromatogram) which indicates there is breakdown to hypoxanthine and ribose. This probably results from the presence of a catalytic amount of an inosine specific nucleosidase.

It was determined by relative concentration (using $A_{260} = \epsilon \cdot c \cdot l$) that the level of adenosine contamination carried throughout the standard purification was $95 \pm 5\%$ prior to treatment with deaminase (suggesting a 1:1 ratio). The observation of any peak consistent with 5'-CIDA in the HPLC analysis is probably carried over from the crystallization condition.

With hindsight it was possible to refine adenosine into the active site of the model using the data obtained when attempting to co-crystallize 5'-CIDA.

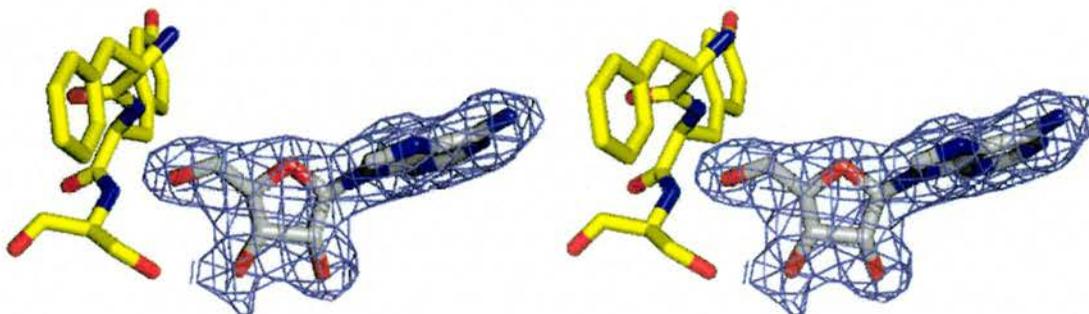


Figure 2.15 – Adenosine refined into the active site of FDAS. *Fo-Fc* (green mesh and $+3.0\sigma$, red mesh and -3.0σ) electron density map, in the active site of FDAS (C = white and yellow; O = red; N = blue). The *2Fo-Fc* (steel-blue mesh and 1.0σ) density map is also shown.

Superimposition of FDA and 'adenosine' models, show both ligands adopt identical positions in the active site. Interestingly the torsion angle ($1'C-2'C-3'C-4'C$)

across the ribose ring ($\sim 13^\circ$) and the 5' hydroxyl atoms (adenosine model) are essentially the same distance from protein interactions as the 5' fluoro group (1qr).

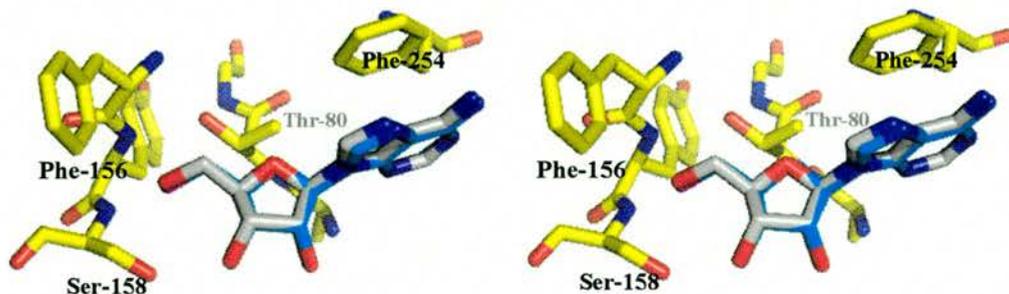


Figure 2.16 – Superimposition of adenosine and 5'-FDA (F= Purple)

Paradoxically recombinant FDAS (which has not been treated for adenosine contamination) does not crystallize and requires some addition of ligand. This suggests full ligand occupancy is essential for well ordered crystallization. It is noteworthy that FDAS expressed in *S. cattleya* is purified with SAM bound and does crystallize in this condition. It has been shown in related studies (McGlinchey, 2006) that adenosine is a substrate of the PNP. Interestingly this raises the possibility that the PNP (which is directly involved in the pathway) may also be involved in regulation of adenosine within the cell. A summary of the data collection and final refinement is shown in Table 2.6

Data collection	<i>045col1-5-CIDA</i>
Wavelength (Å)	1.54
Resolution (Highest Shell) (Å)	37 – 2.3 (2.4 – 2.3)
Space Group	C222 ₁
Cell constants (Å)	a = 75.8, b = 129.9, c = 184.6
Unique reflections	39904 (4906)
Average redundancy	3.4 (2.4)
I/σ	12.8 (5.0)
Completeness (%)	97.4 (83.0)
R _{merge} (%)	6.9 (20.3)
Refinement	<i>---Adenosine---</i>
R	17.9 (19.7)
R _{free}	23.1 (24.5)
Rmsd bond (Å)/angle(°)	0.01 / 1.24
Residues in Ramachandran core (%)	86.9
PDB accession code	----

Table 2.6 – Summary of data collection and refinement statistics

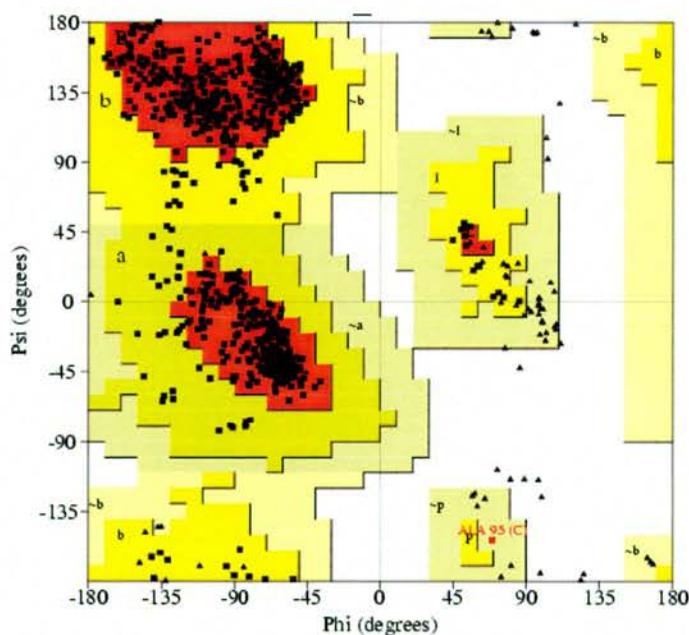


Figure 2.17 – Ramachandran plot of the FDAS model (with bound adenosine)

2.5.4 Co-crystallization and X-ray crystal structure of FDAS complexed with 5'-chloro-5'-deoxyadenosine (5'-CIDA)

Removal of the adenosine contamination allowed the successful solution of 5'-CIDA co-crystallized with FDAS. Crystals were obtained using the optimisation screen (described above) in 32 % PEG 1000, 0.1 M phosphate citrate pH4.2, 0.2 M Li₂SO₄. Data were collected to 2.0 Å at the ESRF (ID14-EH2), and processed as described in the methods. Phases were calculated using molecular replacement with the original FDAS structure, without the ligand. The unbiased Fo-Fc density suggested the presence of two conformers of 5'-CIDA, which were modelled into the density as appropriate. This is shown in Figure 2.18.

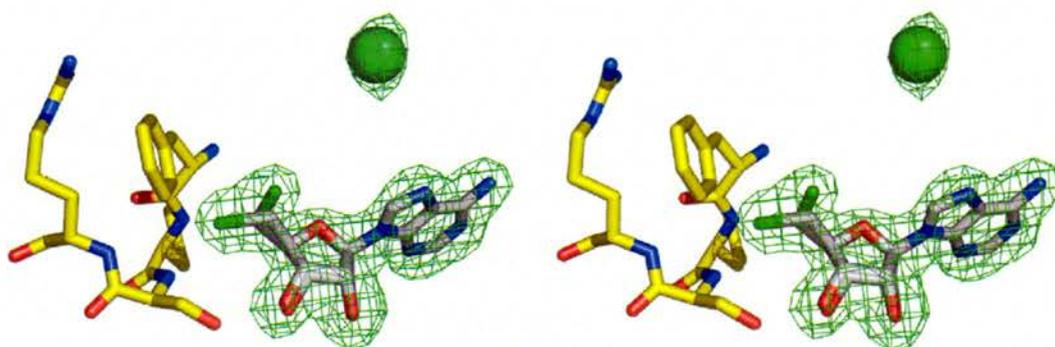


Figure 2.18 – The two conformations of 5'-chloro-5'-deoxyadenosine (5'-CIDA) modelled into the unbiased Fo-Fc (green mesh and +3.0 σ) electron density map, after molecular replacement, in the active site of FDAS (C = white and yellow; O = red; N = blue; Cl = green). A possible chloride ion binding site is also indicated (Green sphere).

The dual conformation of ligand and presence of chlorine atom was conserved in each active site. Final refinement of the occupancies showed that ~30 % of 5'-CIDA in

the crystal pointed towards Ser-158 and ~70 % pointed towards the methionine binding pocket. The final refined model shows the two conformers. The model indicates that the protein does not undergo any major structural modification in response to the ligand; rather the substrate adjusts its position in response to the larger atom at the C5' position. This suggests that the ligand binding site is rigid during binding. The altered position of the halogen does not, however, prevent the enzyme from catalyzing the reaction (as stated in the earlier section). The B-factors for all atoms in the ligand (5CD – 2c2w) are consistent with the B-factors in the remainder of the model.

As the chlorine is in part rotated out of the halogen binding site, this suggests that there is some degree of flexibility. This conformer will not be observed in the situation that L-Met is present. Unfortunately the effect on conformation of the ligand by L-Met could not be observed in this case since pre-incubation of methionine and 5'-CIDA produce SAM in active protein. Attempts to co-crystallize unreactive analogues of L-Met (such as homocysteine, and alanine) were unsuccessful.

The second conformer is coordinated in the same way that 5'-FDA (and adenosine) was shown to be coordinated in 1rqp. The hydrogen bonding interactions of 5'-CIDA to Ser-158 are the same as those observed for 5'-FDA, however the distances are longer. The difference in position was estimated at 1.3 Å. This may be attributed to the larger van der Waals (VDW) radius of the chlorine atom relative to the fluorine atom. The superimposition of 5'-CIDA (one conformer pointing towards Ser-158) and 5'-FDA in the active site of FDAS is shown in Figure 2.19.

Chlorine incorporation in the forward reaction is understandably less favourable than fluorine incorporation. Although the energy necessary for chloride desolvation is significantly lower than fluorine, entry into the active site is less likely due to the tight

binding pocket. This evidence goes some way to explaining FDAS does not act as a hydroxylase to produce adenosine. It does not appear that water is able to enter the ion binding site since the pocket is protected the hydrophobic Phe-156 and there are no residues in the vicinity capable of activating the OH.

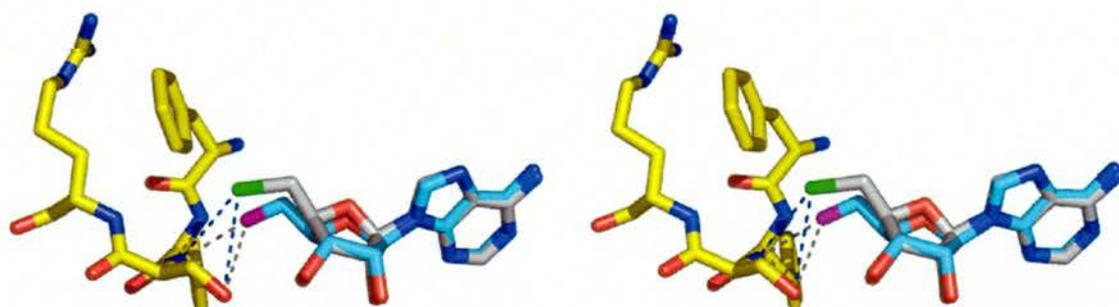


Figure 2.19 – Structure of the 5'-CIDA FDAS co-complex with contacts between the halogens and Ser-158 N and O atoms shown. 5'-CIDA (C = white / yellow, N = blue, O = red and Cl = green.) Shown superimposed is 5'-FDA (C = cyan, F = purple, N = blue). The chlorine atom lies further away from the backbone amide of Ser-158; hydrogen bonds are shown here Cl to N: 3.32 Å and Cl to O: 3.58 Å.

Interestingly a sphere of density was found, located approximately 6 Å above the plane of the ribose ring (Figure 2.18). The sphere, which was too large for a simple water molecule ($+7\sigma$) which was later refined as a chloride ion. The ion is coordinated to Tyr-266, Trp-217 and Asp-210 and is therefore not be expected to interfere with methionine binding. From this it is clear the gap in our understanding is the location of halide ion expected at Ser-158.

A summary of the data collection and final refinement is shown in Table 2.7. The Ramachandran diagram for this model suggests that the quality is good, and is shown in Figure 2.20.

Data collection	<i>5'-CIDA</i>
Wavelength (Å)	0.933
Resolution (Highest Shell) (Å)	45 – 2.0 (2.11 – 2.0)
Space Group	C222 ₁
Cell constants (Å)	a = 75.5, b = 129.2, c = 183.1
Unique reflections	56675 (8767)
Average redundancy	4.7 (4.8)
I/σ	13.6 (2.4)
Completeness (%)	98 (100.0)
R _{merge} (%)	9.5 (45.4)
Refinement	
R	19.0 (25.4)
R _{free}	24.7 (33.3)
Rmsd bond (Å)/angle(°)	0.01 / 1.21
Residues in Ramachandran core (%)	87.0
PDB accession code	2c2w

Table 2.7 – Summary of data collection and refinement statistics

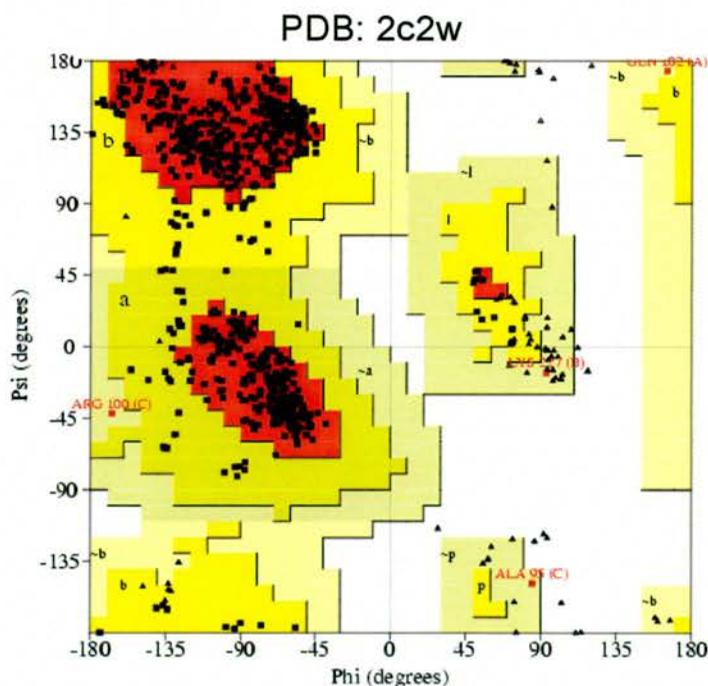


Figure 2.20 – Ramachandran plot of the model 2c2w

2.5.5 Co-crystallization of FDAS with 5'-deoxy-5'-[N-methyl-N-(γ -amino- γ -carboxypropyl)]-adenosine and chloride ion

The first attempt to locate the ion binding site used 5'-deoxy-5'-[N-methyl-N-(γ -amino- γ -carboxypropyl)]-adenosine (Aza-AdoMet), a stable SAM analogue and inhibitor of FDAS. Crystallization was pursued by pre-incubation of FDAS in the presence of ligand [20 mM] and chloride [20 mM NaCl]. Crystals were obtained by vapour diffusion (as outlined in the methods) in 32 % PEG 1K, 0.1 M phosphate citrate pH 4.2, 0.22 M Li_2SO_4 . Data were measured to 2.3 Å using an in-house radiation source. Phases were calculated using molecular replacement with the original FDAS structure, without the ligand. The unbiased F_o-F_c density suggested the presence of the Aza-AdoMet compound in each of the active sites (as shown in Figure 2.21). Refinement showed clear density for Aza-AdoMet but no chloride ion at the predicted binding site and no other chloride site. Attempts to co-crystallize ions at the active site were repeated with increasing concentrations of NaCl (up to 100 mM) showed no addition in observed density.

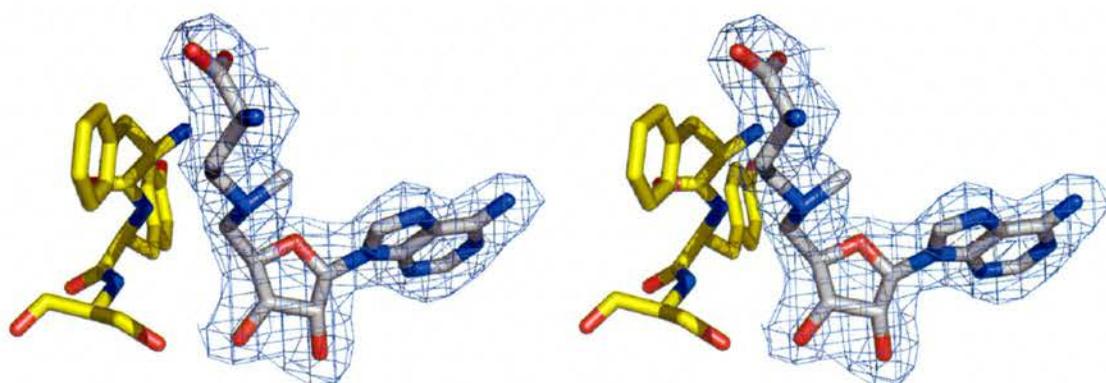


Figure 2.21 – Aza-AdoMet fully refined into the active site of FDAS, showing the final $2F_o-F_c$ electron density map (steel blue mesh and 1.0σ) for the ligand here (C = white and yellow; O = red; N = blue).

As expected, superimposition of the Aza-AdoMet (2c4t) and SAM (1rqp) showed the positions are almost identical. This supports the observation (Schaffrath et al., 2003) that the compound is a good competitive inhibitor ($K_i = \sim 30 \mu\text{M}$) and validates the SAM structure. The torsion angle across $2'\text{O}-2'\text{C}-3'\text{C}-3'\text{O}$ adopts an almost planar conformation which is also observed in the SAM model. This conformation is thought to be important in the catalytic mechanism by introducing the thermodynamically unfavourable conformation. This is achieved by interaction of Asp-16 and orientation of the methionine moiety, which is not observed in the product or product analogues.

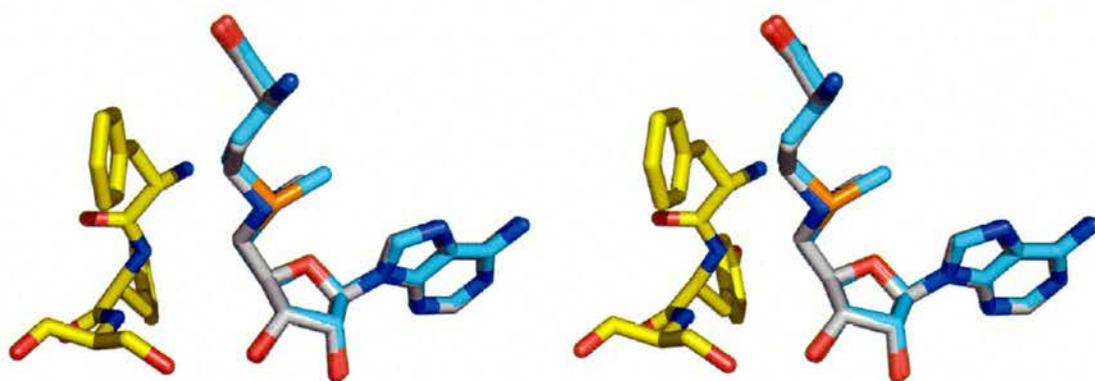


Figure 2.22 – Superimposition of the structures of Aza-AdoMet (2c4t) and S-adenosyl -L-methionine (1rqp)

A summary of the data collection and final refinement is shown in Table 2.8. The Ramachandran plot suggests the model is essentially correct and is shown in Figure 2.23.

Data collection	<i>AZA ADO MET + CI</i>
Wavelength (Å)	1.5418
Resolution (Highest Shell) (Å)	36.8 – 2.3 (2.4 – 2.3)
Space Group	C222 ₁
Cell constants (Å)	a = 76.3, b = 129.5, c = 182.8
Unique reflections	38585 (2844)
Average redundancy	4.8 (4.8)
I/σ	11.4 (2.3)
Completeness (%)	100.0 (100.0)
R _{merge} (%)	14.0 (33.0)
Refinement	
R	18.4 (23.3)
R _{free}	23.9 (31.8)
Rmsd bond (Å)/angle(°)	0.01 / 1.32
Residues in Ramachandran core (%)	87.5
PDB accession code	2c4t

Table 2.8 – Summary of data collection and refinement statistics

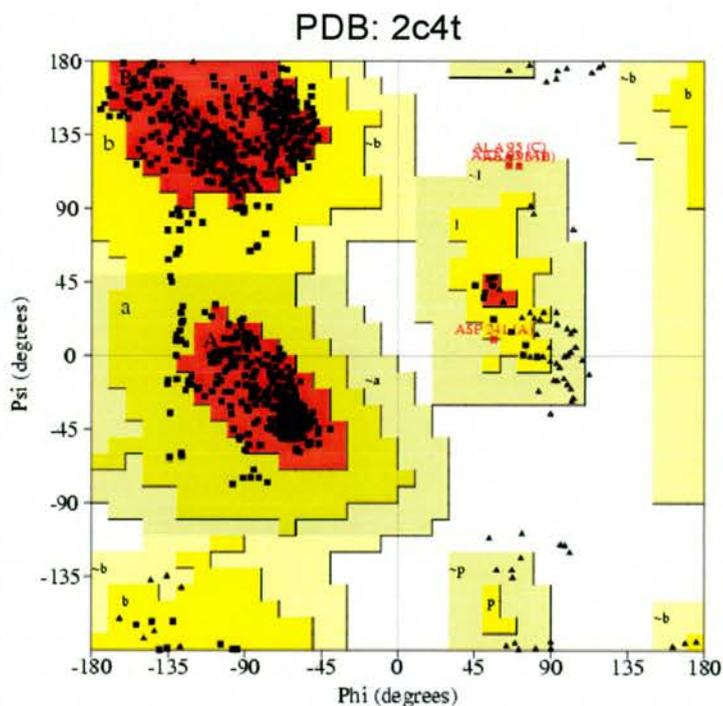


Figure 2.23 – Ramachandran plot of the model 2c4t

2.5.6 Co-crystallization of FDAS complexed with 5'-deoxyadenosine

A stable adenosine analogue was also chosen, as it was expected that this may allow binding of ions in the active sites. Crystallization was carried out after pre-incubation of FDAS in the presence of ligand [20 mM] and NaCl [20 mM]. Crystals were obtained by vapour diffusion (as outlined in the methods) in 24 % PEG 1000, 0.1 M phosphate citrate pH 4.4, 0.2 M Li₂SO₄. Data were measured to 2.0 Å using a synchrotron radiation source (ESRF, ID14 – EH2). Phases were calculated using molecular replacement with the original FDAS structure, without the ligand. The unbiased Fo-Fc density suggested the presence of 5'-deoxyadenosine in each of the active sites. The model was fully refined with the 5'-deoxyadenosine in the active site. The density indicated that there was no chloride ion bound.

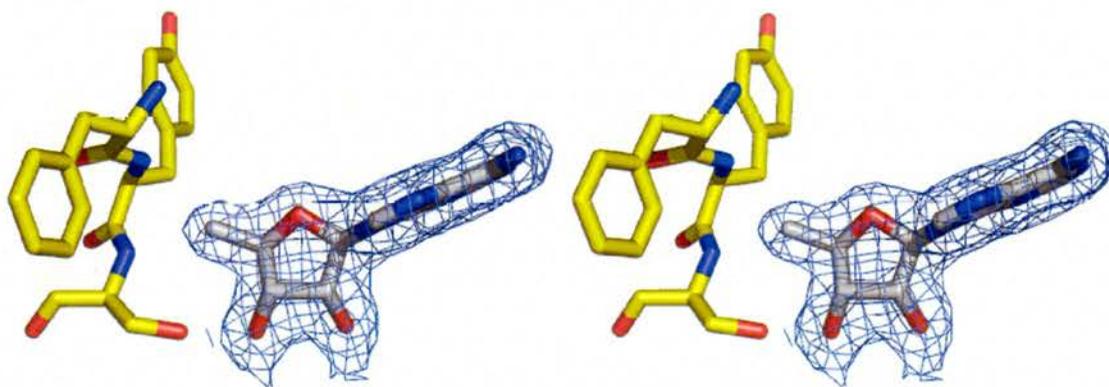


Figure 2.24 – 5'-deoxyadenosine fully refined into the active site of FDAS, showing the final 2Fo-Fc electron density map (steel blue mesh and 1.0 σ) for the ligand here (C = white and yellow; O = red; N = blue).

For a comparison, the model was superimposed with the 5'-FDA model. This is shown in Figure 2.25. There is little difference between both ligands, refining with the

essentially the same torsion angles across the ribose ring including 2' and 3' hydroxyls. This also suggests that the fluorine substituent does not have a unique effect on the overall conformation of the compound, which may be important when considering the catalytic mechanism.

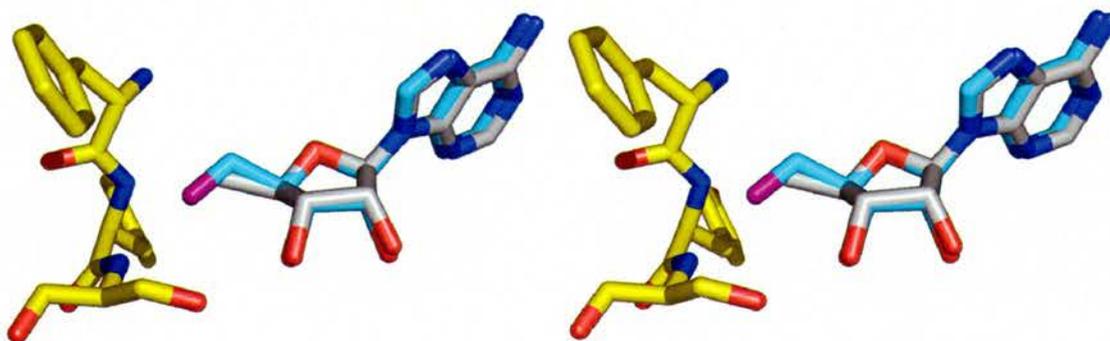


Figure 2.25 – Superimposition of 5'-deoxyadenosine (PDB 2cc2) and 5'-fluoro-deoxyadenosine (PDB 1rqr). (C = white, yellow, cyan; O = red; N = blue, F = purple).

A summary of the data collection and final refinement is shown in Table 2.9. The Ramachandran plot suggests the model is essentially correct and is shown in Figure 2.26.

Data collection	<i>5d adenosine</i>
Wavelength (Å)	0.933
Resolution (Highest Shell) (Å)	35.0 – 2.0 (2.1 – 2.0)
Space Group	C222 ₁
Cell constants (Å)	a = 76.0, b = 129.7, c = 184.5
Unique reflections	43477 (2981)
Average redundancy	2.0 (1.7)
I/σ	9.9 (1.8)
Completeness (%)	74.1 (66.9)
R _{merge} (%)	7.8 (27.2)
Refinement	
R	18.7 (23.2)
R _{free}	24.9 (34.8)
Rmsd bond (Å)/angle(°)	0.014 / 1.12
Residues in Ramachandran core (%)	86.2
PDB accession code	2cc2

Table 2.9 – Summary of data collection and refinement statistics

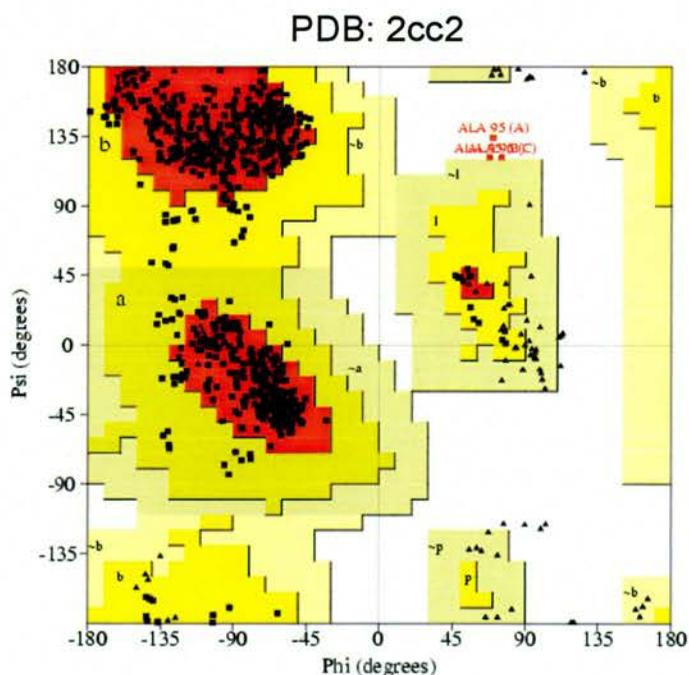


Figure 2.26 – Ramachandran plot of the model 2cc2

2.5.7 Co-crystallization of FDAS complexed with β -D-eythrofuranosyl adenosine and chloride ion

We predicted the novel adenosine homologue, β -D-eythrofuranosyl adenosine (Cadicamo and O'Hagan – Unpublished) would co-crystallize with halide ion in the active site of FDAS. Crystallization was carried out after pre-incubation of FDAS in the presence of ligand [20 mM] and chloride [up to 100 mM NaCl]. Crystals were obtained by vapour diffusion (as outlined in the methods) in 32 % PEG 1000, 0.1 M phosphate citrate pH 4.2, 0.2 M Li_2SO_4 and up to 100 mM NaCl. Data were measured to 2.0 Å using a synchrotron radiation source (SRS PX 14.1). Phases were calculated using molecular replacement with the original FDAS structure, without the ligand. The unbiased F_o - F_c density suggested the presence of the 5' deoxyadenosine compound in each of the active sites.

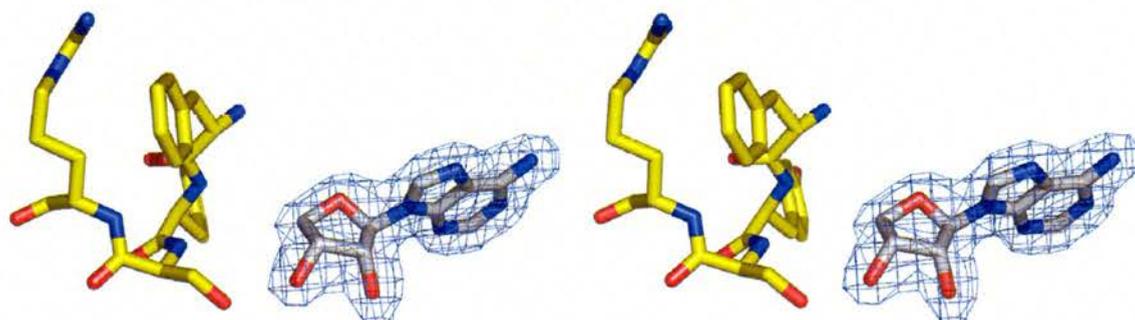


Figure 2.27 – β -D-eythrofuranosyl adenosine fully refined into the active site of FDAS, showing the final $2F_o$ - F_c electron density map (steel blue mesh and 1.0σ) for the ligand here (C = yellow and white; O = red; N = blue).

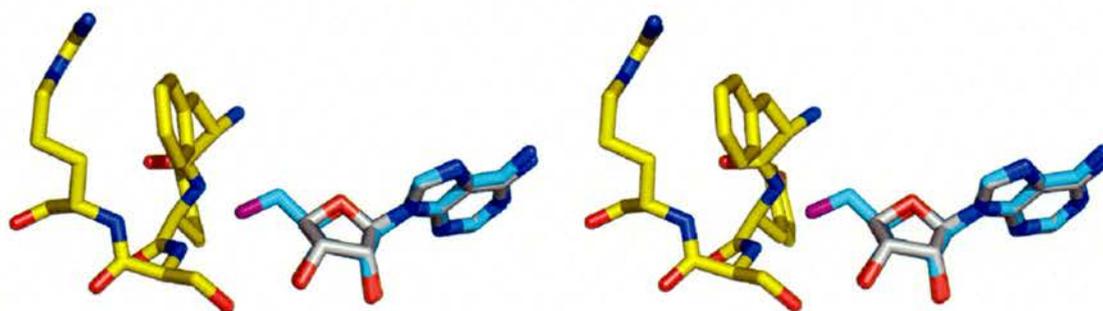


Figure 2.28 – Superimposition of CC5 (PDB 2cbx) and 5'-FDA (PDB 1rqr) (C = yellow, cyan, white; O = red; N = blue; F = purple).

This showed the ligand adopts the same position as 5'-FDA with the same conformation across the ribose hydroxyls. No density was observed that would indicate the presence of a halide.

The experiment was repeated using increasing concentrations of NaCl (or KCl) (up to 100 mM). X-ray diffraction analysis revealed no extra density which could be attributed to a halide. Efforts to increase halide concentration over 100 mM resulted in disordered precipitation of FDAS. New crystallization conditions may be required which will permit saturation.

A summary of the data collection and final refinement is shown in Table 2.10. The Ramachandran plot is shown in Figure 2.29, suggesting the model quality is good.

Data collection	<i>3-cc+Cl</i>
Wavelength (Å)	0.87
Resolution (Highest Shell) (Å)	52.5 – 2.0 (2.1 – 2.0)
Space Group	C222 ₁
Cell constants (Å)	a = 74.82, b = 128.1, c = 181.1
Unique reflections	54334 (2898)
Average redundancy	3.7 (3.7)
I/σ	14.7 (3.1)
Completeness (%)	97.6 (97.1)
R _{merge} (%)	7.2 (38.5)
Refinement	
R	17.5 (18.6)
R _{free}	23.8 (26.5)
Rmsd bond (Å)/angle(°)	0.01 / 1.32
Residues in Ramachandran core (%)	87.6
PDB accession code	2cbx

Table 2.10 – Summary of data collection and refinement statistics

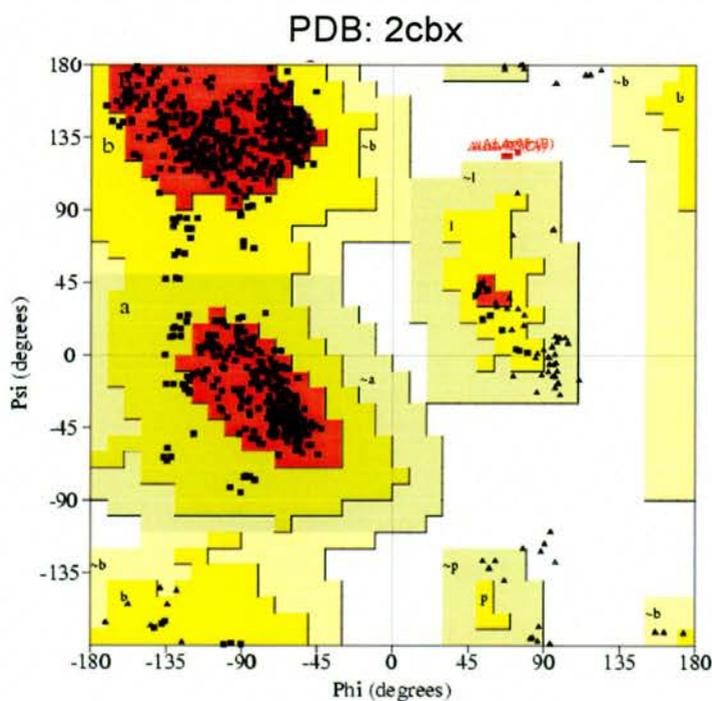


Figure 2.29 – Ramachandran plot of the model 2cbx

2.5.8 Co-crystallization of FDAS complexed with 5'-fluoro-2',5'-dideoxyadenosine

To probe the role of Asp-16 binding on the ribose ring, crystallization was pursued by pre-incubation of FDAS in the presence of 5'-fluoro-2', 5'-dideoxyadenosine (2'-d-FDA) [20 mM] and L-Met [50 mM]. Crystals were obtained by vapour diffusion (as outlined in the methods) in 32 % PEG 1000, 0.1 M phosphate citrate pH 4.2, 0.2 M Li_2SO_4 . Data were measured to 2.5 Å using an in-house radiation source. Phases were calculated using molecular replacement with the original FDAS structure. The unbiased Fo-Fc density suggested the presence of the 2'-d-FDA compound and density for the methionine in each of the active sites. The ligands were modelled and refined into the well fitting density. Figure 2.30 shows the final refined model of the ligand and methionine.

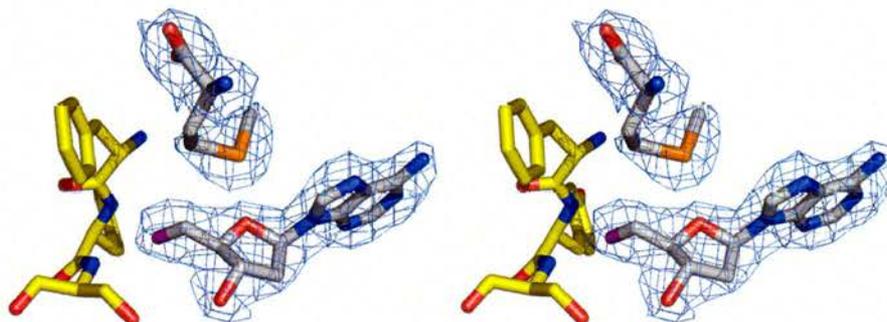


Figure 2.30 – 5'-fluoro-2',5'-dideoxyadenosine (2'-d-FDA) fully refined into the active site of FDAS (C = yellow and white; O = red; N = blue; F = purple; S = orange).

Superimposition of the models (Figure 2.31 below) reveals the 2'-d-FDA (5F1 in PDB entry 2c5b) is restrained in a grossly similar position to 5'-fluoro-5'-deoxyadenosine (5'-FDA in PDB entry 1rqr); the adenine portion is relatively identical however the 2-deoxy ribose ring is restrained differently to the ribose ring of 5'-FDA. The geometry of

the ring is clearly altered as the 3'-hydroxyl moves to accommodate a bifurcated hydrogen bond with between O1 and O2 of Asp-16. The ribose ring becomes puckered in the 2'-d-FDA (C-C-C-C torsion 30.4°), relative to the planar situation found with 5'-FDA (C-C-C-C torsion 1°). The 3'-hydroxyl still retains, albeit longer, hydrogen bonding contacts to the OH of Ser-158 and the NH of Tyr-77. Overall this results in the C5' altering into a different position by about 0.7 \AA which may account for any difference in reactivity observed. It is also expected from this that the corresponding interactions will also be altered in a similar mode. The final relative position of the fluorine is the almost identical for both molecules, suggesting a specific pocket for the fluorine.

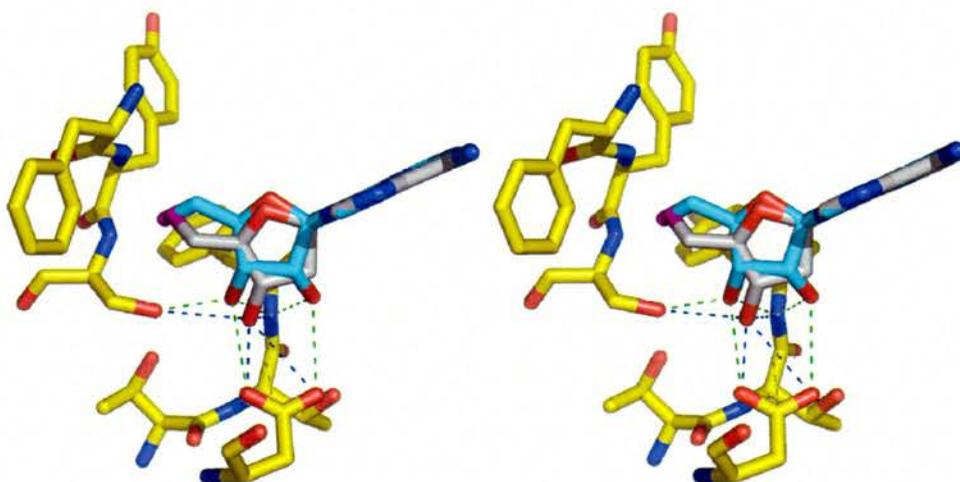


Figure 2.31 – Superimposition of 5'-fluoro-2',5'-dideoxy adenosine (5F1, 2c5b), (C = white and yellow; O = red; N = blue; F = purple) and of 5'-fluoro-5'-deoxy adenosine (5FD, 1rqr) (C = cyan). Interactions of 2'-deoxy ligand to some protein residues are indicated; 03' to Ser-158 (O) = 3.2 \AA ; 03' to Asp-16 (O1) = 2.5 \AA ; 03' to Asp-16 (O2) = 3.2 \AA ; 03' to Tyr-77 (N) = 3.3 \AA

A summary of the data collection and final refinement is shown in Table 2.11. The Ramachandran plot is shown in Figure 2.32, suggesting the model quality is good.

Data collection	<i>2d-fda</i>
Wavelength (Å)	1.5418
Resolution (Highest Shell) (Å)	37.6 – 2.5 (2.64 – 2.5)
Space Group	C222 ₁
Cell constants (Å)	a = 76.1, b = 129.7, c = 183.2
Unique reflections	27454 (3998)
Average redundancy	2.9 (2.8)
I/σ	7.3 (3.4)
Completeness (%)	87 (88)
R _{merge} (%)	13.9 (32.4)
Refinement	
R	18.8 (24.9)
R _{free}	26.7 (35.8)
Rmsd bond (Å)/angle(°)	0.01 / 1.41
Residues in Ramachandran core (%)	84.2
PDB accession code	2c5b

Table 2.11 – Summary of data collection and refinement statistics

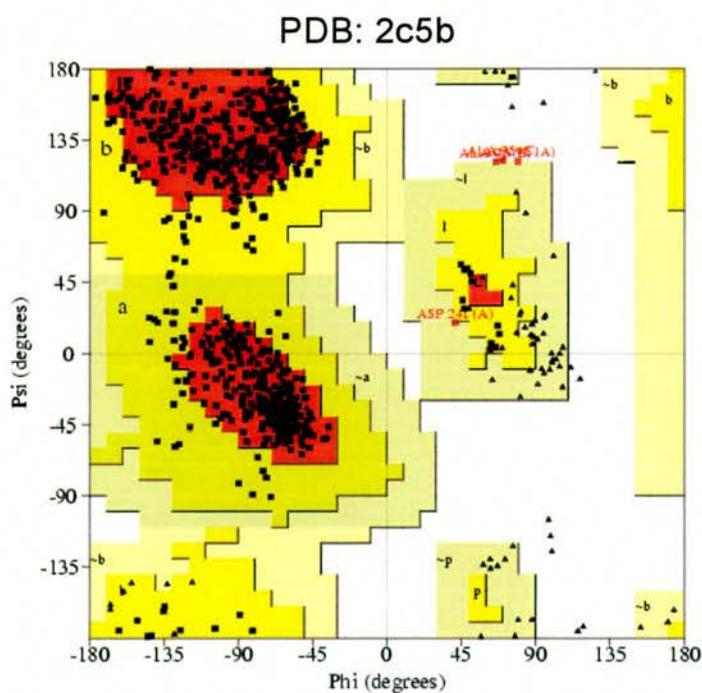


Figure 2.32 – Ramachandran plot of the model 2c5b

2.5.9 Co-crystallization and X-ray crystal structure of FDAS complexed with 2'-deoxyadenosine

Crystallization was pursued by pre-incubation of FDAS in the presence of 2'-deoxyadenosine [20 mM] and L-Met [50 mM]. Crystals were obtained by vapour diffusion (as outlined in the methods) against 32 % PEG 1000, 0.1 M phosphate citrate pH 4.2, 0.2 M Li_2SO_4 . Data were measured to 2.5 Å using synchrotron radiation source (SRS – PX 14-1). Phases were calculated using molecular replacement with the original FDAS structure, without ligand.

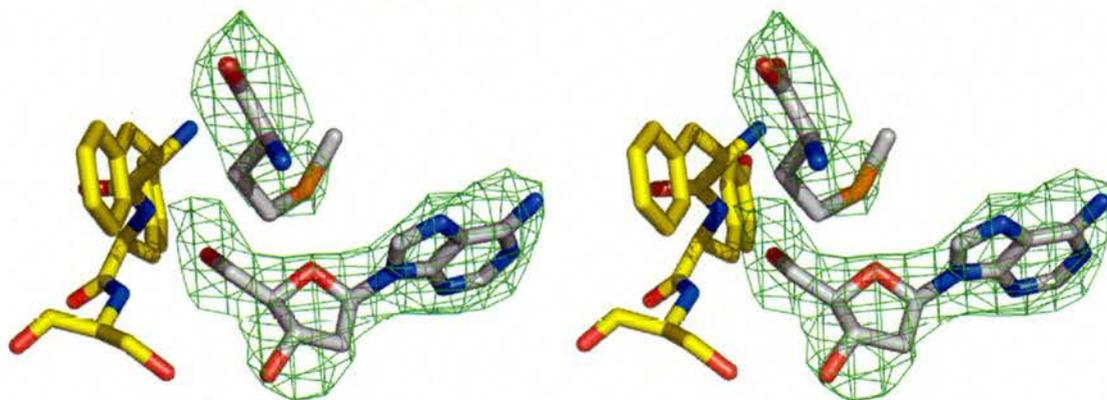


Figure 2.33 – 2'-deoxyadenosine (2DA) and methionine modeled into the unbiased $F_o - F_c$ (green mesh and $+3.0 \sigma$) electron density map, after molecular replacement with *Irqp* (without ligand), in the active site of FDAS (C = white and yellow; O = red; N = blue).

Superimposition shows that the ligands adopt the same conformation for both 2'-deoxy compounds. This is shown in Figure 2.34. The torsion remains puckered and the 3'-hydroxyl forms bifurcated hydrogen bonding to Asp-16. The relative position of the 5' substituent is identical in both molecules suggesting that there is no significant difference between the binding of either compound.

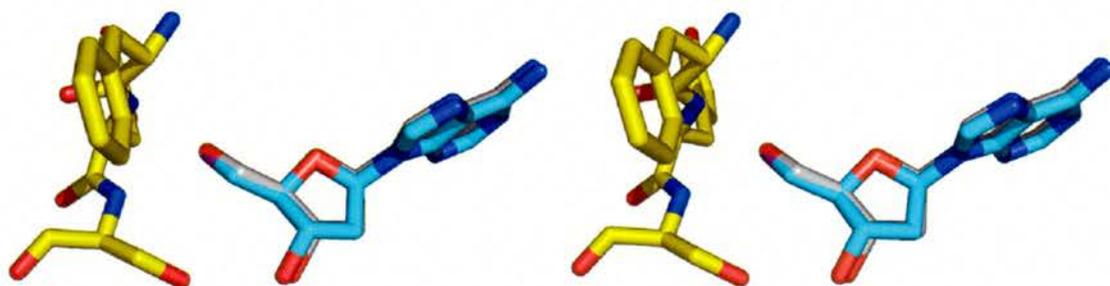


Figure 2.34 – Superimposition of 2'-dFDA (C = cyan; O = red; N = blue and F = purple) with 2'-deoxyadenosine in the active site of FDAS. (C = white and yellow; O = red; N = blue)

A summary of the data collection and final refinement is shown in Table 2.12.

The Ramachandran plot is shown in Figure 2.35, suggesting the model quality is good.

Data collection	<i>2d-ad</i>
Wavelength (Å)	1.488
Resolution (Highest Shell) (Å)	59.8 – 2.5 (2.6 – 2.5)
Space Group	C222 ₁
Cell constants (Å)	A = 74.5, b = 126.7, c = 181.3
Unique reflections	28755 (4289)
Average redundancy	3.2 (3.2)
I/σ	10.9 (1.6)
Completeness (%)	95.6 (98.5)
R _{merge} (%)	8.6 (57.0)
Refinement	
R	22.8 (39.1)
R _{free}	29.3 (51.9)
Rmsd bond (Å)/angle(°)	0.013 / 1.564
Residues in Ramachandran core (%)	82.1
PDB accession code	2c5h

Table 2.12 – Summary of data collection and refinement statistics

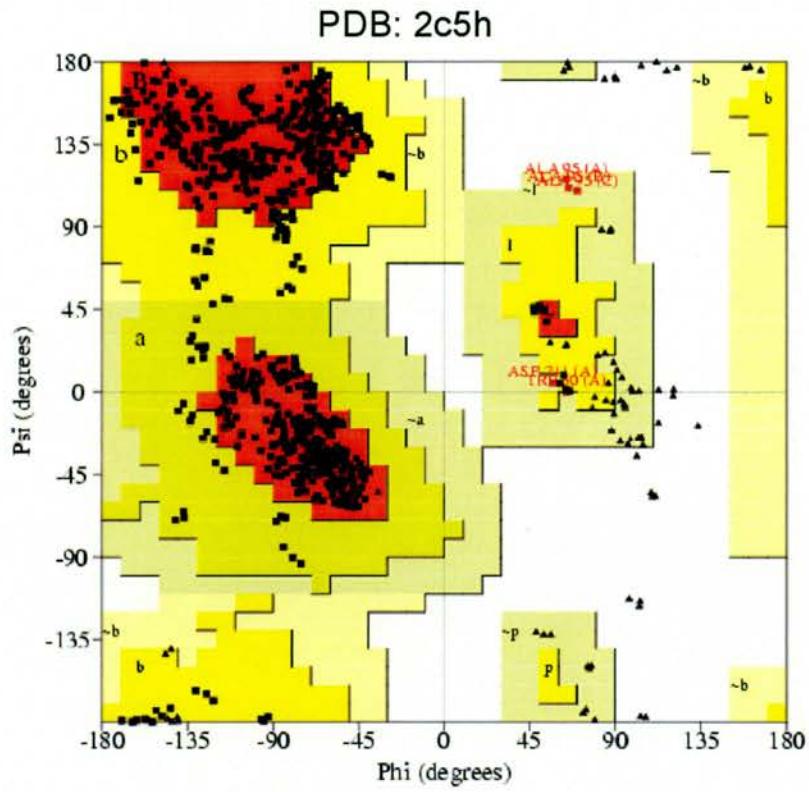


Figure 2.35 – Ramachandran plot of the model 2c5h

2.5.10 Crystallization and structure solution of apo-FDAS

The native apo-protein did not crystallize in the previously reported conditions (Table 2.3) nor during sparse matrix screening using the commercially available screens. Hits were however obtained for the Asn-215-Arg mutant (which was shown to contain no small molecules in Section 2.5.3) in The Pegs™ sparse matrix screen (Nextal). These hits were optimized using the native protein and eventually yielded diffractable crystals (as shown in Figure 2.36).

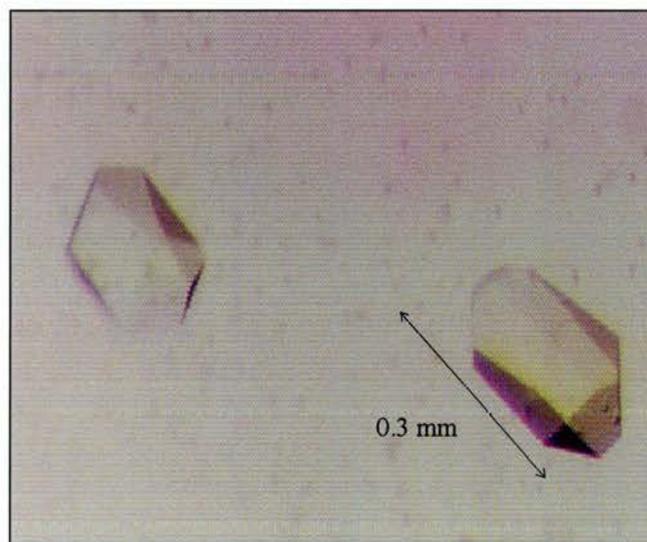


Figure 2.36 – Crystals obtained of the apo-FDAS.

Crystals precipitated in drops containing 0.2 M sodium thiocyanate, 20 % w/v PEG 3350. The crystals grew to full size after 2 to 6 weeks. The crystals are clearly different in morphology compared with crystals grown in the presence of ligand. Data were measured to 2.5 Å using synchrotron radiation source (ESRF – ID29). The data was processed in $P2_12_12_1$. The structure was solved using molecular replacement which located two trimers in the ASU, which constructed the hexamer. The Fo-Fc map

confirmed that this was the apo-form of FDAS, since there was clearly no density that could be attributed to a ligand (Figure 2.37).

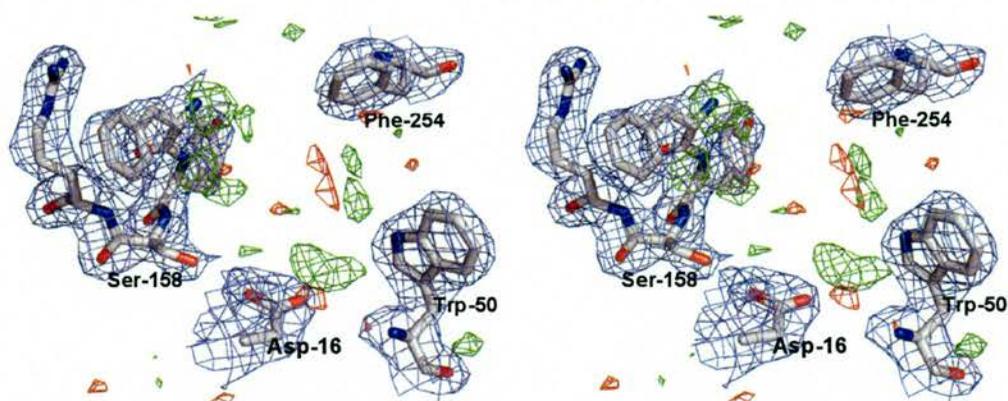


Figure 2.37 – Diagram of the unbiased F_o-F_c density (green mesh and $+3.0 \sigma$; red mesh and -3.0σ) in the active site of the apo crystals. The initial $2F_o-F_c$ density obtained is shown here also (steel blue mesh and 1.0σ). (C = white, O = red, N = blue).

The final density calculated revealed several glycerol molecules bound to the protein as a result of cryo-protection. No ligand density became apparent at any stage of the refinement. A summary of the data collection and final refinement is shown in Table 2.13. The Ramachandran plot is shown in Figure 2.38, suggesting the model quality is good.

Data collection	APO
Wavelength (Å)	0.953
Resolution (Highest Shell) (Å)	73.0 – 2.5 (2.6 – 2.5)
Space Group	P2 ₁ 2 ₁ 2 ₁
Cell constants (Å)	a = 103.9, b = 128.1, c = 146.5
Unique reflections	64329 (4760)
Average redundancy	3.6 (3.6)
I/σ	12.2 (3.6)
Completeness (%)	99.5 (100.0)
R _{merge} (%)	11.0 (64.0)
Refinement	
R	21.8 (28.1)
R _{free}	27.2 (33.9)
Rmsd bond (Å)/angle(°)	0.012 / 1.47
Residues in Ramachandran core (%)	86.4
PDB accession code	2c4u

Table 2.13 – Summary of data collection and refinement statistics

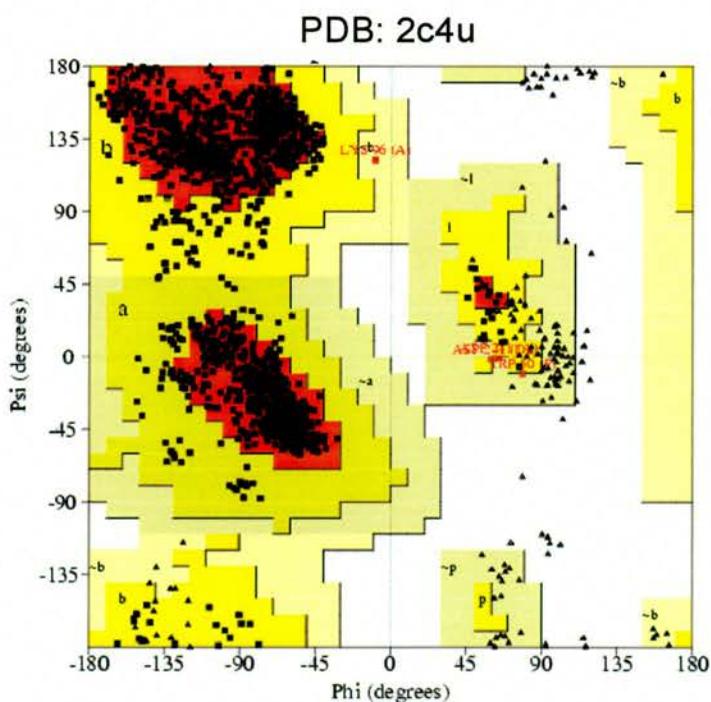


Figure 2.38 – Ramachandran plot of the model 2c4u

2.5.11 Overall structure comparisons

The α -carbon trace of 1rqp (SAM), 1rqp (FDA) and 2c4u (the apo-structure) were compared as shown in Figure 2.39 (below).

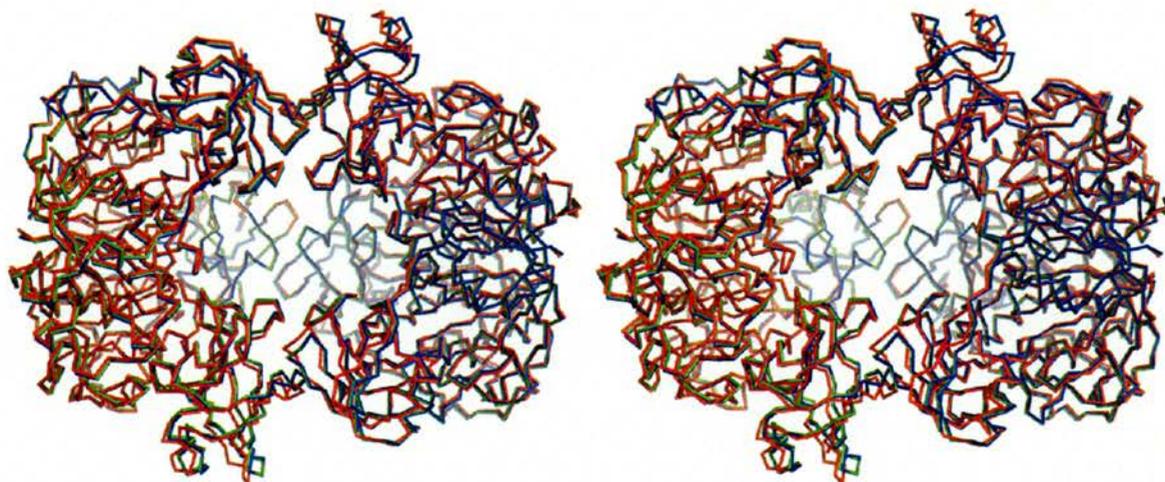


Figure 2.39 – Superimpositions of the carbon α trace of 1rqp (SAM co-complex: blue), 1rqp (FDA co-complex: green) and 2c4u (apo-model: red), as the hexamer.

These models were chosen as they represent the three stages of the enzyme. The superimposition of these models reveals no significant structural differences in the overall structures. It had been expected that the apo-form of the enzyme would adopt an open conformation to accommodate the ligand binding; however this is not observed from the crystal structure obtained here.

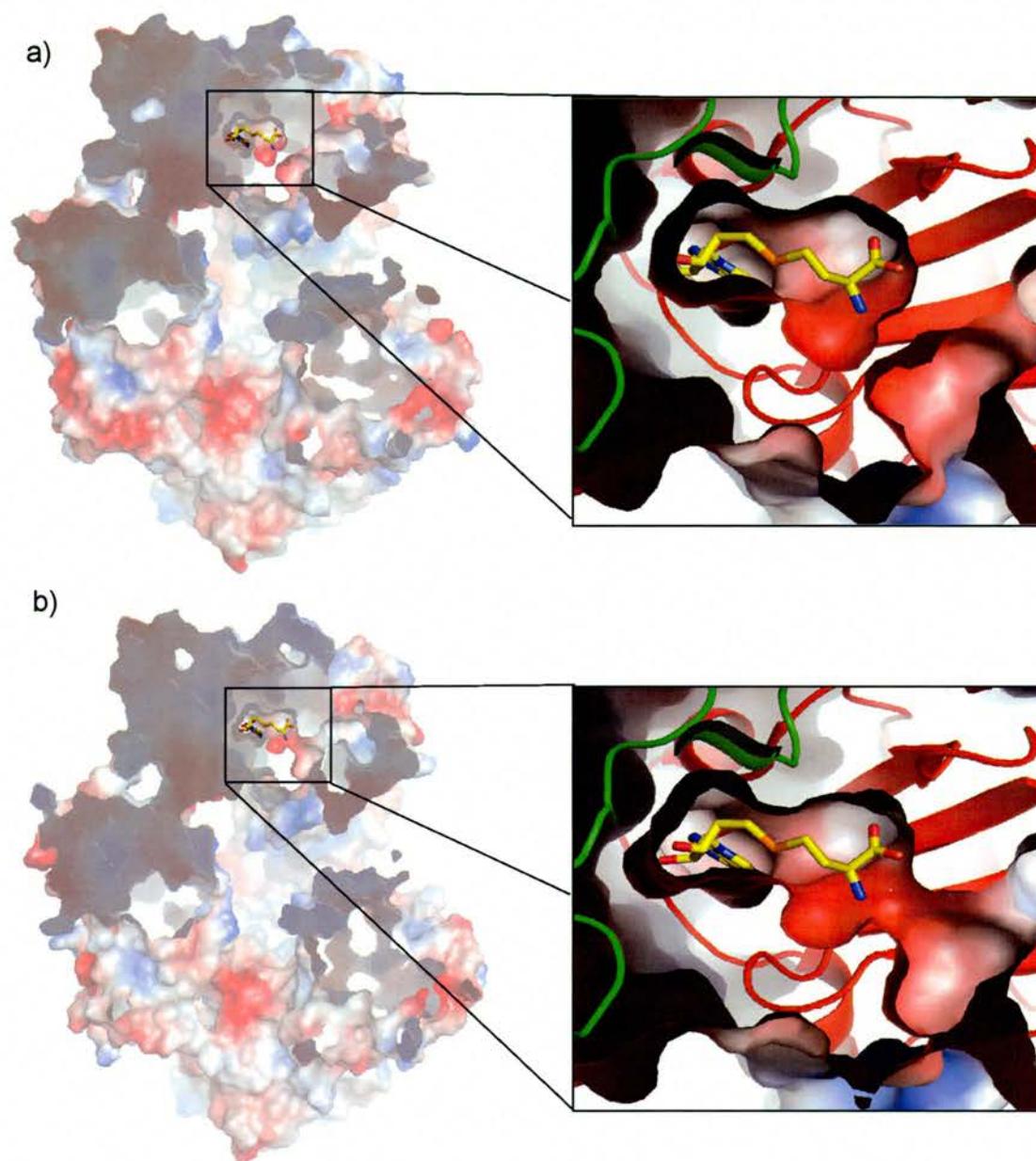


Figure 2.40 – Dissection of the surface potential (negative in red, neutral in white, positive in blue) and reveals the buried active site of FDAS, SAM has been modelled into each (C = yellow, S = orange, N = blue, O = red). a) Relates to the co-crystal structure of substrate (PDB: 1rqp); b) Relates to the apo-structure (PDB: 2c4u) SAM was modelled into the binding site. Red and green cartoon models are chain A and B respectively.

The above diagram (Figure 2.40) shows a section through the calculated electrostatic surface potentials for models 1rqp and 2c4u, to reveal the buried active site of FDAS. Interestingly in all the co-crystal structures, the active site is completely isolated from the surface of the protein. Although subtle, the refined structures were different enough to suggest a potential entry site for ligands of the FDAS. This occurs due to the accumulation of slight differences across several residues, such as Asp-21, Arg-270 and Asp-210 which surround the ‘opening’ of the active site. Arg-270 in the apo-model appears to move away by ~ 1.7 Å compared with 1rqp. This could be an effect of the crystals, as protein will adopt an energetically favourable conformation for crystallization. This however does suggest that the apo-protein undergoes some change to permit entry into the buried active site.

With respect to the ligands, overall structural comparisons can give some insight to the mechanism of the reaction. Firstly the organo-halogen binding site is generally consistent between models. This is clear from the models containing fluorinated ligands and indicates that the organo-fluorine is well restrained by the protein interactions. The disparity observed when comparing the fluorine and chlorine can be accounted for, by the relative size of the atoms. The models containing 5'-hydroxyls (adenosine and 2'-deoxyadenosine) confirm that the fluorine permits a hydrogen like bonding to the protein, as both groups adopt identical conformations. This also indicates that hydroxyl groups are accommodated in the halide binding site however reaction of OH⁻ is not possible as there are no residues capable of activating the anion.

The torsional restraints observed in the product and substrate models across the ribose ring and between 2' and 3' hydroxyls appear to be conserved with respect to their analogous structures. The restraints observed in SAM are also found in the Aza-AdoMet

model, such that the 2' and 3' hydroxyls adopt an unfavourable planar conformation. Similarly, 5'-FDA, 5'-CIDA and adenosine adopt a more energetically favourable conformation. Overall these observations indicate that a torsional change is an effect of reaction and suggest that there may be energetic gain with respect to the reaction. The structures of the 2'-deoxy analogues present a third conformation, mainly as a result of change in interactions with Asp-16. This allows us to explain the reduced activity towards the 2'-deoxy analogues.

2.5.12 Isothermal Calorimetry

FDAS-ligand interactions were analyzed using isothermal calorimetry (ITC), which measures thermal changes during titration between molecules. In this case, an ITC titration experiment is carried out by introducing ligand solution into protein solution. As the species react, depending on the interaction, heat may be released or absorbed (Pierce et al., 1999). By measurement of the heats of interaction, the binding constant (K_a), reaction stoichiometry (n), and thermodynamic parameters such as enthalpy (ΔH) and entropy (ΔS) can be determined ($\Delta G^\circ = -RT \ln Ka = \Delta H^\circ - T\Delta S^\circ$).

Several ligands were titrated against FDAS as described previously (Section 2.4.10). Initial experiments were carried out in Tris-HCl buffer however due to the presence of chloride some titrations were repeated in phosphate to study the effects of the buffer ($T = 298 \text{ K}$). It was also shown there are uncertainties in the observed stoichiometry of the titrations with $n < 1$ in all cases, this may be explained by partially inactive protein in the sample (possibly due to unfolding, contamination or miscalculation of concentrations). To ensure that this was minimised the protein was re-purified by size exclusion chromatography prior to use, to reduce unfolded protein or contamination. The concentration of the protein was confirmed by UV_{280} and Bradford assay (as before). The concentration of ligands were initially calculated by weight and confirmed by UV_{260} .

Figures 2.41 to 2.48 show the raw ITC thermograms (TOP) and ΔH plot (BOTTOM) with fitting function parameters for the titrations of ligand (as listed) against FDAS, where stoichiometry is unrestrained. The results are summarised in Table 2.14.

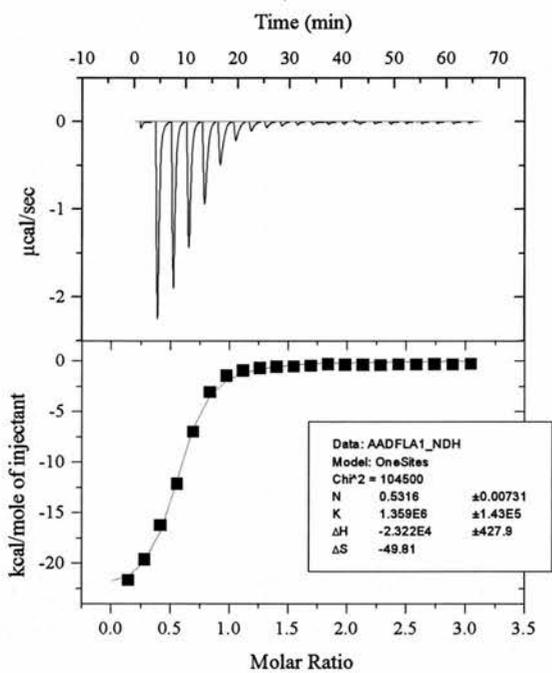


Figure 2.41 – Adenosine (Tris)

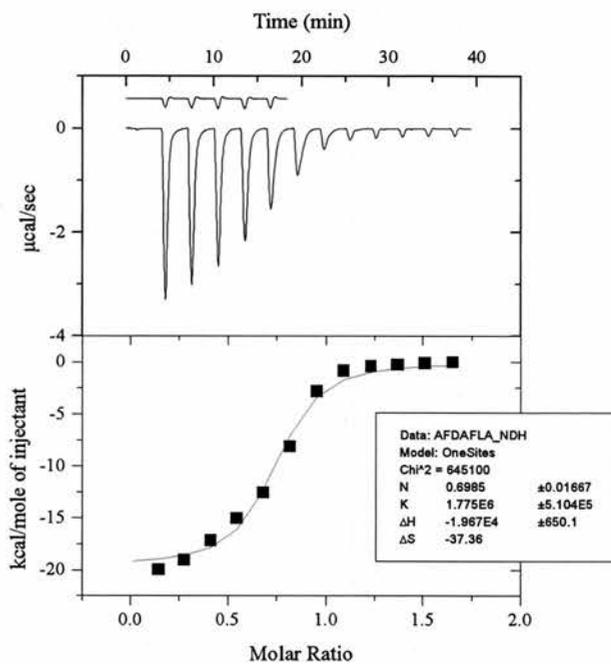


Figure 2.42 – FDA (Tris)

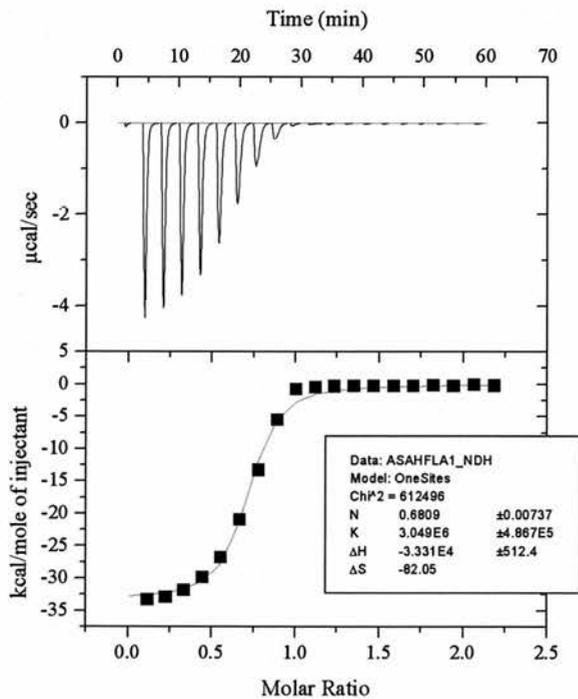


Figure 2.43 – SAH (Tris)

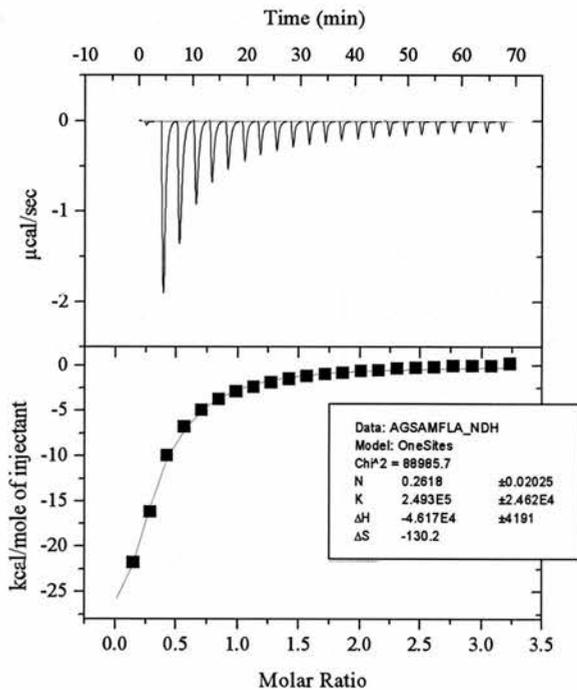


Figure 2.44 – SAM (Tris)

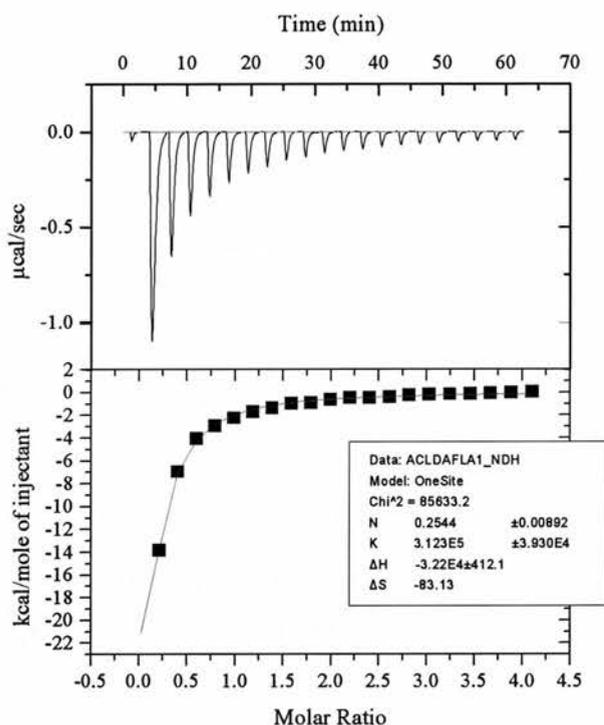


Figure 2.45 – 5'-CIDA (Tris)

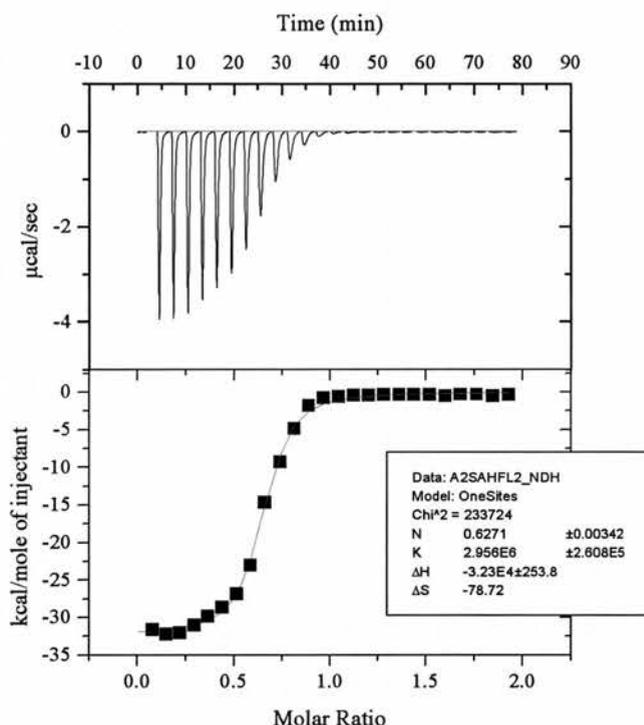


Figure 2.46 – SAH (Phosphate)

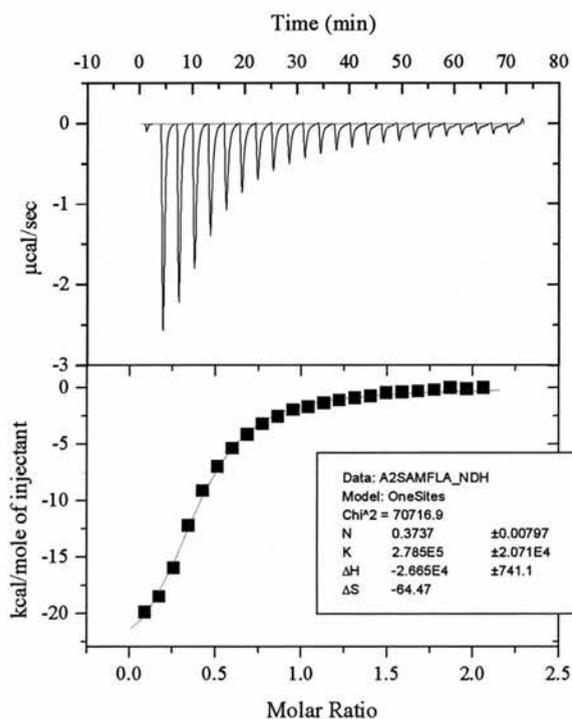


Figure 2.47 – SAM (Phosphate)

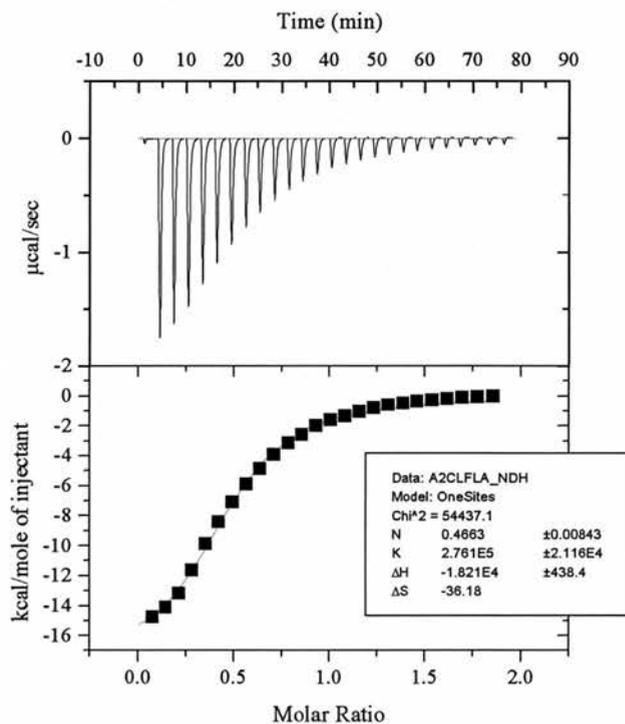


Figure 2.48 – 5'-CIDA (Phosphate)

Ligand	Buffer	Stoichiometry (n)	K_a (app) ($\times 10^8 \text{ M}^{-1}$)	ΔH° (kcal mol $^{-1}$)	ΔS° (cal mol $^{-1}$ K $^{-1}$)	$T\Delta S^\circ$ (kcal mol $^{-1}$)	ΔG° (kcal mol $^{-1}$)
Ad	Tris(HCl)	0.53 ± 0.007	1.36 ± 0.14	-23.2 ± 0.04	-49.8 ± 0.4	-14.8 ± 0.2	-8.4 ± 0.1
FDA	Tris(HCl)	0.70 ± 0.016	1.77 ± 0.51	-19.6 ± 0.06	-37.3 ± 0.8	-11.1 ± 0.4	-8.5 ± 0.1
SAH	Tris(HCl)	0.68 ± 0.007	3.05 ± 0.48	-33.1 ± 0.05	-82.0 ± 0.9	-24.4 ± 0.8	-8.7 ± 0.2
SAM	Tris(HCl)	0.26 ± 0.020	0.25 ± 0.02	-46.1 ± 0.41	-130.0 ± 1.4	-38.7 ± 1.1	-7.4 ± 0.2
CIDA	Tris(HCl)	0.25 ± 0.009	0.31 ± 0.03	-32.2 ± 0.12	-83.1 ± 0.6	-24.7 ± 0.3	-7.5 ± 0.2
SAH	Phosphate	0.63 ± 0.003	2.96 ± 0.26	-32.3 ± 0.02	-78.7 ± 0.2	-23.4 ± 0.1	-8.9 ± 0.1
SAM	Phosphate	0.37 ± 0.008	0.28 ± 0.02	-26.6 ± 0.07	-64.5 ± 0.4	-19.2 ± 0.2	-7.4 ± 0.1
CIDA	Phosphate	0.47 ± 0.008	0.28 ± 0.02	-18.2 ± 0.04	-36.2 ± 0.4	-10.8 ± 0.2	-7.4 ± 0.1

Table 2.14 – Summary of ITC data results (according to one site model), with unrestrained stoichiometry and experimental concentrations.

Clearly the order of binding affinity (K_a) according to the data obtained is SAH > FDA > Adenosine > CIDA ≥ SAM. Unexpectedly the binding profiles of SAH and SAM are very different. For SAH, the enthalpies are the same in each buffer ruling out any binding induced protonation of either the protein or the ligand. The situation with SAM and 5'-CLDA is more complicated, as although the binding affinities are not affected by the buffer, the apparent heats of binding are significantly more exothermic in Tris compared to phosphate. This behaviour is typical of binding-induced release of protons. The ΔH° values suggest that at least one proton (possibly more) is displaced during SAM or 5'-CIDA binding which are absorbed by the buffer to give an extra heat contribution (as H^+ binding to Tris is more exothermic than for phosphate). This could be a result of a significant change in pK_a of the protein or ligand. These observations may be related to the slight changes in protein conformation between the apo and ligand bound models.

We decided to recalculate the data such that $n = \sim 1$, by adjustment of protein concentrations (Figures 2.49 to 2.53)

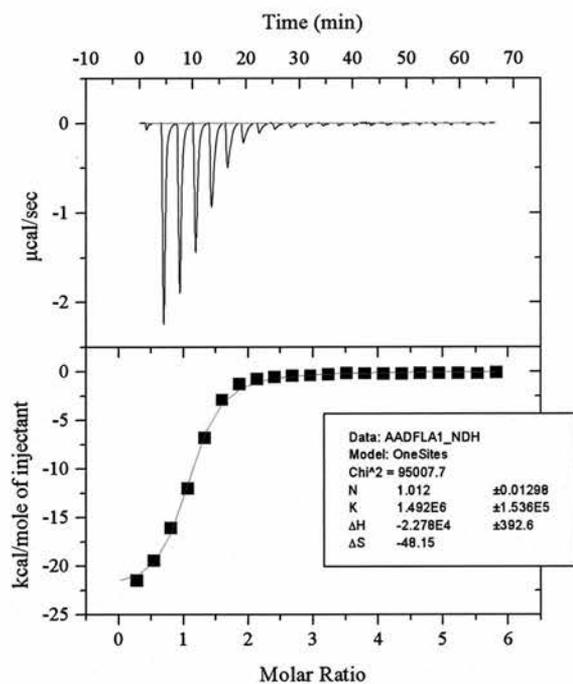


Figure 2.49 – Adenosine (Tris)

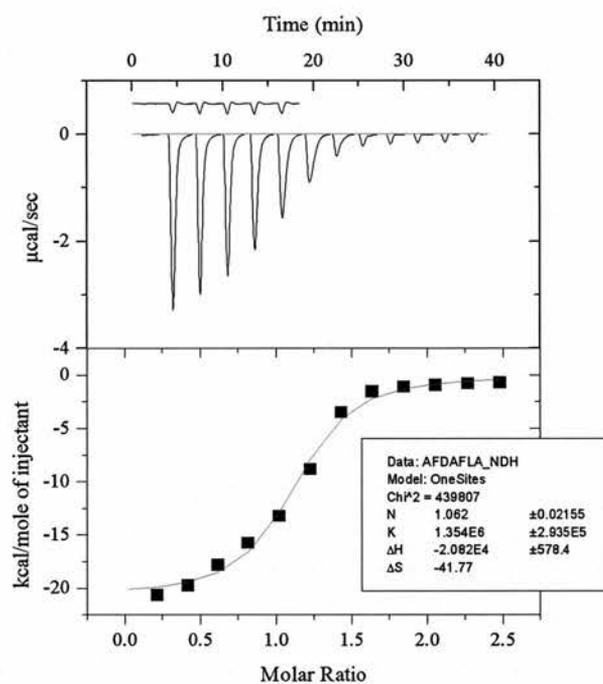


Figure 2.50 – FDA (Tris)

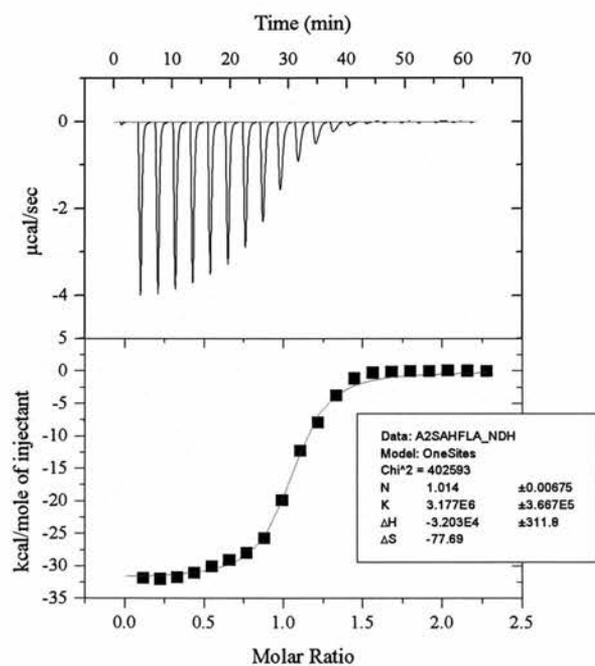


Figure 2.51 – SAH (Phosphate)

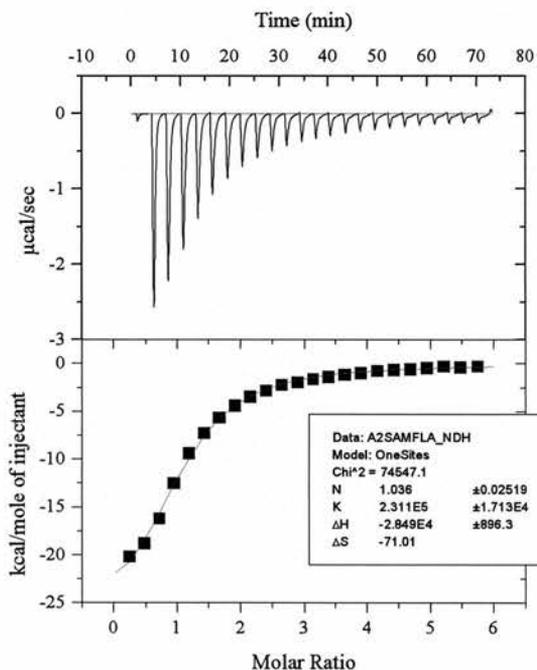


Figure 2.52 – SAM (Phosphate)

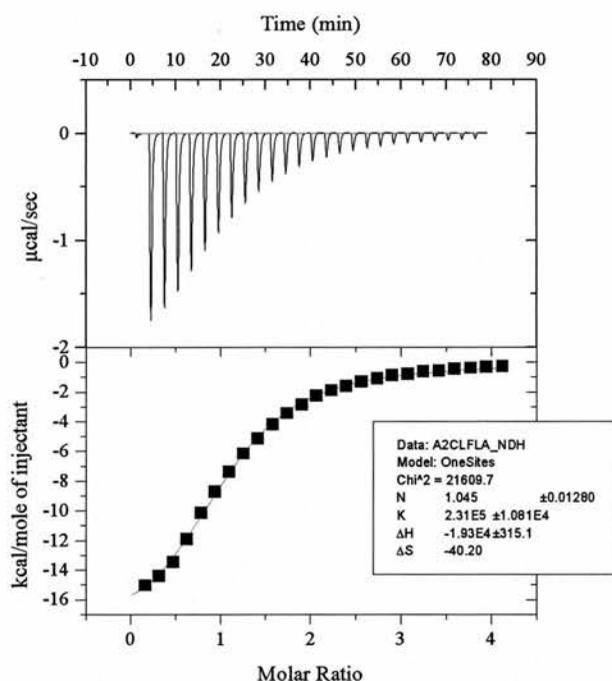


Figure 2.53 – 5'-CIDA (Phosphate)

Ligand	Buffer	Stoichiometry (n)	K_a (app) ($\times 10^6 \text{ M}^{-1}$)	ΔH° (kcal mol $^{-1}$)	ΔS° (cal mol $^{-1}$ K $^{-1}$)	$T\Delta S^\circ$ (kcal mol $^{-1}$)	ΔG° (kcal mol $^{-1}$)
Ad	Tris(HCl)	1.014 ± 0.013	1.49 ± 0.15	-22.7 ± 0.04	-48.2 ± 0.4	-14.3 ± 0.2	-8.4 ± 0.1
FDA	Tris(HCl)	1.062 ± 0.021	1.35 ± 0.29	-20.8 ± 0.06	-41.8 ± 0.5	-12.4 ± 0.4	-8.4 ± 0.1
SAH	Phosphate	1.014 ± 0.006	3.17 ± 0.37	-32.0 ± 0.03	-77.7 ± 0.3	-23.1 ± 0.1	-8.9 ± 0.1
SAM	Phosphate	1.036 ± 0.025	0.23 ± 0.02	-28.5 ± 0.09	-71.0 ± 0.8	-21.1 ± 0.2	-7.4 ± 0.1
CIDA	Phosphate	1.045 ± 0.013	0.23 ± 0.01	-19.3 ± 0.03	-40.2 ± 0.4	-11.9 ± 0.2	-7.4 ± 0.1

Table 2.15 – Summary of ITC data results (according to one site model) recalculated so that $n = \sim 1.0$

Correction of the stoichiometry by adjusting protein concentration does not change the ΔH° or K values obtained for binding affinities significantly. Alternatively the ligand concentration could be altered which would result in bigger changes. Since there is less ambiguity in calculation of ligand concentration, this can be disregarded.

Overall the dissociation constants ($K_d = 1/K_a$) can be obtained also for the ligands; SAH is 0.32 nM, adenosine is 0.66 nM, FDA is 0.74 nM, SAM is 3.17 nM and CIDA is 3.17 nM. This indicates why we find adenosine as a contaminant carried through FDAS purification and why SAH is a good competitive inhibitor. There appears to be a difference in binding profile between SAH (or adenosine and FDA) with respect to SAM (or CIDA).

2.6 CONCLUSIONS

To conclude, this study revealed the X-ray crystal structure of several ligands co-crystallized into the active site of FDAS. The structures were only obtainable after identification and removal of the adenosine contaminant from recombinant FDAS. These structures revealed (in combination with other work) interesting features of the structure and mechanism of FDAS. Firstly the 5'-CIDA structure (2c2w) revealed that FDAS can accept 5'-CIDA and the chlorine at the 5' position will orientate in a similar conformation to the fluorine of 5'-FDA. This is significant as it confirms the halide binding site observed in the FDA structure. The final position is ~ 1.3 Å different relative to fluorine (as a result of the van der Waals radius). A second conformer is also observed which orients itself into the methionine binding site, but should not be observed in the presence of L-Met. The structures of Aza-AdoMet in FDAS show the ligand adopts an essentially identical conformation to that of the natural substrate (SAM). The torsions angles across the ribose ring are the same as for those in the SAM model. The structure is the first example of this compound found in the PDB. The model of 5'-

deoxyadenosine (an adenosine analogue) was also obtained. This structure revealed that it binds in the same manner as 5'-FDA and the modification at the 5' position is not essential for the conformation adopted by the ribose ring. The structure of β -D-eythrofuransyl adenosine (a novel adenosine analogue prepared within the O'Hagan group) was obtained with aim to identify the halide binding site. No density was obtained which could be accounted for by a halide ion despite several attempts at various concentrations of halide in the co-crystallization condition.

The compounds 5'-fluoro-2', 5'-dideoxyadenosine (2'd-FDA) and 5'-chloro-2',5'-dideoxyadenosine (2'd-CIDA) had been previously identified as a substrates of FDAS. The X-ray co-crystal structure of 2'd-FDA was obtained. This showed significant structural difference (when compared with 5'-FDA), which can account for the substrate specificity. The structural changes are dictated by the accommodation of a bifurcated hydrogen bonding interaction between the 3'-hydroxyl to Asp-16 and a rearrangement in ring conformation. The torsion in the ring changes from a strained planar form (5'-FDA) to a puckered system (2'd-FDA). Interestingly all attempts to co-crystallize 3' deoxy compounds were unsuccessful. These compounds are not substrates of FDAS, but should bind to the protein. The 2'-hydroxyl group of the ligand is thought to be essential as part of the mechanism, forming a hydrogen bond interaction towards the halide ion.

Crystallization conditions for apo-FDAS were identified, which lead to solution of the apo-structure to 2.5 Å. It was expected that there may be some domain movement to allow access to the buried active site however the model revealed no major structural differences. A possible tunnel was identified, as a result of slight changes over several

residues, which leads to the surface of the protein and could be responsible for access of substrates to the active site.

Isothermal calorimetry was performed on selected ligands. The thermodynamic profiles for SAH, adenosine, 5'-FDA, SAM and 5'-CIDA were determined. The titrations of SAM and 5'-CIDA have a similar profile. The interaction of Ad, SAH and 5'-FDA profiles were similar to each other but clearly different to SAM and 5'-CIDA. Analysis of the differences observed between reactions carried out in different buffers indicated that this may be explained by ligand induced release of proton(s). Clearly all of the ligands appear to bind (in the titration) at less than a stoichiometric ratio. This can be corrected for by altering the protein concentration so that $n = \sim 1$. Adenosine is clearly strongly bound to the ligand ($K_d = 0.66$ nM), and this observation goes some way to explaining why the ligand is co-purified from the recombinant system. SAH ($K_d = 0.32$ nM) clearly is the tightest binding ligand in the study which is consistent with it being a good competitive inhibitor.

2.7 FUTURE WORK

It became clear that the crystallization was dependent on ligand being present within the crystallization conditions, until the point at which the 'apo' structure was solved. Time constraints, unfortunately, restricted the pursuit of the halide binding site. Mutational analysis was initiated during the investigation; to date several mutants have been prepared and are undergoing analysis (within the group). However conclusions cannot be drawn from the early results. ITC revealed interesting differences in interactions of some ligands during binding which cannot be accounted for at this stage and requires further investigation. An analysis using FDAS mutants may also reveal some properties of ligand binding. The K_d values for 2'-deoxy and 3'-deoxy compounds should also be determined.

3

**An Investigation into Sso6206;
A Highly Conserved Protein in Archaea with Unknown Function.**

3.1 SUMMARY

Sso6206 is a small acidic protein of unknown function. The 10.4 kDa protein from *Sulfolobus solfataricus* (Sso), has been overexpressed, purified and crystallized. The protein crystallizes in the space group $P6_{1/5}22$ with cell parameters $a = b = 157.8 \text{ \AA}$, $c = 307.3 \text{ \AA}$. The crystals are hexagonal bipyramids and a data set to 2.4 \AA resolution has been collected. Molecular replacement cannot be attempted since no convincing model can be identified. The selenomethionine (SeMet) derivative was prepared which crystallizes with the space group $P2_12_12_1$ with cell parameters $a = 119.2 \text{ \AA}$, $b = 124.1 \text{ \AA}$, $c = 128.2 \text{ \AA}$. This suggests a large oligomer. A data set was collected to $\sim 5 \text{ \AA}$ however the selenium sites could not be located. Heavy atom soaks were attempted by soaking various solutions containing heavy atoms into native crystals. The crystals were screened for anomalous diffraction however no signal was observed. Finally sulfur phasing was attempted using the native crystals. Data sets were collected to $\sim 2.7 \text{ \AA}$ resolution at a wavelength suitable for sulfur phasing. Although an anomalous signal was observed to $\sim 4 \text{ \AA}$, attempts to find the sulfur sites were unsuccessful.

3.2 EXPERIMENTAL INTRODUCTION

3.2.1 Sso 6206

Sso6206 is a small acidic protein (Mr 10483 Da, pI 4.5) encoded in the genome of the crenarchaeon *Sulfolobus solfataricus* (She et al., 2001). The protein has no known function, but is nearly ubiquitous in the archaea, including crenarchaea and euryarchaea, hyperthermophiles, halophiles, psychrophiles and methanogens. In addition homologues have been annotated in two bacterial species, *Shewanella frigidimarina*, a marine member of the gamma proteobacteria (Bowman *et al.*, 1997) and *Desulfotalea psychrophila*, a sulfate-reducing delta proteobacterium from arctic sediments (Knoblauch *et al.*, 1999).



Figure 3.1 – Sequence alignment of Sso6206 and archaeal and bacterial homologues. The four conserved acidic residues are highlighted in red, and the conserved histidine in blue. Sso6206 from *Sulfolobus solfataricus* (Sso); Pto, *Picrophilus torridus*; Mka, *Methanopyrus kandleri*; Pae, *Pyrobaculum aerophilum*; Pfu, *Pyrococcus furiosus*; Mth, *Methanothermobacter thermautotrophicum*; Mma, *Methanococcus maripaludis*; Dps, *Desulfotalea psychrophila*; Sfr, *Shewanella frigidimarina*.

The limited distribution in bacteria suggests these two species have obtained the gene by lateral gene transfer from an archaeon (Nelson *et al.*, 1999). The amino acid sequence of the protein is highly conserved (Figure 3.1), with greater than 40 % identity observed between archaeal family members. Perhaps significantly there are four conserved acidic residues and a histidine (coloured red and blue, respectively, in Figure 3.1), which are often ligands for bound metal ions. The strong conservation hints at an important role for this protein in the archaea that is currently not understood, making it an interesting target for structural studies. The protein belongs to pfam2680 a family which has yet to be characterised. There are no protein structures deposited in the PDB with significant sequence homology as determined by BLAST (Altschul *et al.* 1997), indicating that the protein is likely to have a novel fold.

Sso6206 can be assessed for suitability towards high throughput using some *in silico* techniques. Entering Sso6206 protein sequence into the regional order neural network (RONN) application indicated that there are no regions of disorder (Figure 3.2).

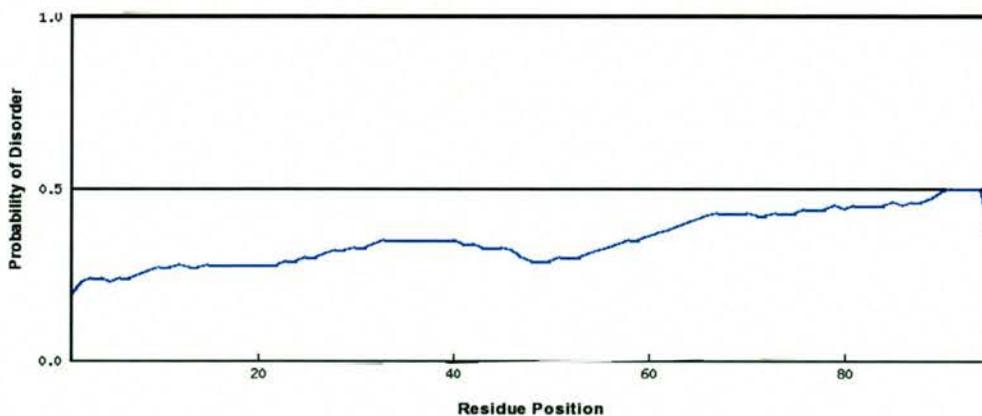


Figure 3.2 – RONN plot of probability of disorder against residue number. The threshold value (0.5) for disorder is indicated. Residues with a probability above 0.5 are predicted to be disordered, while those below are predicted to be ordered.

This is significant because disordered regions are typically unfavourable for crystallization (Yang *et al.*, 2005). Identification of potential disorder by such methods can reduce the efforts spent on proteins which may be insoluble or otherwise uncrystallizable (Oldfield *et al.*, 2005).

In a study of the *Thermotoga maritima* proteome, three protein clusters are observed when pI is plotted against the grand average of hydropathy (GRAVY) index. The cluster A is occupied by proteins which have a higher chance of producing diffraction quality crystals than for proteins which fall into cluster B or C (Canaves *et al.*, 2004). Figure 3.3 shows the cluster analysis plot and the predicted position of Sso6206. Since it is predicted to fall into cluster A, it can be inferred that Sso6206 stands a good chance of successful crystal structure determination.

The aim of this investigation was to determine the X-ray crystal structure of Sso6206.

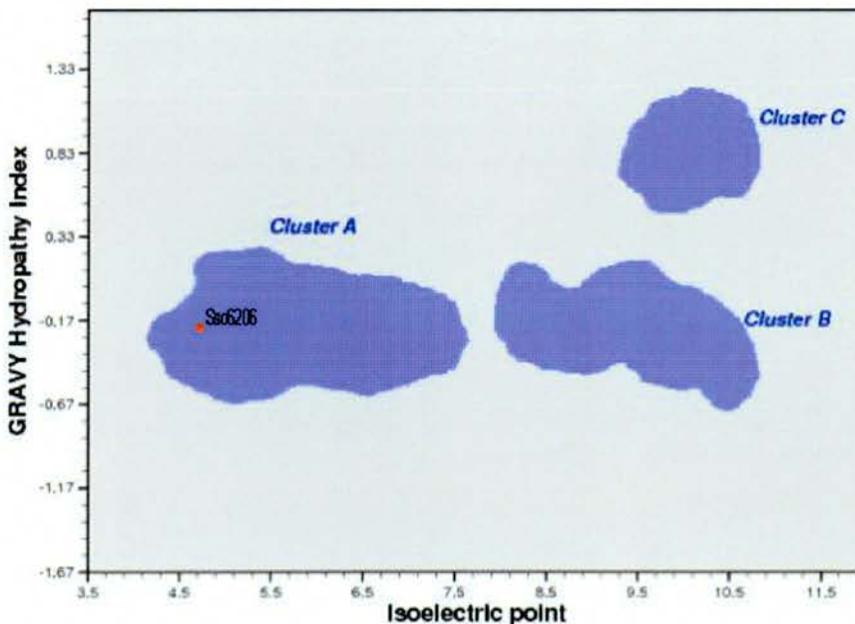


Figure 3.3 – Cluster analysis of Sso6206 indicating that the protein lies in cluster A since predicted pI = 4.632 and GRAVY = 0.194.

3.3 MATERIALS AND METHODS

3.3.1 Cloning of 6206 gene

The Sso6206 gene was amplified using the genomic DNA of *Sulfolobus solfataricus* P2 and the oligonucleotide primers containing *NcoI* and *BamHI* sites shown in Figure 3.4. The amplified DNA fragment was digested with *NcoI* / *BamHI* restriction enzymes and then ligated into a *NcoI* / *BamHI* linearised pEHISTEV vector (Naismith and Liu, 2002) such that six histidine residues and a tobacco etch virus (TEV) protease site was added to the N-terminus of the protein in order to facilitate automated protein purification.

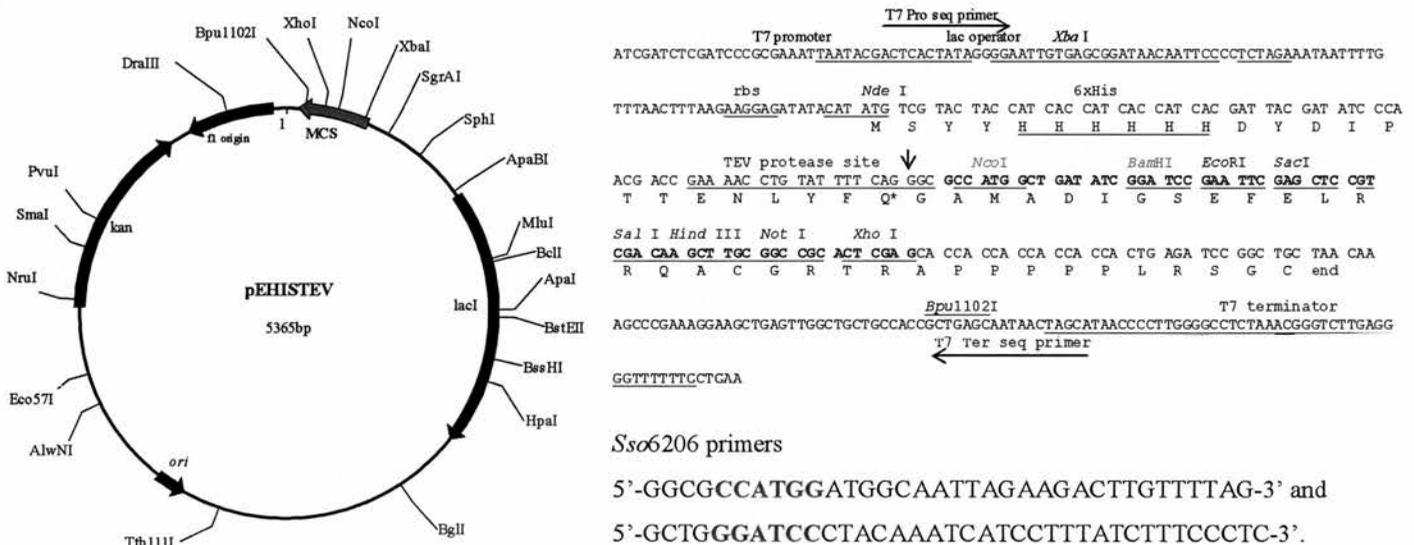


Figure 3.4 – Vector map of pEHISTEV and Sso6206 primers

The construct was amplified by use of RapidTrans™ TAM1 chemically competent cells (Active Motif), purification of the plasmid DNA was from overnight cultures (10 mL) as directed in the protocol of the Qiagen miniprep kit. The DNA was

eluted into 40 μL $d\text{H}_2\text{O}$. Purity was determined by calculation of absorbance ratio at $\lambda = 260$ to $\lambda = 280$ nm. The concentration was determined from the absorbance of $\lambda = 260$ nm. The construct was stored at 253 K until required. To confirm correct insertion of the gene, DNA sequencing was carried out at the University of Dundee (www.dnaseq.ac.uk), using T7 forward and reverse primers.

3.3.2 Typical overexpression and purification of Sso6206

The pEHISTEV-Sso6206 construct was transformed into *Escherichia coli* BL21 (DE3) cells (Novagen) and four 500 mL cultures were grown in Luria-Bertani medium (Formedia) supplemented with 30 $\mu\text{g mL}^{-1}$ kanamycin at 310 K until A_{600} reached 0.6. The cultured cells were induced with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) and incubated for 4 h at 298 K. The cells were harvested by centrifugation (5000 rpm, 15 min, 277 K), washed with PBS and re-spun. The cells were stored at 193 K until required. For purification, the cell pellet was resuspended in lysis buffer [25 mM Tris-HCl pH 7.75, 0.6 M NaCl, 30 mM imidazole 20 μM lysozyme and 20 $\mu\text{g mL}^{-1}$ DNase I] and lysed by a Constant Systems cell disruptor at 298 K. The crude lysate was centrifuged (15000 rpm, 25 min, 277 K) and the supernatant fraction was passed through a 0.22 μm PES syringe filter (Millipore). The supernatant was passed over a Ni-NTA affinity column and the protein eluted in a single step with elution buffer [25 mM Tris-HCl pH 7.75, 0.6 M NaCl, 250 mM imidazole]. The eluted protein was then incubated with 1 mL of 1 mg mL^{-1} tobacco etch virus protease to remove the polyhistidine tag. The protein was then applied to a Sephacryl S-200 size exclusion chromatography column and fractions were assayed by SDS page gel electrophoresis.

The fractions containing pure Sso6206 were combined and dialysed (1000 fold) against buffer [25 mM Tris-HCl pH 7.75, 50 mM NaCl]. In total it was estimated 20 mg pure Sso6206 was obtained from 2 L of cell culture. The protein was concentrated to 6 mg mL⁻¹ according to a Bradford assay curve. The identity and integrity of the protein was confirmed by mass spectrometry (University of St. Andrews).

3.3.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was performed to verify protein folding, prior to crystallization (Kelly *et al.*, 2005) (Kelly and Price, 2000). The CD scans were carried out at the University of St Andrews using 1 mg mL⁻¹ protein in 25 mM phosphate buffer pH 7.75 using a JASCO J-810 CD spectropolarimeter across a 1 mm cell, and processed using JASCO Spectra Manager™.

3.3.4 Crystallization of native Sso6206

Initial conditions were obtained from a sitting drop vapour diffusion screen of commercial sparse matrix crystallisation conditions. Three protein concentrations were used, 6, 3 and 1 mg mL⁻¹. Drops of 0.1 µL of protein and 0.1 µL precipitant was prepared using a nano-drop crystallization robot (Cartesian HONEYBEE™) as apart of the Hamilton-Thermo Rhombix™ system. Optimization of the initial twelve most promising hits, including at least one at each concentration, were performed to confirm crystals were protein. The largest protein crystals were obtained from the hanging drop vapor-diffusion method. Improved crystals could be obtained from a number of conditions, however the best crystals, judged by size and regular shape, were obtained

from variation of Wizard™ I (Emerald Biosystems) conditions 31 [20 % (w/v) PEG-8000, 0.1 M phosphate-citrate pH 4.2, 0.2 M NaCl] and 36 [1.0 M sodium citrate, 0.1 M imidazole pH 8.0]. The crystals obtained from both conditions were protein with hexagonal bipyramidal morphology (0.4 x 0.4 x 0.4 mm). The crystals appear as spherical precipitate within two days and gradually take more distinct edges over the course of a month. Diffraction analysis showed that crystals obtained from a precipitant of 0.6 M sodium citrate, 0.1 mM imidazole pH 7.75 with a protein concentration of 5 mg mL⁻¹ were optimal (Table 3.1).

		0.1 M Imidazole				
		7.25	7.50	7.75	8.00	8.25
Sodium Citrate	0.4 M					
	0.6 M					
	0.8 M					
	1.0 M					
	1.2 M					
	1.4 M					
	1.6 M					
	1.8 M					

Table 3.1 – Sso6206 crystallization grid screen for crystal growth, where optimization of sodium citrate concentration is varied against pH of 0.1 M imidazole

3.3.5 Typical overexpression and purification of SeMet 6206

Since molecular replacement is not possible (McEwan *et al.*, 2006), the selenomethionine (SeMet) variant of the protein was obtained. The SeMet protein was expressed in PASM-5052 auto inducing, methionine substitution medium (Studier, 2005). The PASM-5052 medium consists of 50 mM Na₂HPO₄, 50 mM KH₂PO₄, 25 mM (NH₄)₂SO₄, 2 mM Mg₂SO₄, 0.5 % glycerol, 0.05 % glucose, 0.2 % lactose, and 200 µg

mL⁻¹ of each of 17 amino acids (not C, Y or M), 10 µg mL⁻¹ methionine, 125 µg mL⁻¹ SeMet and 100 nM vitamin B₁₂.

BL21 (DE3) cells transformed with the pEHISTEV-Sso6206 construct were grown in LB-broth overnight (10 mL, 310 K) to produce a stock culture. The cells were spun down (5000 rpm, 15 min, 277 K), washed with PBS to remove excess LB and re-spun, these were suspended using PASM-5052 media. The resuspended culture was transferred to 4 x 2 L baffled flasks containing 500 mL PASM-5052 media. The cultures were incubated with shaking (300 rpm, 293 K) till an OD₆₀₀ of ~3.0 to ~5.0 was reached. This was reached within 36 to 48 h. The cells were harvested by centrifugation (5000 rpm, 15 min, 277 K), washed with PBS and re-spun. The cells were stored at 193 K until required. For purification, the cell pellet was resuspended in lysis buffer [25 mM Tris-HCl pH 7.75, 0.6 M NaCl, 30 mM imidazole 20 µM lysozyme and 20 µg mL⁻¹ DNase I and 0.7 mM DTT] and lysed by a Constant Systems cell disruptor at 298 K. The protein was purified as for native protein. The reducing agent DTT (0.7 mM) was added to all buffers to prevent oxidation of the SeMet. The fractions containing pure Sso6206 were combined. In total it was estimated 4 mg pure SeMet Sso6206 was obtained from 2 L of cell culture. The protein was concentrated to 6 mg mL⁻¹ according to a Bradford assay curve. The identity and integrity of the protein was confirmed by mass spectrometry (University of St. Andrews). Mass spectrometry confirmed full incorporation of selenium and did not indicate that any oxidation has occurred.

3.3.6 Crystallization of SeMet derivative of Sso6206

SeMet Sso6206 did not to crystallize in the optimized screen (Table 3.1). SeMet protein which was not treated with reducing agent also did not yield crystals. Sparse

matrix screening was also carried out on the SeMet protein at different concentrations (6, 5, 4 and 2 mg mL⁻¹) in commercially available screens (as detailed in appendix A.I) including Crystal screen 1 and 2 (Hampton), Wizard I and II (Emerald Biosciences), the ‘Pegs’ (Nextal) and Premier (In-house). The screening was performed at the SSPF (University of St. Andrews) using 0.1 µL of protein and 0.1 µL precipitant using a nono-drop crystallization robot (Cartesian HONEYBEE™) as part of the Hamilton-Thermo Rhombix™ system. Some hits were identified however none were improved with optimization.

Small crystalline precipitate was obtained using a re-optimized version of the original screen, however this did not appear until after two months of setting up the drops and yielded poorly diffracting crystals. The crystals obtained grew to 0.1 x 0.2 mm and were different to those obtained for native protein. The screen was extended away from the precipitation (Table 3.2) so as to provide conditions which may yield diffracting crystals. Replacing the buffer (with sodium cacodylate or Bis-Tris), so that the pH could also be extended, did not yield crystals.

		<i>0.1 M Imidazole</i>				
		5.8	6.0	6.2	6.4	6.6
Sodium Citrate	100 mM					
	200 mM					
	300 mM					
	400 mM					
	500 mM					
	600 mM					
	700 mM					
	800 mM					

Table 3.2 – *SeMet Sso6206 crystallization grid screen for crystal growth, where optimization of sodium citrate concentration is varied against pH of 0.1 M imidazole*

3.3.7 Heavy atom replacement

Since frozen crystals are non-isomorphous, we require to perform multi-wavelength anomalous diffraction (MAD) or single-wavelength (SAD) techniques. It was decided to insert heavy atoms by soaking and co-crystallization into the native protein to obtain phases for the native data. Several heavy atom salts were used in an attempt to obtain derivative crystals, including mersalyl acid, 4-(hydroxymercuri) benzoic acid, thallium acetate, potassium hexachloroplatinate (IV), europium (III) nitrate hexahydrate, potassium tetra-bromo-aurate (III), platinum (II) ethylenediamine dichloride, ammonium hexabromoosmate (IV), uranium (VI) dinitrate oxide hexahydrate, samarium (III) acetate hydrate and platinum potassium iodide. Co-crystallization using the grid screen for native protein was prepared containing 2 mM and 20 mM of the heavy atoms listed. Drops were set up as for native protein however crystallization did not occur. The grid screens were also used as soaking solutions for native crystals, these were diluted by half with buffer [25 mM Tris-HCl pH 7.75, 50 mM NaCl] to correct for concentration. The crystals were transferred drop to drop into the soaking solution and assessed for stability. Crystals which survived soaking using low concentrations (1 mM) were also soaked using the higher concentration screen (10 mM). The crystals were screened for diffraction and if possible screened for anomalous diffraction.

3.3.8 Native crystal data collection

Native data sets were recorded using in-house, and synchrotron X-ray sources. The initial 2.7 Å dataset was collected in-house at 130 K from a single crystal mounted

on a loop. The crystals were soaked in a 15 % ethylene glycol stabilization buffer before transferring to cryogenic conditions and data collection. Data were collected as 20 min 0.5° oscillations using a Rigaku HTC image plate detector from a rotating anode x-ray source and with Osmic mirrors. The crystal to detector distance was 220 mm. The initial indexing showed the crystals to be primitive hexagonal, with unit-cell parameters of $a = b = 158.1$, $c = 308.3$ Å; $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$. A second native dataset was collected at Daresbury which showed a similar space group and resolution. A third dataset was collected at the European Scientific Research Facility (ESRF) Grenoble of the same space group and cell dimensions with improved resolution

3.3.9 SeMet derivative crystal data collection

SeMet crystals were frozen in-house at 130 K from a single crystal mounted on a loop. The crystal was soaked in 20 % glycerol in stabilization buffer before transferring to cryogenic conditions. Screening for diffraction was performed in-house using a Rigaku HTC image plate detector from a rotating anode X-ray source (Cu $K\alpha$). Diffracting crystals were transferred to the ESRF (Grenoble). Fluorescence scans were performed using the beamline (ID23-1) mounted fluorescence detector (Rontec, *Si* drift chamber detector). SeMet data were recorded using synchrotron radiation (ID23 – ESRF, Grenoble) using radiation at $\lambda = 0.979$ Å, corresponding to the peak for selenium. The indexing showed the crystals to be primitive orthorhombic, with unit cell parameters of $a = 119.24$, $b = 124.08$, $c = 128.25$ Å; $\alpha = \beta = \gamma = 90^\circ$.

3.3.10 Heavy atom derivative data collection

For the purpose of screening for an anomalous signal several data sets were collected using in-house, and synchrotron (SRS - Daresbury) X-ray sources. The data were collected at 130 K from a single crystal mounted on a loop. The crystal were removed from solution, containing heavy atom, and back soaked to remove excess heavy atom. The crystal was then soaked in 15 % ethylene glycol in stabilization buffer before transferring to cryogenic conditions and data collection. Crystals were assessed for diffraction and a dataset collected from those which survived.

3.3.11 Sulfur data collection

A sulfur dataset was collected at the ESRF Grenoble, using $\lambda = 1.77 \text{ \AA}$. A single 2.7 \AA resolution dataset was collected at 130 K from a single crystal mounted on a loop. The crystal was subject to a soaking in 15 % ethylene glycol in stabilization buffer before transferring to cryogenic conditions and data collection. To ensure the dataset was highly redundant the data were measured over 600° , at two orientations of κ to increase redundancy.

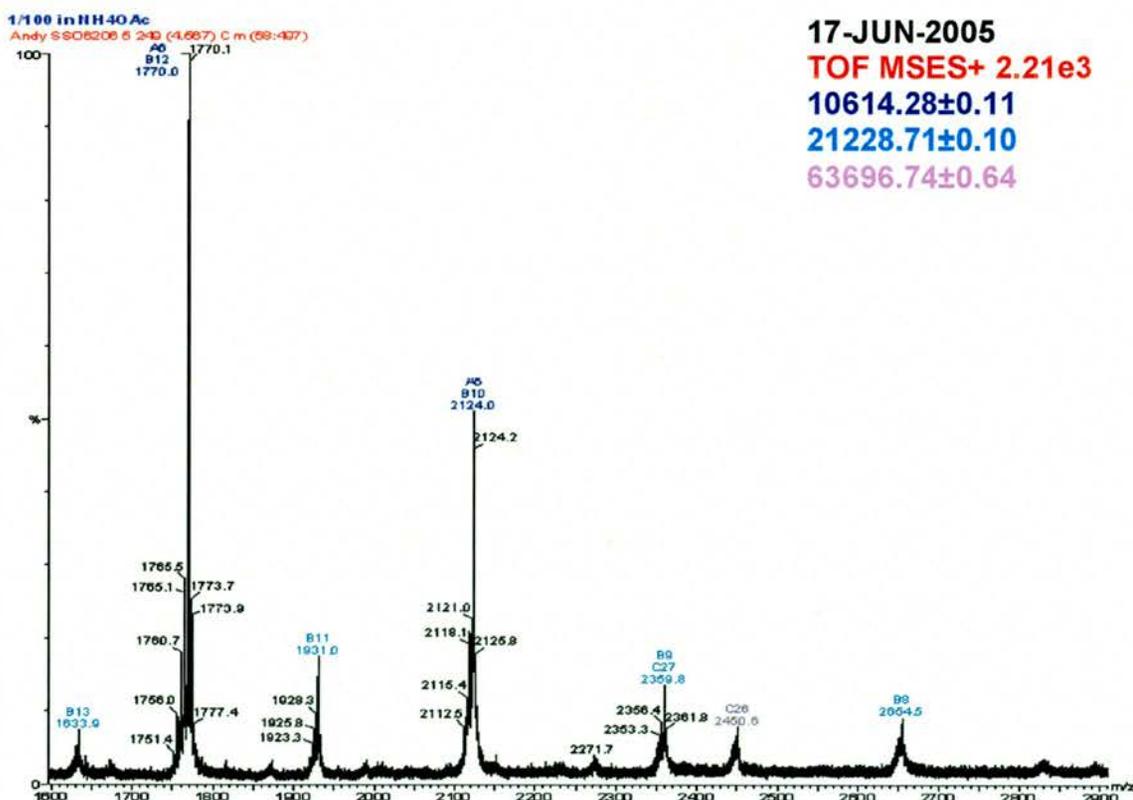
3.3.12 Data processing

Native data was integrated using MOSFLM (Leslie, 1992) and merged with SCALA (Evans, 1997) as implemented in the CCP4i suite (Potterton *et al.*, 2003). The program POINTLESS (Evans, 2004) was used to confirm point group and symmetry.

3.4 RESULTS AND DISCUSSION

3.4.1 Initial purification and crystallisation

The protein Sso6206 was purified to homogeneity proven by SDS-PAGE gel and mass spectrometry. During purification, fractions containing Sso6206 could only be assessed for purity using Coomassie stained SDS-PAGE, since the protein does not have an A_{280} (as predicted by ProtParam - ExPASy). The concentration of the pure protein was therefore determined by Bradford assay. Interestingly Sso6206 elutes very close to the void volume of the S-200 gel filtration as a broad peak and certainly very far from the expected elution point for a 10 kDa protein. On the Sephacryl S-200 column globular proteins in excess of 200 kDa elute in the void volume.



17-JUN-2005
TOF MSES+ 2.21e3
10614.28±0.11
21228.71±0.10
63696.74±0.64

Figure 3.5 – Soft ionization mass spectrometry (Sso6206) indicating presence of monomer, dimer and hexameric peaks.

Molecular weight calibration of the column suggest Sso6206 is over 100 kDa, more accurate analysis is not possible due to the broad nature of the peak. This suggests that Sso6206 contains in excess of 10 monomers per protomer (100 kDa). Soft ionization mass spectrometry shows peaks which correspond to multimers (Figure 3.5) (up to hexamer can be distinguished) as well as the peak corresponding to monomeric protein.

The optimum concentration for crystallization was predicted using PCT (Hampton) as 5 mg mL⁻¹. Sparse matrix crystal trails were prepared through automated crystallization at the SSPF and further trials were set up by hand for comparison purposes. From the several hits obtained (i.e. Figure 3.6) only two were able to be optimized and one gave diffracting crystals.

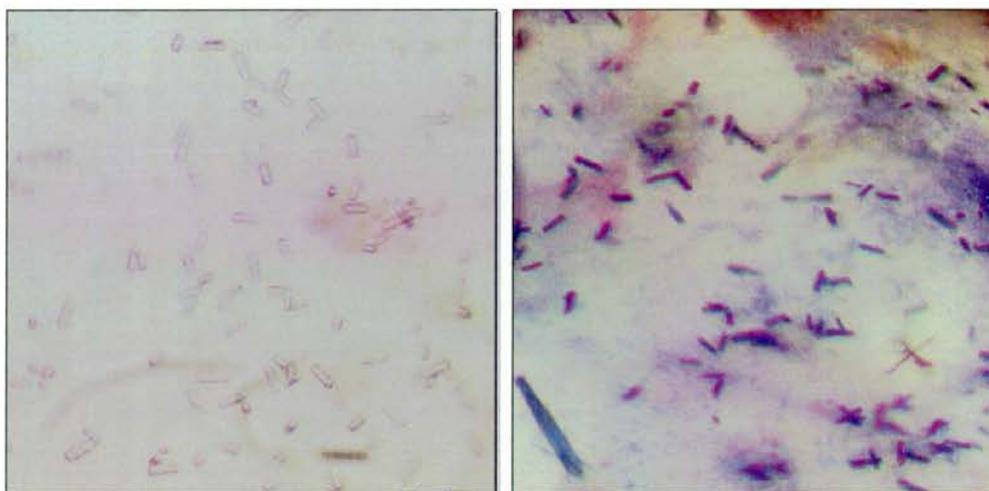


Figure 3.6 – A) Initial native crystals obtained in Wizard™ I # 36, and B) the same crystals dyed with Izit™ solution (Hampton research).

The final diffracting crystals were obtained from 0.6 M sodium citrate, 0.1 mM imidazole pH 7.75 (Figure 3.7). These were subsequently used for native (Section 3.4.2), heavy atom (Section 3.4.3) and sulfur SAD (Section 3.4.6) experiments.

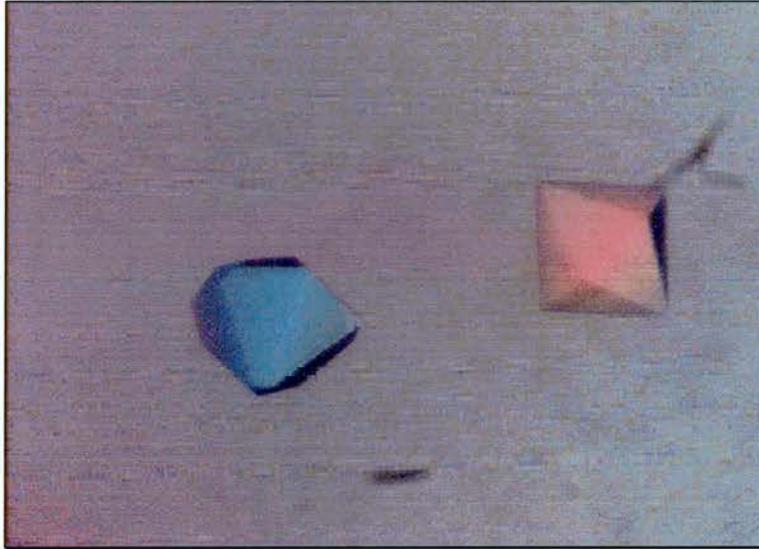


Figure 3.7 – Picture of a typical Sso6206 crystal grown in optimized conditions. The crystals appear coloured under a polarizing microscope filter (used for measuring birefringence).

3.4.2 Native Sso6206 data

The crystals obtained were assessed for their diffraction using an in-house X-ray source at 130 K. Several cryo-protectant solutions were prepared into which single crystals were transferred (Garman, 1999). The crystals which survived were flash frozen in a stream of nitrogen. The crystals diffracted reasonably well using 15 % to 20 % ethylene glycol and an initial data set was collected to 2.8 Å. Two further datasets were obtained with improving resolution at the SRS (Daresbury) and the ESRF (Grenoble). A typical image of the diffraction is shown in Figure 3.8.

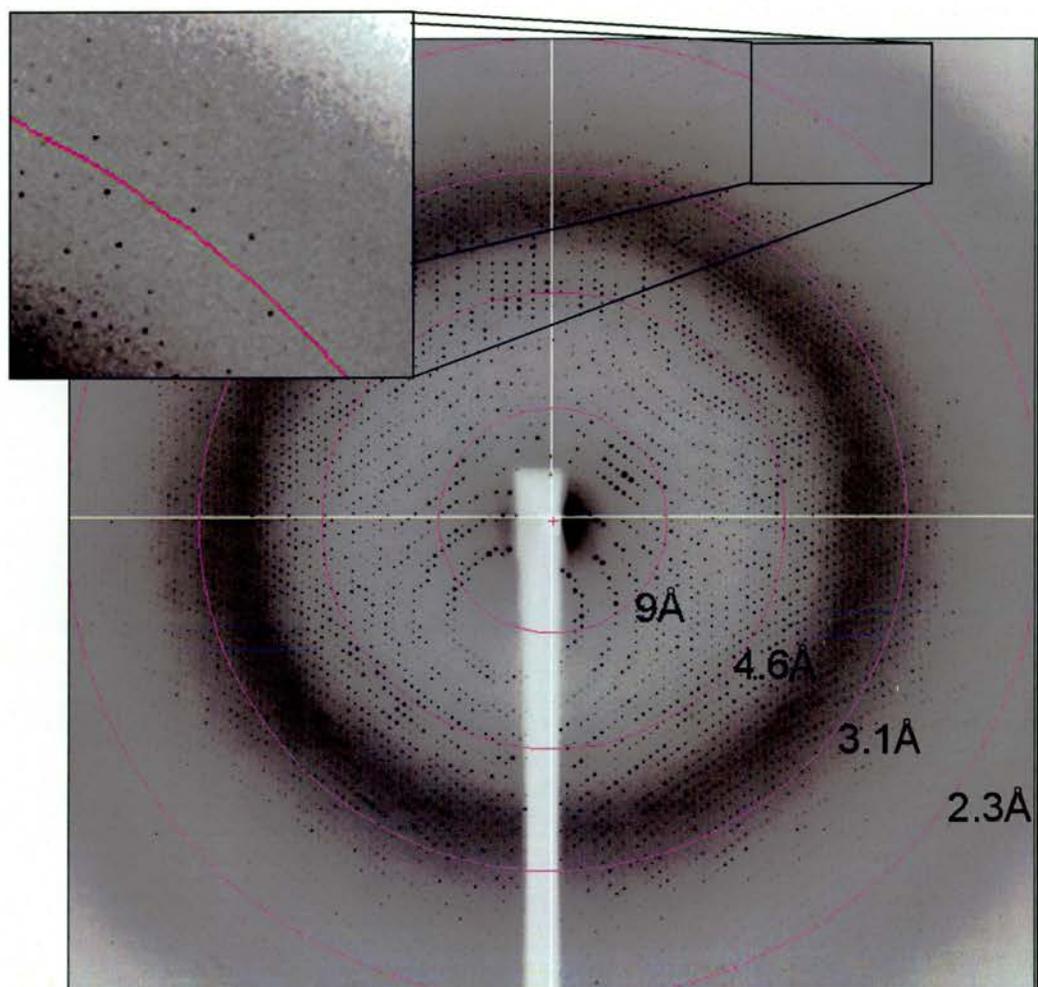


Figure 3.8 – Diffraction pattern of Sso6206 collected at the ESRF (Grenoble). The resolution at the edge of the detector is 2.3 Å.

Merging of the data indicates that the spacegroup belongs to the higher symmetry 6/mmm Laue class. Analysis of over a hundred 00l axial reflections shows a clear $l = 6n$ condition, indicating the spacegroup is $P6_122$ or its polar enantiomorph $P6_522$. The cell dimensions also suggest that each ASU contains between 15 (V_m 3.50 Da Å⁻³, solvent content 65 %) and 30 (V_m 1.77 Da Å⁻³, solvent content 30 %) monomers. This suggests

that the protein may exist as a large multimeric protomer. A summary of the data collections are shown in Table 3.3.

	ESRF (ID 29)	SRS (14-2)	In-house
Wavelength (Å)	0.98	0.87	1.54
Resolution (Å)	76.47 – 2.40 (2.53 – 2.40)	69.0 – 2.5 (2.6 – 2.5)	45.13 – 2.80 (2.95 – 2.80)
Space group	P6 _{1/5} 22	P6 _{1/5} 22	P6 _{1/5} 22
Temperature (K)	130	130	130
Detector	MAR 225 CCD	ASDC Q4 CCD	Rigaku HTC IP
Unit-cell parameters (Å)	a = b = 157.8, c = 307.3 $\alpha = \beta = 90^\circ \gamma = 120^\circ$	a = b = 155.2, c = 302.7 $\alpha = \beta = 90^\circ \gamma = 120^\circ$	a = b = 158.1, c = 308.3 $\alpha = \beta = 90^\circ \gamma = 120^\circ$
Solvent (%)	49.10	46.25	49.12
Unique reflections	89232 (12793)	59867 (8561)	56502 (8126)
$I/\sigma(I)$	26.6 (8.0)	28.5 (6.6)	19.7 (6.6)
Average redundancy	15.8 (16.1)	12.6 (12.9)	8.4 (8.5)
Data completeness (%)	100.0 (100.0)	100.0 (100.0)	99.8 (100.0)
Rmerge †	0.085 (0.356)	0.084 (0.367)	0.098 (0.303)

Table 3.3 – Sso6206 crystal data and data collection statistics. Values in parentheses refer to the highest resolution shell. († $R_{\text{merge}} = \frac{\sum \sum I(h)_i - I(h)}{\sum I(h)_i}$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations). Solvent calculated for 22 monomers per ASU.

3.4.3 Phasing using heavy atom soaks

Co-crystallization (in the previous screen) using 2 and 20 mM (1 and 10 mM) heavy atom solution were attempted after pre-incubation with native protein. This

typically resulted in amorphous precipitation or no precipitation at all. It was therefore decided to concentrate on soaking experiments. It has been shown in other studies that changing the crystal buffer environment from high salt and low in PEG to one which may be low in salt and high in PEG may permit a more optimal environment for heavy atom soaking (Sun *et al.*, 2002). Any significant alteration of the buffer properties however resulted in clear deterioration of Sso6206 crystal integrity. Despite this, attempts were made to directly incorporate heavy atoms by direct soaking methods (Garman and Murray, 2003).

	Buffer	Concentration	Crystal survived	Data screening
1	stabilization buffer	n/a	+ 4 h	n/a
2	mersalyl acid	2 mM	short soak	poor diffraction
3	4-(hydroxymercuri) benzoic acid	2 mM	no	no
4	thallium acetate	20 mM	+ 4 h	yes
5	potassium hexachloroplatinate (IV)	20 mM	yes	poor diffraction
6	europium (III) nitrate hexahydrate	2 mM	+ 4 h	yes
7	potassium tetra-bromo-aurate (III)	20 mM	no	no
8	platinum (II) ethylenediamine dichloride	20 mM	+ 4 h	yes
9	ammonium hexabromoosmate (IV)	2 mM	short soak	yes
10	uranium (VI) dinitrate oxide hexahydrate	20 mM	no	no
11	samarium (III) acetate hydrate	20 mM	+ 4 h	yes
12	platinum potassium iodide	2 mM	no	no

Table 3.4 – Summary of heavy atom soaking experiments

Crystals were soaked using solutions of heavy atoms as detailed in Table 3.4. Crystals which survived soaking were removed from the soaking solution and back

soaked using stabilization buffer to remove any excess heavy atoms. The crystals were cryo-protected before freezing, as for native protein.

	SRS 14.1 - HA4	SRS - HA6	SRS - HA9s	Inhouse - HA11	Inhouse - HA8
Wavelength (Å)	1.488	1.488	1.488	1.54	1.54
Resolution (Å)	70.1 – 2.7 (2.85 – 2.7)	81.0 – 2.9 (3.1 – 2.9)	81 – 3.0 (3.1 – 3.0)	49.0 – 2.9 (3.1 – 2.9)	70 – 2.9 (3.0 – 2.9)
Space group	P6 _{1/5} 22	P6 _{1/5} 22	P6 _{1/5} 22	P6 _{1/5} 22	P6 _{1/5} 22
Temperature (K)	130	130	130	130	130
Detector	ASDC Q4 CCD	ASDC Q4 CCD	ASDC Q4 CCD	Rigaku HTC IP	Rigaku HTC IP
Unit-cell parameters (Å)	a = b = 157.2, c = 307.7	a = b = 156.3, c = 306.7	a = b = 157.3 c = 307.3	a = b = 158.0, c = 308.8	a = b = 157.9, c = 308.6
Solvent (%)	48.2	47.8	48.2	50.0	48.9
Unique reflections	62344 (8911)	49463 (6812)	45802 (6542)	51057 (7286)	51138 (7314)
<i>I</i> / $\sigma(I)$	24.3 (4.4)	13.7 (2.0)	24.8 (7.8)	18.1 (3.8)	13.1 (3.1)
Average redundancy	11.8 (11.8)	7.0 (6.4)	13.4 (12.9)	14.5 (14.6)	10.4 (10.6)
Data completeness (%)	100.0 (100.0)	99.4 (96.1)	100.0 (100.0)	99.7 (99.4)	100.0 (100.0)
<i>R</i> _{merge} †	0.09 (0.50)	0.10 (0.84)	0.09 (0.32)	0.15 (0.77)	0.19 (0.62)
Anomalous					
Completeness	100.0 (100.0)	99.4 (96.1)	100.0 (100.0)	99.9 (99.7)	100.0 (100.0)
Multiplicity	6.2 (6.1)	3.7 (3.3)	7.1 (6.7)	7.7 (7.7)	5.4 (5.4)
CC-Anom	-0.15 (-0.02)	-0.11 (0.01)	-0.07 (0.04)	-0.05 (-0.4)	0.06 (0.01)
Mid slope	1.00	0.98	1.00	0.93	1.19

Table 3.5 – Sso6206 crystal data and data collection statistics for heavy atom screening. Values in parentheses refer to the highest resolution shell. († $R_{\text{merge}} = \frac{\sum \sum I(h)_i - (I(h)) / \sum I(h)_i}{\sum I(h)_i}$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations). Solvent calculated for 22 monomers per ASU.

Table 3.6 shows a typical example (crystal – HA8) of the analysis of anomalous statistics as output from SCALA, showing there is little or no anomalous data.

>HA8												
N	1/resol^2	dmax	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CC_Imean	N_Imean
\$\$												
1	0.0119	9.17	0.156	1128	0.099	482	1.280	1121	1.191	472	0.997	1610
2	0.0238	6.48	0.098	2388	0.078	621	1.077	2368	1.074	613	0.996	3009
3	0.0357	5.29	0.046	3223	-0.056	630	1.043	3201	0.968	623	0.991	3853
4	0.0476	4.59	0.052	3872	0.037	636	1.042	3849	0.998	626	0.992	4508
5	0.0595	4.10	0.145	4442	0.103	628	1.121	4417	1.049	622	0.992	5070
6	0.0713	3.74	0.069	4934	0.094	630	1.097	4919	1.004	619	0.974	5564
7	0.0832	3.47	0.065	5424	0.044	632	1.035	5397	0.905	623	0.966	6056
8	0.0951	3.24	-0.010	5940	0.012	640	0.998	5824	0.995	627	0.925	6495
9	0.1070	3.06	-0.040	6256	0.014	641	0.984	6238	0.980	622	0.853	6897
10	0.1189	2.90	-0.015	6611	-0.023	644	0.999	6573	0.968	609	0.710	7255
\$\$												
Overall			0.064	44127	0.050	6190	1.060	43907	1.031	6056	0.992	50317

Table 3.6 – Table of anomalous data as output from SCALA. The CC_anom is the anomalous correlation coefficient which indicates a weak anomalous signal.

3.4.4 Overexpression and crystallization of SeMet Sso6206.

The identity and integrity of the selenomethionine protein was confirmed by mass spectrometry (Figure 3.9) and SDS gel electrophoresis. This confirmed full incorporation of five selenomethionine's in place of methionine.

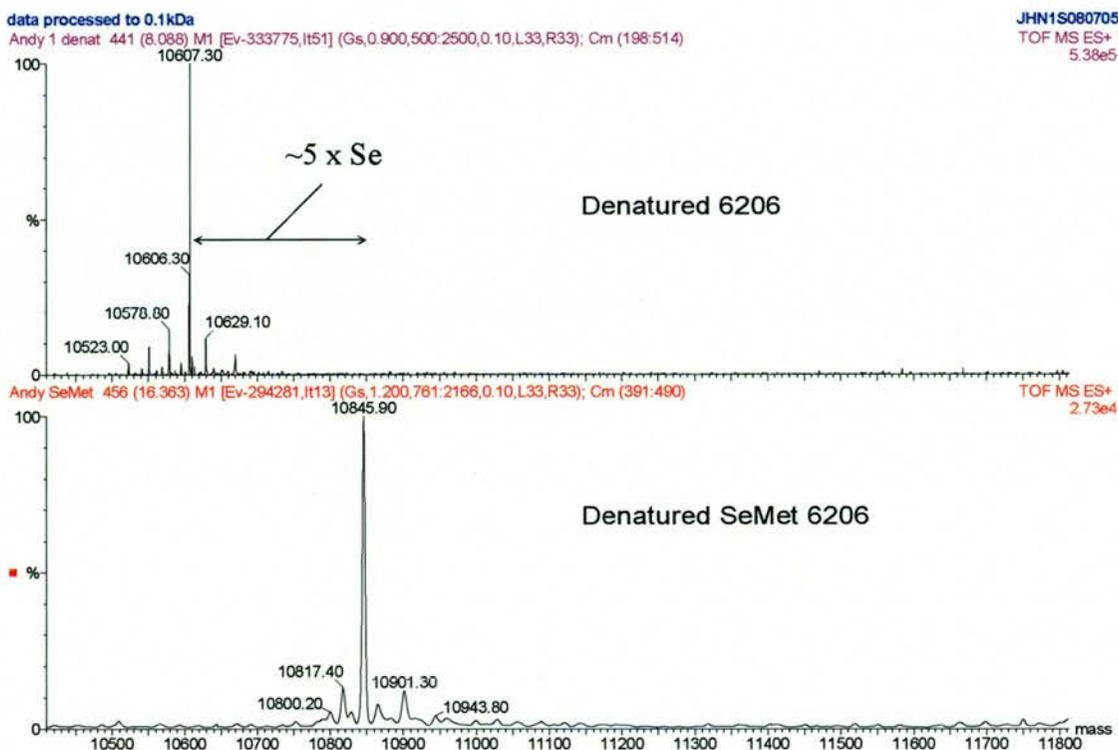


Figure 3.9 – Mass spectra of the SeMet derivative of Sso6206

Initial attempts to crystallize the selenomethionine variant of the protein either in the native conditions or in a re-screen of sparse matrix conditions was unsuccessful. Purification was repeated such that no reducing agent (DTT) was added to the buffers and crystal trials were repeated. Re-optimization of the initial conditions, which yielded diffracting crystals for the native protein, eventually produced crystals (Figure 3.10).

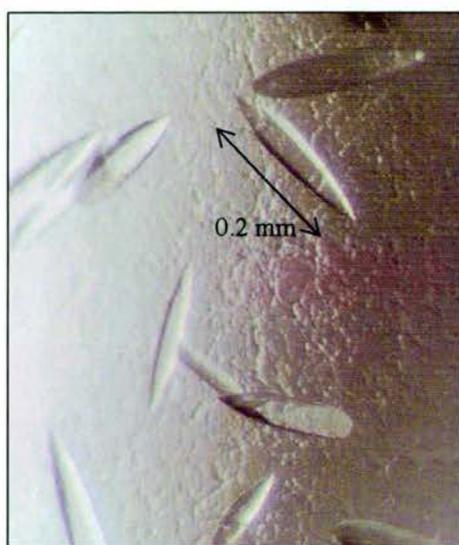


Figure 3.10 – SeMet crystals grown in 0.2 M sodium citrate and 0.1 M imidazole pH 6.0

The crystals obtained grew over two months, the morphology suggested that the crystals were different to those obtained from native protein. Circular dichroism spectrometry suggests the selenomethionine protein is not significantly different and the protein has a similar profile on size exclusion chromatography (Figure 3.11).

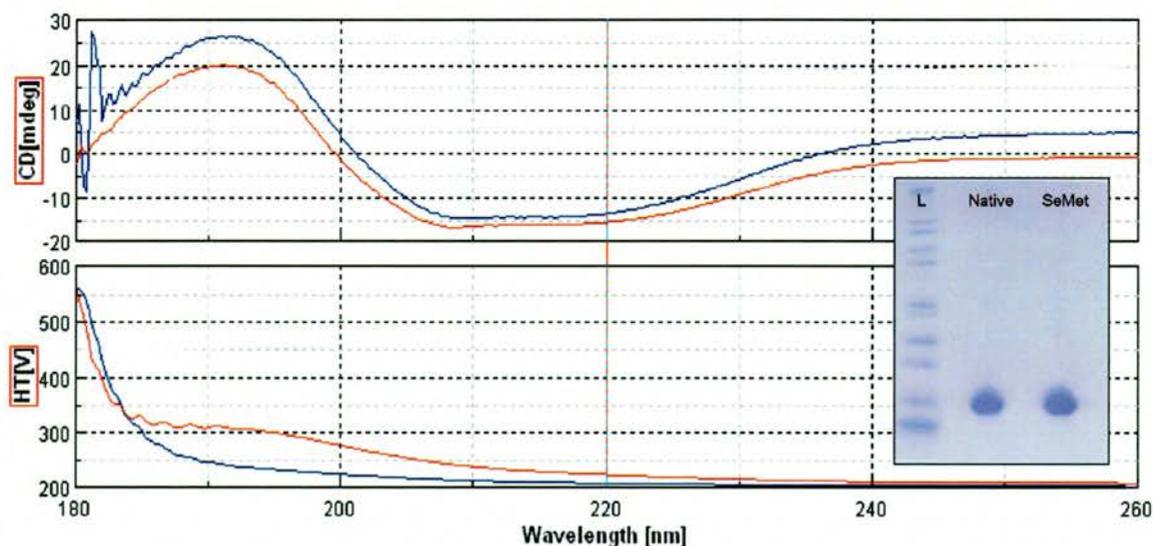


Figure 3.11 – Circular dichroism and SDS gel of the SeMet derivative of Sso6206 (red) compared with native Sso6206 (blue).

3.4.5 X-ray data collection of SeMet Sso6206

Initial diffraction using an in-house X-ray source indicated that the crystals did diffract however only to $\sim 8 \text{ \AA}$ (Figure 3.12). The crystals were frozen and transferred to the ESRF (Grenoble) where the same crystals were found to diffract to $\sim 5 \text{ \AA}$ (as shown in Figure 3.13).

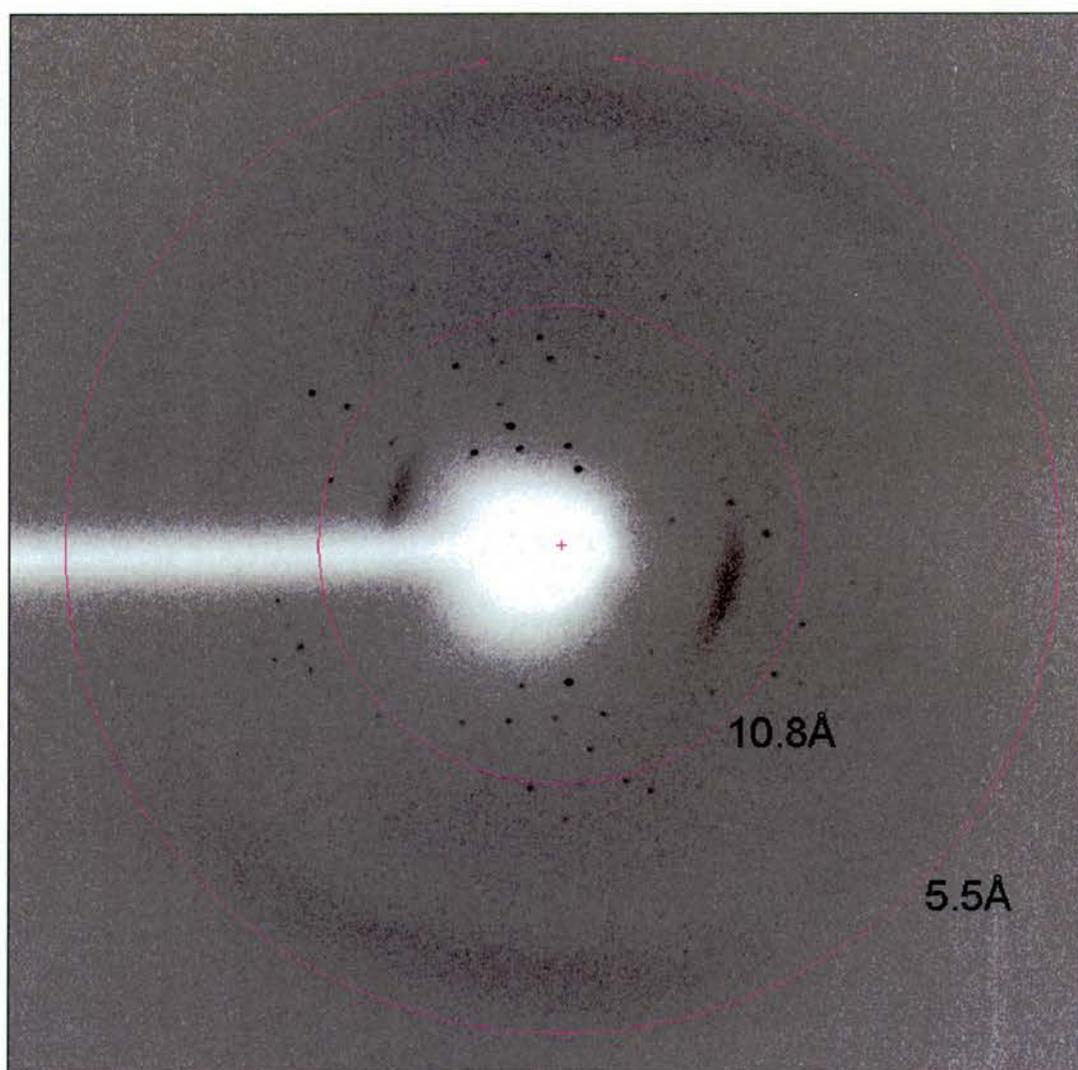


Figure 3.12 – Diffraction pattern of SeMet Sso6206 collected in-house

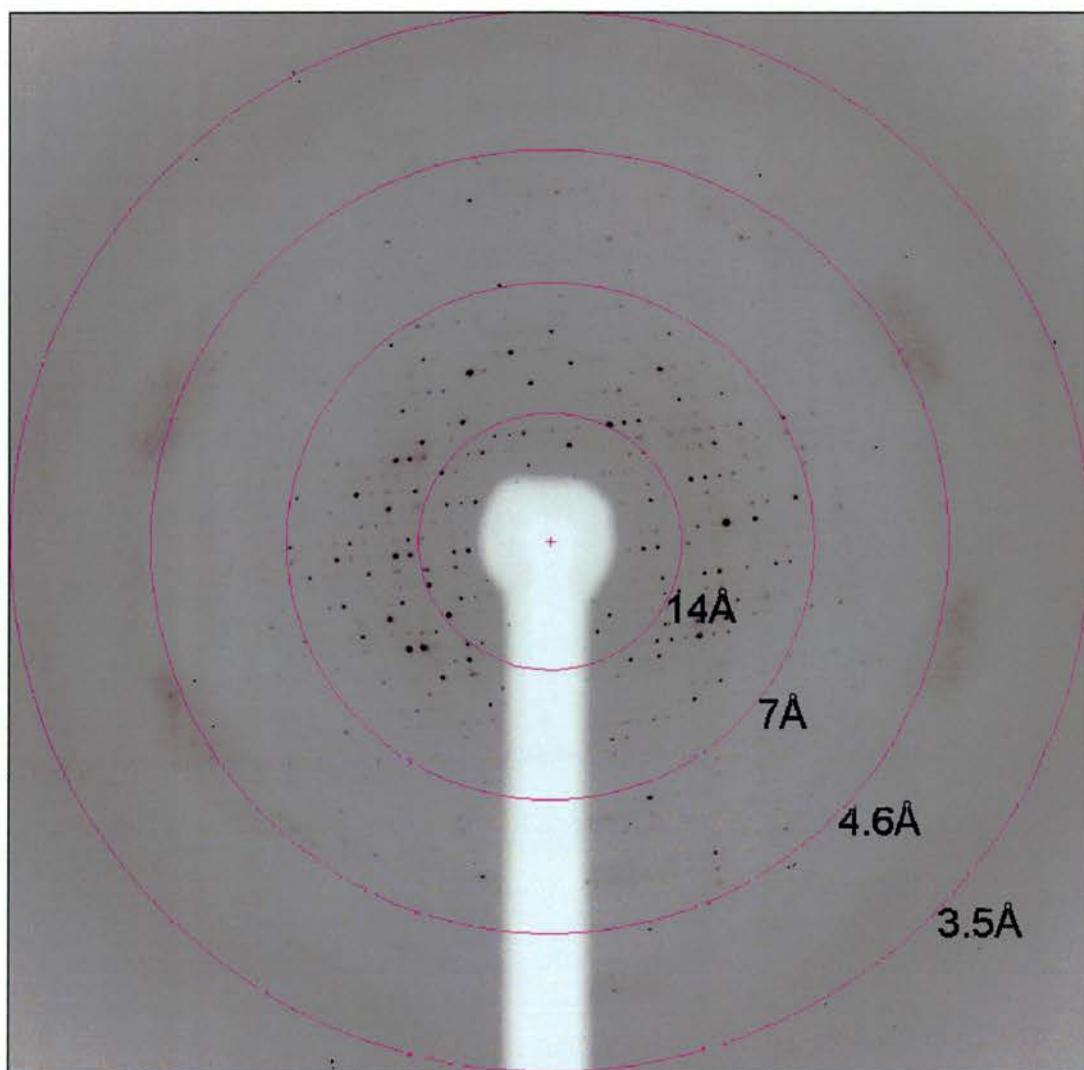


Figure 3.13 – Diffraction pattern of SeMet Sso6206 collected at the ESRF Grenoble

A fluorescence scan (Figure 3.14) was performed on the crystals indicating the presence of selenium within the crystal. A full data set was collected at the peak of the selenium signal using 1° oscillations over 200° . The crystal did not survive long enough to collect the inflection data, intended for MAD. Table 3.7 shows the data collection statistics.

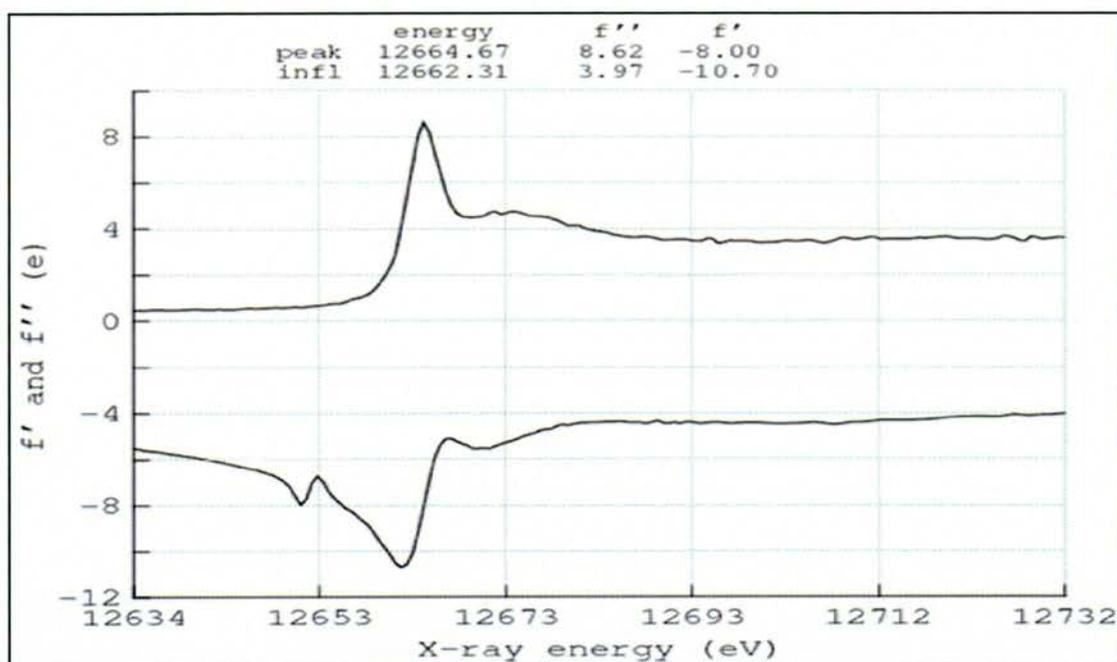


Figure 3.14 - Fluorescence scan (Se) performed on SeMet Sso6206 crystals

SeMet - ID23	
Wavelength (Å)	0.979
Resolution (Å)	128.0 – 5.0 (5.27 – 5.0)
Space group	P2 ₁ 2 ₁ 2 ₁
Temperature (K)	130
Detector	ADSC Q315R CCD
Unit-cell parameters (Å)	a = 119.24 b = 124.08 c = 128.25
Solvent (%)	53.3 (for 18 monomers)
Unique reflections	8678 (1222)
I / σ(I)	17.7 (5.6)
Average redundancy	7.8 (8.0)
Data completeness (%)	100.0 (100.0)
R _{merge} †	0.08 (0.33)
Anomalous	
Completeness	100.0 (100.0)
Multiplicity	4.2 (4.2)
f/f''	-8.0 / 8.62
Correlation	0.581 (-0.01)
Mid slope	1.20

Table 3.7 – Summary of data statistics for the SeMet data collection.

Table 3.8 shows an output from SCALA indicating the extent of anomalous signal ($\sim 6 \text{ \AA}$). We were unable to locate the SeMet sites using SHELX (Schneider and Sheldrick, 2002), SOLVE (Terwilliger, 2004) or SNB (Howell *et al.*, 2000).

>SEMET													
N	l/resol^2	dmax	CC_anom	N_anom	CC_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CC_Tmean	N_Tmean	
\$\$													
1	0.0040	15.81	0.289	173	-0.333	95	1.407	172	0.710	95	0.998	268	
2	0.0080	11.18	0.787	391	-0.536	123	2.385	388	0.799	120	0.999	514	
3	0.0120	9.13	0.349	526	0.360	127	1.454	519	0.934	115	0.999	653	
4	0.0160	7.91	0.342	629	0.799	125	1.351	617	1.082	119	0.998	754	
5	0.0200	7.07	0.472	732	-0.322	129	1.495	720	0.939	121	0.997	861	
6	0.0240	6.46	0.297	808	0.020	127	1.316	800	0.888	121	0.996	935	
7	0.0280	5.98	0.128	891	-0.201	125	1.140	881	1.075	119	0.996	1016	
8	0.0320	5.59	0.075	968	-0.028	129	1.073	964	0.961	126	0.985	1097	
9	0.0360	5.27	0.000	1026	-0.041	129	1.063	1015	1.041	125	0.987	1155	
10	0.0400	5.00	-0.012	1081	0.299	124	1.016	1072	1.234	118	0.985	1205	
\$\$													
Overall			0.581	7225	0.041	1233	1.792	7148	0.816	1179	0.999	8458	

Table 3.8 – Table of anomalous data as output from SCALA. The CC_anom is the anomalous correlation coefficient (red box) indicates an anomalous signal to $\sim 5.5 \text{ \AA}$.

The unit cell dimensions for SeMet crystals are also large suggesting the presence of between 14 ($V_m 3.39 \text{ Da \AA}^{-3}$, solvent content 64 %) and 26 ($V_m 1.82 \text{ Da \AA}^{-3}$, solvent content 32 %) per ASU.

3.4.6 Sulfur phasing

We aimed to measure phases using the sulfur anomalous signal. Data was collected to 2.7 \AA using $\lambda = 1.771$; as this has been shown to produce an sulfur anomalous signal (Ramagopal *et al.*, 2003). The only way to obtain the weak sulfur anomalous signal is by single wavelength anomalous diffraction, as MAD does not lead

to differences in scattering at this energy (as it is far from its absorption edge). The data collection requires to be highly redundant therefore data were measured over 600°.

<i>Sulfur SAD – ID23</i>	
Wavelength (Å)	1.771
Resolution (Å)	78.5 – 2.7 (2.85 – 2.7)
Space group	$P6_1c22$
Temperature (K)	130
Detector	MarCCD
Unit-cell parameters (Å)	a = b = 157.3, c = 307.4
Solvent (%)	
Unique reflections	62420 (8922)
$I/\sigma(I)$	44.7 (9.3)
Average redundancy	48.0 (48.0)
Data completeness (%)	100.0 (100.0)
Rmerge †	0.096 (0.514)
Anomalous	
Completeness	100.0 (100.0)
Multiplicity	25.5 (25.5)
Correlation	0.029 (0.015)
Mid slope	1.099

Table 3.9 – Summary of data statistics for the sulfur SAD experiment.

Sulfur SAD												
N	1/resol ²	dmax	CC_ancm	N_ancm	CC_cen	N_cen	RCR_ancm	N_ancm	RCR_cen	N_cen	CC_Imean	N_Imean
1	0.0137	8.54	0.318	1475	-0.124	778	1.502	1473	1.089	765	1.000	2253
2	0.0274	6.04	0.525	2985	0.031	786	1.745	2979	1.032	770	1.000	3771
3	0.0412	4.93	0.026	3988	-0.050	789	1.109	3943	0.973	761	0.999	4777
4	0.0549	4.27	-0.183	4793	-0.076	787	0.910	4678	1.089	705	0.999	5580
5	0.0686	3.82	0.043	5493	0.202	795	1.097	5414	0.921	740	0.999	6288
6	0.0823	3.49	0.073	6106	-0.222	789	1.067	6061	1.022	753	0.999	6895
7	0.0960	3.23	0.065	6672	0.075	791	1.041	6630	0.961	753	0.999	7463
8	0.1097	3.02	0.002	7192	0.083	785	1.006	7166	0.998	757	0.998	7977
9	0.1235	2.85	0.009	7704	-0.008	787	1.003	7673	0.967	735	0.995	8491
10	0.1372	2.70	0.015	8131	-0.082	791	1.015	8112	0.994	753	0.984	8922
§§												
Overall			0.029	54539	-0.072	7878	1.111	54129	1.052	7492	1.000	62417

Table 3.10 – Table of anomalous data (Sulfur SAD), the column containing the correlation coefficient for anomalous data is highlighted.

Processing indicates an anomalous signal to around 6 Å (Table 3.10), however no sites could be located.

3.5 CONCLUSIONS

The structure of this protein should give some insight into the function of Sso6206 and why it is so highly conserved within archaea. There is no convincing sequence match between Sso6206 and any structure in the PDB, thus ruling out molecular replacement as a phasing strategy. Several attempts to incorporate heavy atoms by soaking into existing crystals or by co-crystallization with recombinant protein have proven unsuccessful in that crystals which grew, or survived soaking, did not show a clear anomalous diffraction when tested at the SRS Daresbury. Selenomethionine derivative protein was prepared which crystallized in the spacegroup $P2_12_12_1$ ($a = 199.2$, $b = 124.1$, $c = 128.2$) and data were collected (to $\sim 4 \text{ \AA}$) which showed some anomalous data, however phases were not obtained. Data were also collected on native crystals using radiation ($\lambda = 1.77 \text{ \AA}$) with the aim to obtain phases calculated from the sulfur anomalous signal, however this was also unsuccessful.

The cell dimensions of native crystals are consistent with an asymmetric unit content varying from 15 ($V_m 3.50 \text{ Da \AA}^{-3}$, solvent content 65 %) to 30 ($V_m 1.77 \text{ Da \AA}^{-3}$, solvent content 30 %). The relatively low diffraction resolution of the crystals suggests a higher solvent content. Mass spectrometry has been able to identify trimers and hexamers. Gel filtration suggests that the protein is in excess of 100 kDa since it eluted at the void volume. Together these data suggest that Sso6206 forms a large multimeric protomer. If such an oligomer is indeed found for the protein it is likely to have important functional consequences.

3.6 FUTURE WORK

Suitable phases must be obtained for solution of Sso 6206. This will be achieved through the ongoing work by the SSPF, through continuation of heavy metal derivatisation, either by soaking or by SeMet incorporation. The SSPF are also pursuing new leads from sparse matrix screening. The structure should provide an insight into the biological role of this highly intriguing protein.

4

**An Investigation into KDPGal Aldolase;
A Stereospecific Pyruvate Dependent Class I Aldolase**

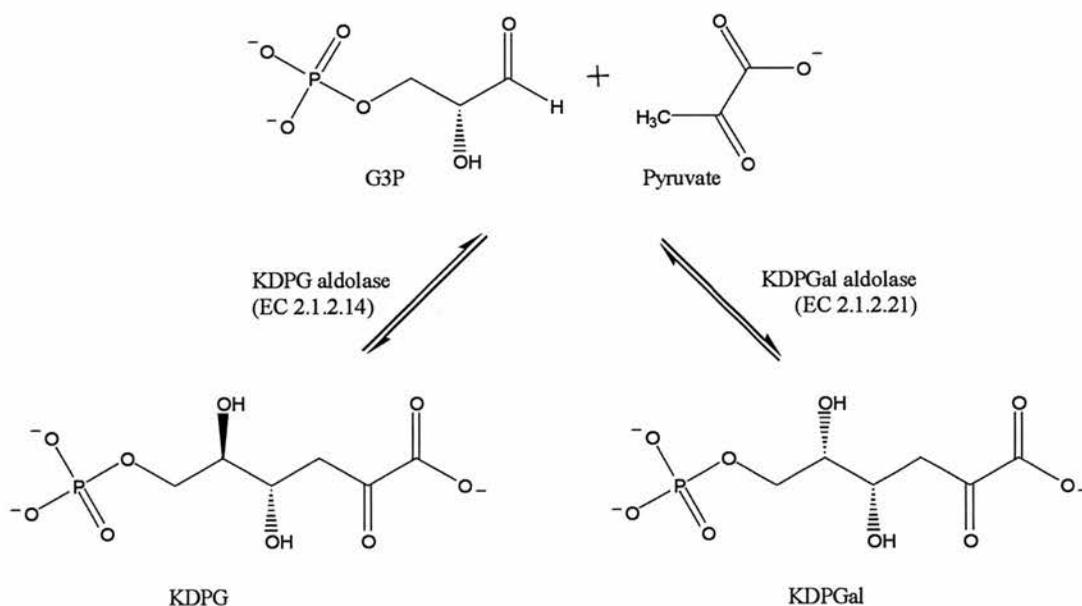
4.1 SUMMARY

The X-ray crystal structure of *E. coli* KDPGal aldolase has been determined within our group. Despite being structurally similar to the KDPG aldolase homologue, it catalyses the formation of 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) rather than the 2-keto-3-deoxy-6-phosphogluconate (KDPG). Comparison of the native (apo) structure does not clearly identify the reason for this difference. In order to rationalize this difference an investigation was carried out to determine the crystal structure of substrates and products in the enzyme. Models were achieved between 2.6 Å and 1.9 Å resolution, showing Schiff base formation, and binding of G3P from soaks with G3P and KDPGal. These provide evidence for the stereo-selectivity of the aldolases.

4.2 INTRODUCTION

Aldolases are important biocatalysts as they are capable of stereo specific carbon – carbon bond formation. The catalytic mechanism and stereo-specificity of these enzymes is widely studied (Samland and Sprenger, 2006). Recent work within the lab has obtained X-ray crystal structures of some aldolase enzymes; the *Escherichia coli* and *Thermotoga maritima* KDPG aldolase (EC 4.1.2.14) (Fullerton *et al.*, 2006) and the *E. coli* KDPGal aldolase (Srikannathasan and Naismith, 2005). The KDPG aldolase is a constitutive enzyme responsible for the reversible conversion of glyceraldehyde-3-phosphate (G3P) and pyruvate to 2-keto-3-deoxy-6-phosphogluconate (KDPG) and is part of the Entner-Doudoroff pathway found in bacteria (Conway *et al.*, 1991). The

KDPGal aldolase is inducible during growth in the presence of galactose or galactonate and it can convert G3P and pyruvate to 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) (Griffiths *et al.*, 2002; Henderson *et al.*, 1998). The transformations are shown in Scheme 4.1.



Scheme 4.1 – KDPG aldolase reaction versus the KDPGal aldolase. The enzymes are responsible for the breakdown of the sugars in the cell.

A sequence alignment of these aldolases (Figure 4.1) shows that KDPGal aldolase is 23.9 % identical and 44.5 % homologous with the KDPG aldolase from *E. coli*. The active site lysine (red arrow), which is responsible for formation of the Schiff base intermediate is conserved throughout these proteins.

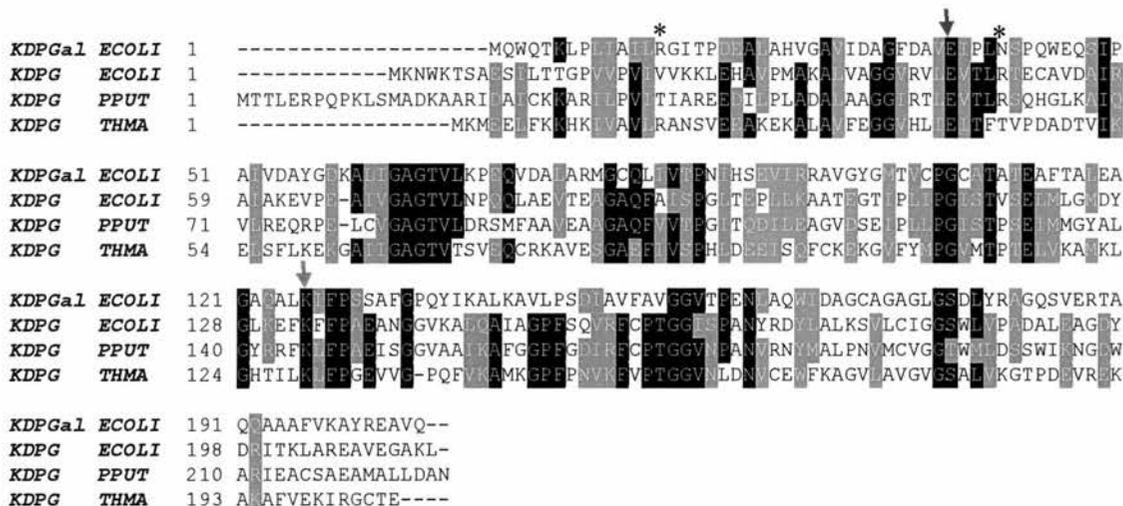


Figure 4.1 – Sequence alignment (ClustalW) KDPGal aldolase from *ECOLI*; *E. coli*, KDPG aldolase from *E. coli*, PPUT; *Pseudomonas putida*, THMA; *Thermotoga maritima*. A conserved active site lysine (red arrow) and glutamic acid (blue arrow) are indicated. The asterisk indicates the position of the active site arginines.

The structure of KDPG aldolase was first reported using protein isolated from *Pseudomonas putida* (Mavridis *et al.*, 1982; Mavridis and Tulinsky, 1976). Later work on protein from *E. coli* confirmed that the protein adopts an $(\alpha/\beta)_8$ fold which forms a propeller like trimer complex. The covalent Schiff base intermediate between Lys-133 and pyruvate was identified. The glutamic acid residue (Glu-45) was implicated as involved in catalysis (Allard *et al.*, 2001). A cartoon representation of the KDPG aldolase from *E. coli* is shown in Figure 4.2.

A detailed mechanism of the class I KDPG aldolase was proposed from work carried out on proteins from *E. coli* and *Thermotoga maritima* (Fullerton *et al.*, 2006). Structures were obtained using native and mutant (Glu-45-Asn) protein. The study showed that reactivity depends on the catalytic lysine and a glutamate/aspartic acid/base.

There were two catalytic waters identified in the active site are proposed to be essential for activity; acting as proton shuttles during intermediate/product formation. Two waters are thought to be involved in the overall mechanism proposed for class I aldolase enzymes as shown in Scheme 4.2. Similar ideas were first suggested from a structural study into the mechanism of DERA (Heine *et al.*, 2001), although there are several differences in the key residues (Fullerton *et al.*, 2006).

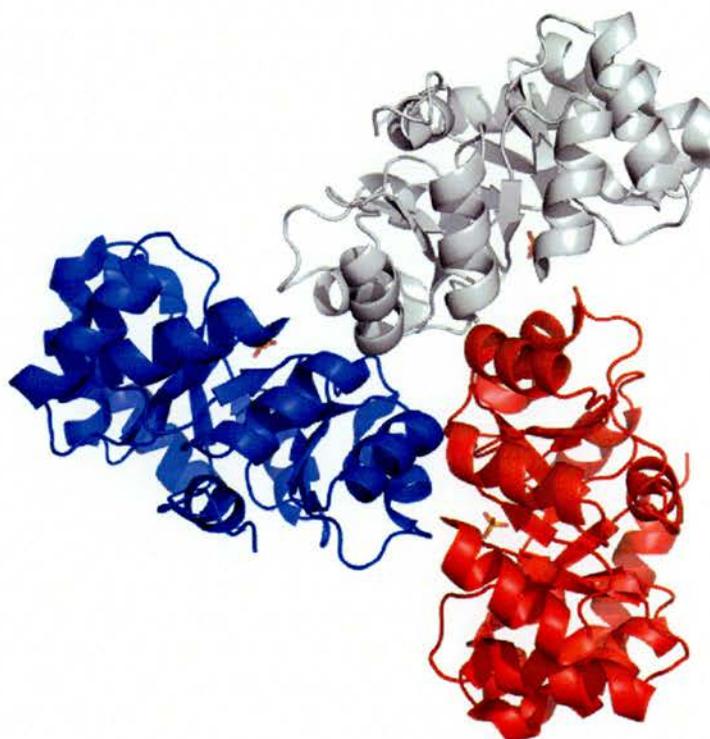
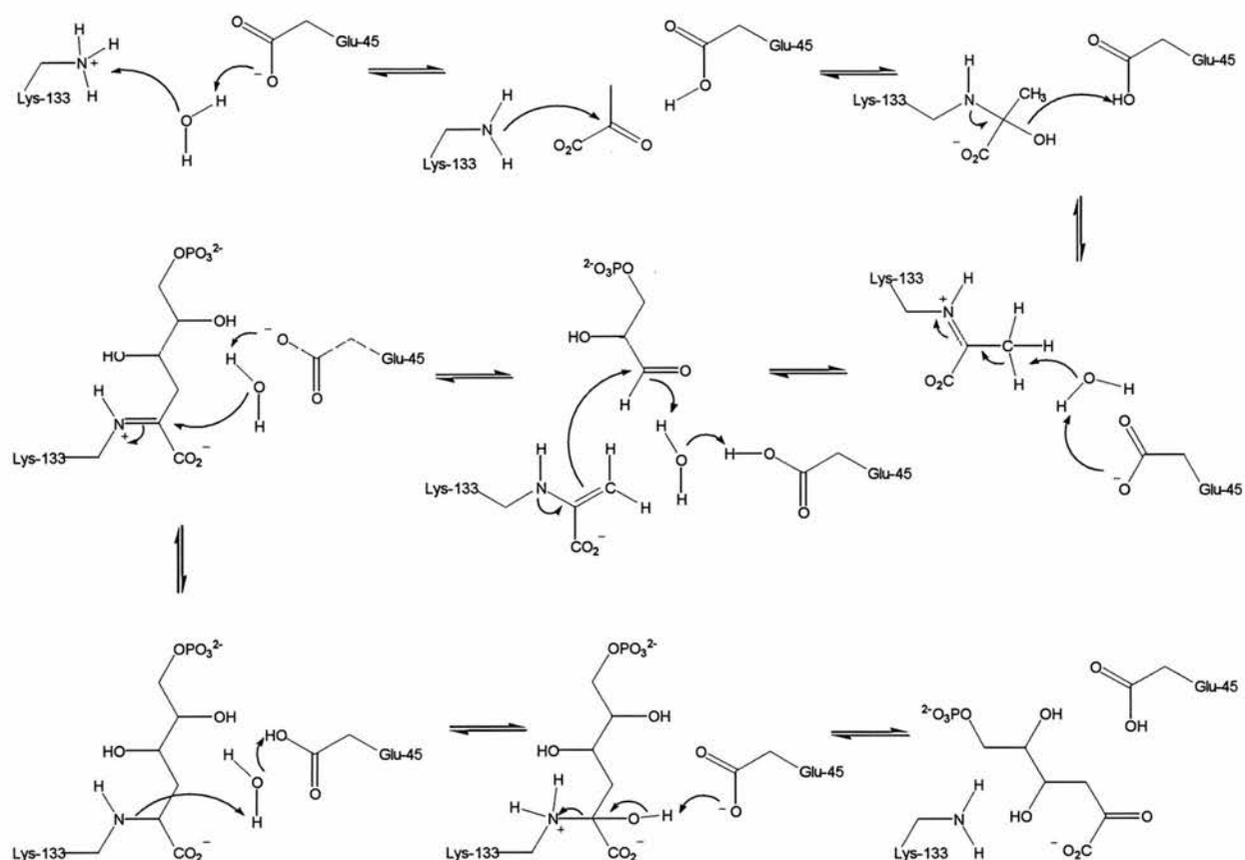


Figure 4.2 – KDPGal aldolase trimer.



Scheme 4.2 – The mechanism of KDPGal Aldolase (Fullerton *et al.*, 2006).

KDPGal aldolase is interesting since it is clearly similar, in terms of the sequence, to the KDPGal aldolase, however it confers different stereo-selectivity at the C4 stereogenic centre (Cotterill *et al.*, 1998). The *E. coli* enzyme has been purified, crystallized and a crystal structure determined (Figures 4.3 and 4.4). The structure was solved by single wavelength anomalous diffraction (Srikannathasan and Naismith, 2005).

The native structure showed the catalytic water (Figure 4.3), which was proposed as the proton shuttle (Fullerton *et al.*, 2006) between Glu-37 (β -strand 2) and Lys-126 (β -strand 6).

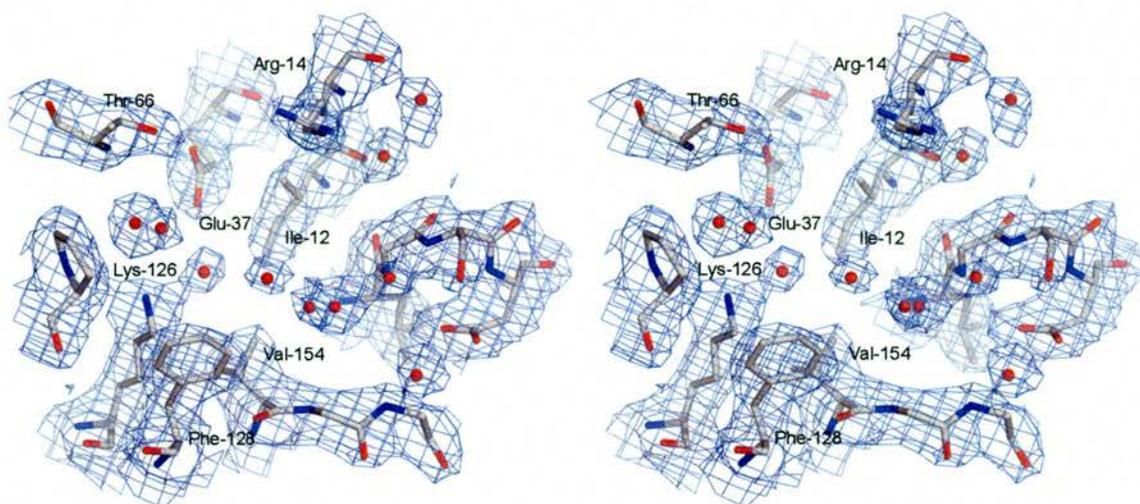


Figure 4.3 – A representation of the active site of the KDPGal aldolase, a slice of observed density are shown (blue mesh = 1.0 σ). Residues located at the active site have been labelled, several waters were located (red spheres).

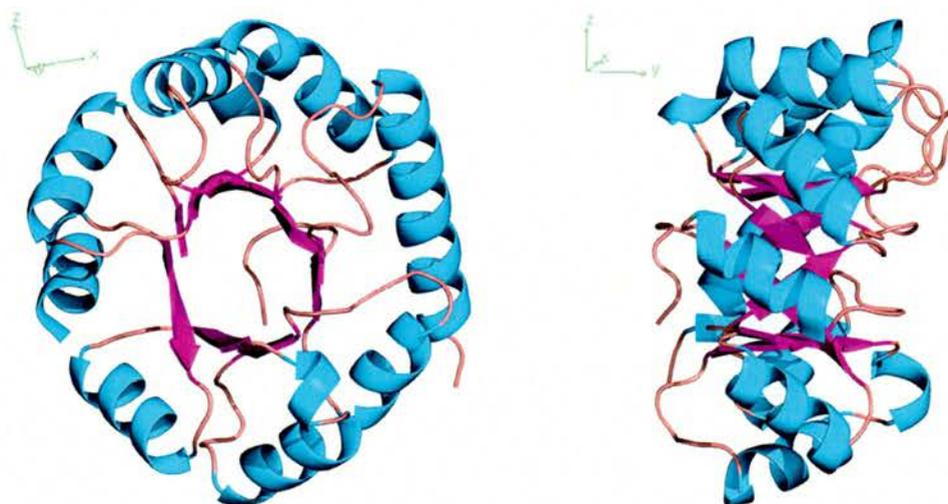


Figure 4.4 – Cartoon representation of the KDPGal aldolase monomer at different orientations.

The biological unit can be reconstructed from crystallographic symmetry and adopts the same general trimeric structure as the KDPG aldolases. Superimposition of

the structure (Figure 4.5), revealed that the KDPGal aldolase shares fold homology with the KDPG aldolase models.

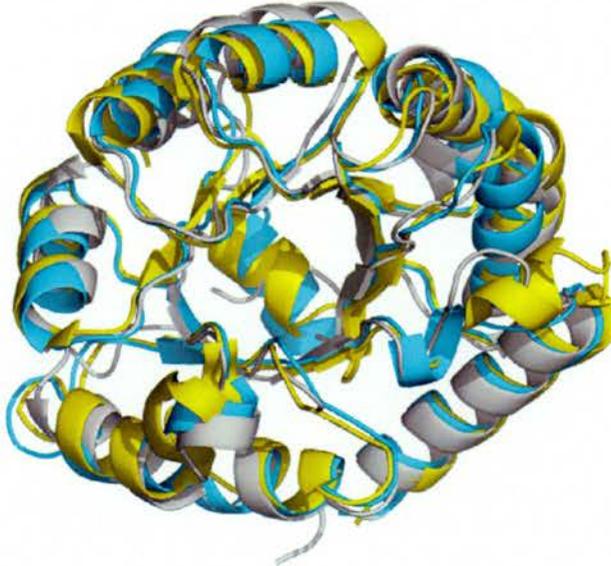


Figure 4.5 – Superimposition of KDPGal aldolase and superimposition of the *E. coli* KDPG (Yellow, PDB: 2c0a), *Thermotoga maritima* (Cyan, PDB: 1wa3) and *E. coli* KDPGal aldolase (Grey) monomer models.

The active site of the both KDPG and KDPGal aldolases are located inside the β -barrel. The active sites share many similar features (Figure 4.6). Comparison of *E. coli* and KDPGal aldolase structures reveal Arg-49 in KDPG is replaced by Asn-41 in KDPGal. A compensation mutation is seen in which Val-22 in KDPG, is replaced by Arg-14 in KDPGal aldolase. This is unlikely to be functionally significant as the *Thermotoga maritima* KDPG is located on the same strand (β -strand 1) as KDPGal aldolase.

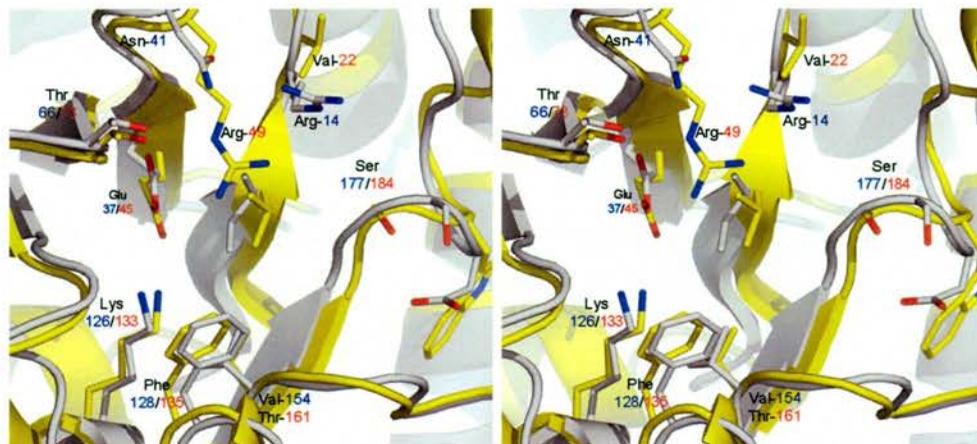


Figure 4.6 – Superimposition of the active site of KDPG aldolase (PDB 1EUN – yellow cartoon, red labels) and KDPGal aldolase (grey cartoon, blue labels).

4.3 EXPERIMENTAL AIMS

It was planned to develop a structural understanding of stereo-selectivity of KDPGal aldolase from *E. coli* using structural studies into substrate and product complexes.

4.4 MATERIALS AND METHODS

4.4.1 Transformation and purification

Plasmid DNA (pET28a-KDPGal) (Eric Toone – Duke University) was transformed into BL21 (DE3) (Novagen). An aliquot (1 mL) of overnight culture was used to inoculate TPB (*Tryptone phosphate broth*) (500 mL), treated with carbenicillin ($100 \mu\text{g mL}^{-1}$), in a baffled flask (2 L) and grown (298 K, 200 rpm) to an $\text{OD}_{600} = 0.6$. Protein expression was induced using IPTG (0.5 mM) with constant temperature and agitation. The cells were harvested after an overnight induction by centrifugation (8000

rpm, 20 min, 277 K, Beckman Avanti J20-XP, JL8.100 rotor). The cells were resuspended in PBS to remove any excess growth media, and re-spun (8000 rpm, 20 min, 277 K, Beckman Avanti J20-XP, JL8.100 rotor). The cells were stored at 253 K until required. The cells from four 500 mL cultures were resuspended in lysis buffer [20 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole, 20 uM lysozyme and 20 ug mL⁻¹ DNase I, one (EDTA free) complete protease inhibitor cocktail (Roche)] and stirred for 30 min at 298 K. The cells were disrupted by Constant Systems cell disruptor (2 cycles, 20 KPSI, 298 K). The disrupted product was centrifuged (15000 rpm, 20 min, 277 K, Beckman Avanti J20-XP, JA 25.50 rotor). The supernatant was collected and passed through a 0.22 µm syringe filter membrane (Millipore) prior to application to a (15 mL) Ni⁺ charged, Ni-NTA metal chelating column (Amersham Biosciences). The column was washed using 15 CV of wash buffer [20 mM HEPES pH 7.5, 300 mM NaCl and 30 mM imidazole]. The protein was eluted using a single step of elution buffer [20 mM HEPES pH 7.5, 300 mM NaCl and 500 mM imidazole]. The fractions containing the pure target protein were determined by SDS-PAGE, pooled and concentrated to 5 mg mL⁻¹ (determined by Bradford assay). Some precipitant was removed by centrifugation (15000 rpm, 20 min, 277 K, Beckman Avanti J20-XP, JA 25.50 rotor). The protein was dialyzed (2000x) overnight against a final buffer [20 mM HEPES pH 7.5 and 100 mM NaCl]. The protein purity was polished using size exclusion gel filtration using a Superdex™ 200 column. The protein integrity and identity were confirmed by mass spectrometry. The purity of the sample was determined by Coomassie™ Blue stained SDS-PAGE gel. In total 60 mg of protein was purified from the 2 L cell culture.

4.4.2 Crystallization

Crystallization conditions for apo-crystals had been identified previously (Srikannathasan and Naismith), using 0.1 M MES pH 6.0, 30 % PEG 200 and 5 % PEG 3000. An optimisation screen was prepared based on this.

		0.1 M MES				
		5.6	5.8	6.0	6.2	6.4
PEG 200	22 %					
	24 %					
	26 %					
	28 %					
	30 %					
	32 %					
	34 %					
	36 %					

Table 4.1 – Grid screen for *E. coli* KDPGal aldolase crystal growth, where optimization of % w/v PEG 200 is varied against pH of 0.1 M MES and 5 % PEG 3000 is used throughout.

Apo-protein (concentrated to 5 mg mL⁻¹) was crystallized using the sitting drop vapour diffusion method, by 2 µL (protein) + 2 µL drops (mother liquor), over the course of a day. Crystals were used for data collection between one day to 2 months after they reached full size.

4.4.3 Compound library

Several compounds were chosen as targets for co-crystallization and soaking experiments; as detailed in Table 4.2.

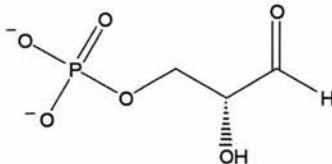
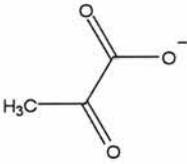
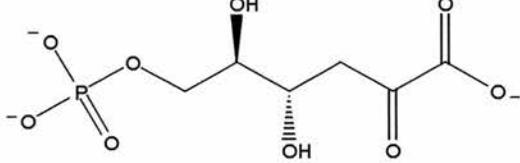
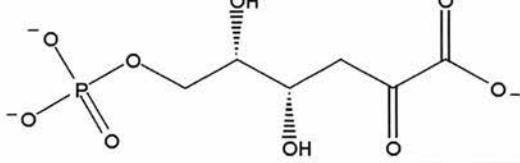
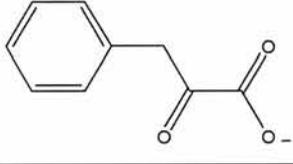
Name (Abbreviation) Source	Structure
glyceraldehyde-3-phosphate (G3P) Sigma Chem. Co. Ltd., UK	
pyruvate (Pyr) Sigma Chem. Co. Ltd., UK	
2-keto-3-deoxy-6-phospogluconate (KDPG) E. Toone	
2-keto-3-deoxy-6-phospogalactonate (KDPGal) E. Toone	
phenyl pyruvate PP E. Toone	

Table 4.2 – Compounds used for co-crystallization and soaking experiments.

4.4.4 Co-crystallization

Co-crystallization was attempted using 10 mM ligand as listed in Table 4.2. (above). The protein was pre-incubated with the ligand solution for at least 6 h at 298 K. The mixture was centrifuged (13000 rpm for 2 min, Eppendorf Centrifuge 5415D) to remove any insoluble ligand. Crystallization was attempted using the vapour diffusion method (as described previously). Crystals did not appear from these attempts and sparse matrix screening was performed to identify other conditions.

4.4.6 Ligand soaking

Complexes were obtained by soaking native crystals obtained (at 298 K) immediately prior to data collection. The native crystals were transferred from the growth condition into mother liquor containing 10 mM of each of the compounds (described in Table 4.2). Crystals usually tolerate soaking of ligand, however in some cases the crystals began to dissolve into the solution, or cracked. Those which survived were assessed by diffraction.

4.4.7 Data collection and processing

Crystals were directly frozen in the cryo-stream since the crystallization condition does not require additional cryo-protectant. Data were collected using a rotating anode X-ray generator (Rigaku), which produced Cu $K\alpha$ radiation ($\lambda = 1.5418 \text{ \AA}$) with Osmic mirrors. The detector was a Raxis IV⁺⁺ image plate (Rigaku). Data were indexed and integrated using MOSFLM (Leslie, 1999; Leslie, 1992) and merged using SCALA (Evans, 1997). The structures were solved by molecular replacement using MOLREP (Vagin and Teplyakov, 2000) as part of the CCP4i suite (Potterton *et al.*, 2002), with protein coordinates (trimer) derived from the existing model. Structures were refined using Refmac5 (Murshudov *et al.*, 1997), the ligands were modelled into the unbiased Fo-Fc density. The final models were validated and Ramachandran plots were prepared using PROCHECK (Laskowski *et al.*, 1996). The details of refinement and validation are discussed in the results for each structure. Crystallographic figures were produced using PyMOL™ (Version 0.99 rc6 – DeLano Scientific) unless otherwise stated. Ligand libraries were prepared using PRODRG (Schuttelkopf and van Aalten, 2004). The

covalently modified lysine (in the pyruvate soak model) was replaced as an atypical amino acid (LYP) and the refinement libraries modified as necessary.

4.5 RESULTS AND DISCUSSION

4.5.1 Purification and crystallization

Protein was purified to homogeneity (confirmed by MS and SDS gel) and concentrated to $\sim 5 \text{ mg mL}^{-1}$ for crystallization. Crystallization of the protein was rapid ($< 8 \text{ h}$) and appeared from precipitant (Figure 4.7), the crystals grew to typically 0.1 to 0.4 mm^3 .

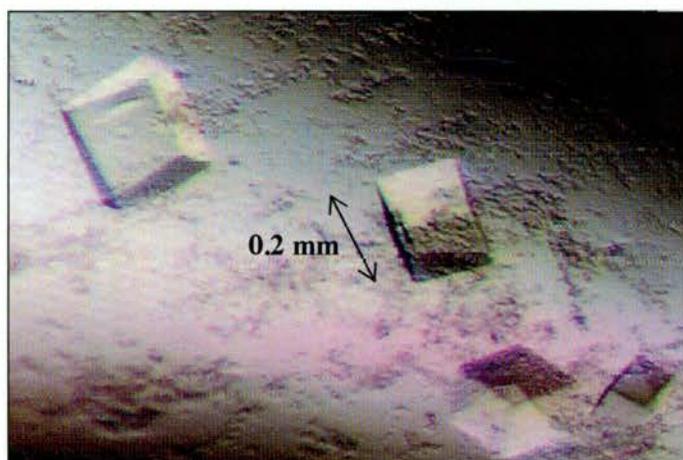


Figure 4.7 – *KDPGal aldolase crystals grown from precipitant in conditions containing 30 % w/v PEG 200, 0.1 M MES ph 6.2 and 5 % PEG 3000.*

Native crystals were transferred to drops containing $\sim 10 \text{ mM}$ ligand and soaked for approximately 10 min. Some crystals appeared to survive longer soaks however

allowing the crystals to soak for too long resulted in loss of integrity. Diffraction from the pyruvate complex is shown in Figure 4.8.

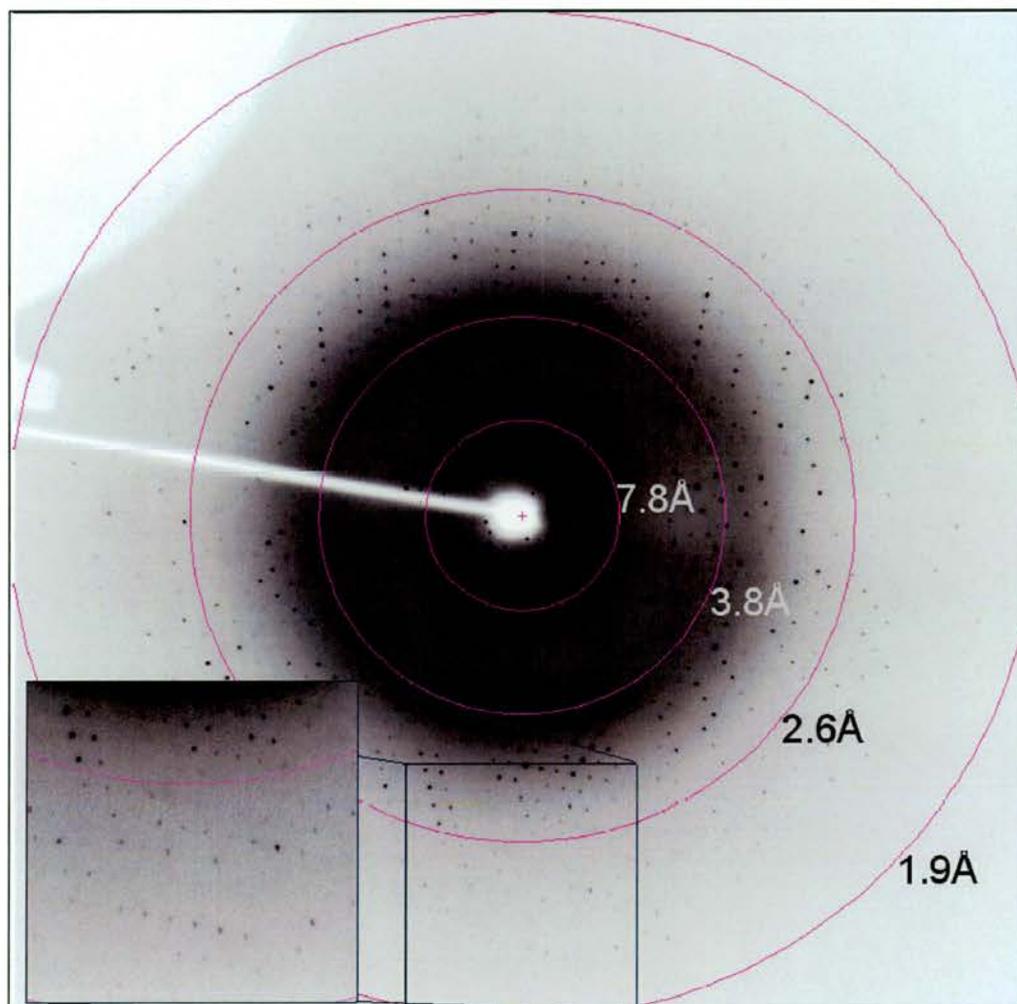


Figure 4.8 – A diffraction image of the KDPGal aldolase crystals after soaking.

4.5.2 Crystal structure of KDPGal aldolase and pyruvate.

Native protein crystals were obtained and transferred to drops containing 10 mM sodium pyruvate (Sigma Chem. Co. Ltd) for 10 min at 298 K immediately prior to flash freezing. A dataset was collected to 1.8 Å resolution from a single crystal. The data

were scaled and processed and the structure determined by molecular replacement from the existing model. The electron density obtained showed clearly that pyruvate was covalently bound to the active site lysine. Density consistent with a phosphate was seen close to Arg-14. Figure 4.9 shows pyruvate modified lysine and phosphate modelled into the relative density and the new conformation of Arg-14. The refined model showed that the protein residues, pyruvate modified lysine, phosphate ion and waters fitted well (Figure 4.9b).

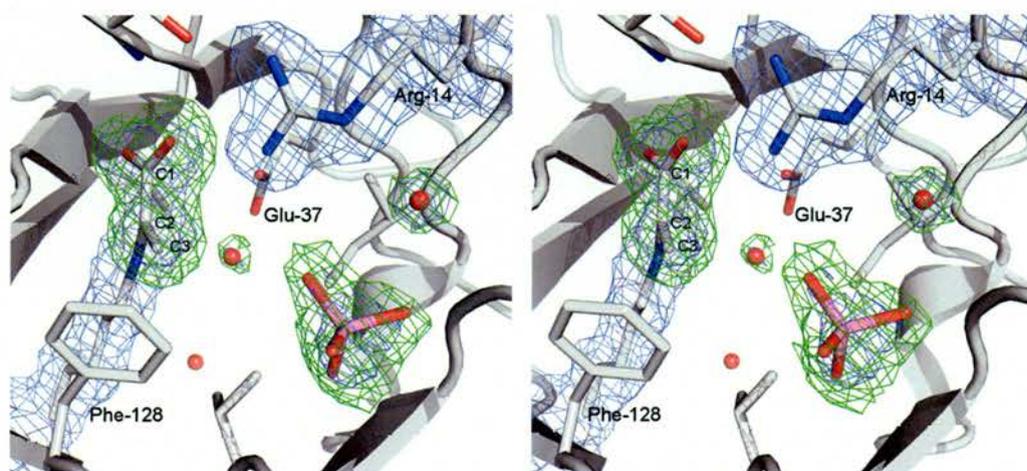


Figure 4.9a – Initial density in the active site of crystals soaked with pyruvate solution. The pyruvate (covalently bound to the lysine) and phosphate and waters (red spheres) are modelled into the F_o-F_c (green mesh and $+3.0 \sigma$) electron density map after molecular replacement. (C = white; O = red; N = blue).

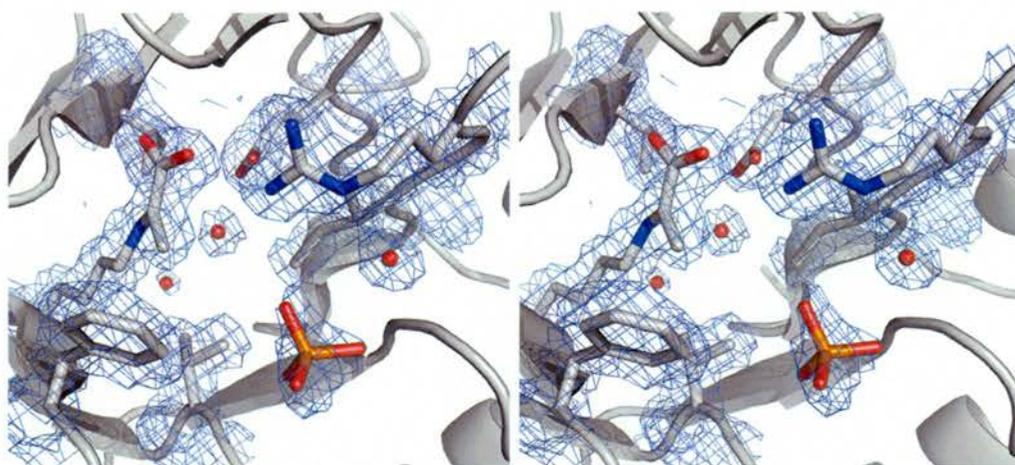


Figure 4.9b – Fully refined active site showing the ligands as refined into the $2F_o-F_c$ maps (blue mesh and 1.0σ).

The orientation of the pyruvate is consistent with the expected delocalisation of electrons across several bonds such that all of the atoms of the pyruvate (including N from lysine) are almost planar ($C3-C2-C1-O1$ torsion = $\sim 14^\circ$). The libraries constructed for the pyruvate complex were not restrained to allow the ligand to adopt the conformation as observed from the density. The CH_2 of the pyruvate points towards Phe-128 which may stabilize the anion. Arg-14 folds into the active site making a hydrogen bonding network between Asn-41, Thr-66 and a carboxyl oxygen of the pyruvate complex. These interactions may be important in positioning the pyruvate. Further pyruvate interactions include a nitrogen of Thr-66 the carboxylic group of Thr-86. The shuttling water molecule makes a hydrogen bond to C2 ($\sim 2.7 \text{ \AA}$). The interactions between pyruvate and the protein are shown in Figure 4.10 and the LIGPLOT representation in Figure 4.11. The data collection and refinement statistics are shown in Table 4.3. The Ramachandran plot (Figure 4.13) suggests that the model quality is excellent.

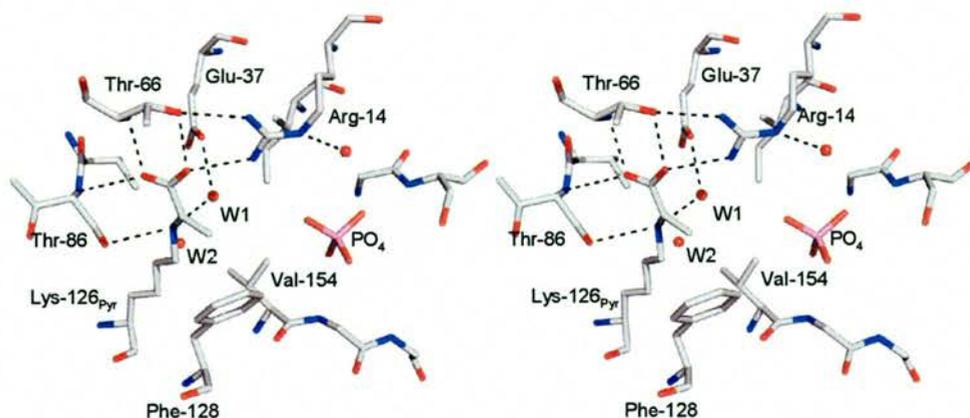


Figure 4.10 – Key interactions (black dashes) in the active site between pyruvate (covalently bound to Lys-126) and the protein. The waters (labelled W1 and W2) are those expected to be involved in proton shuttling (Fullerton et al., 2006).

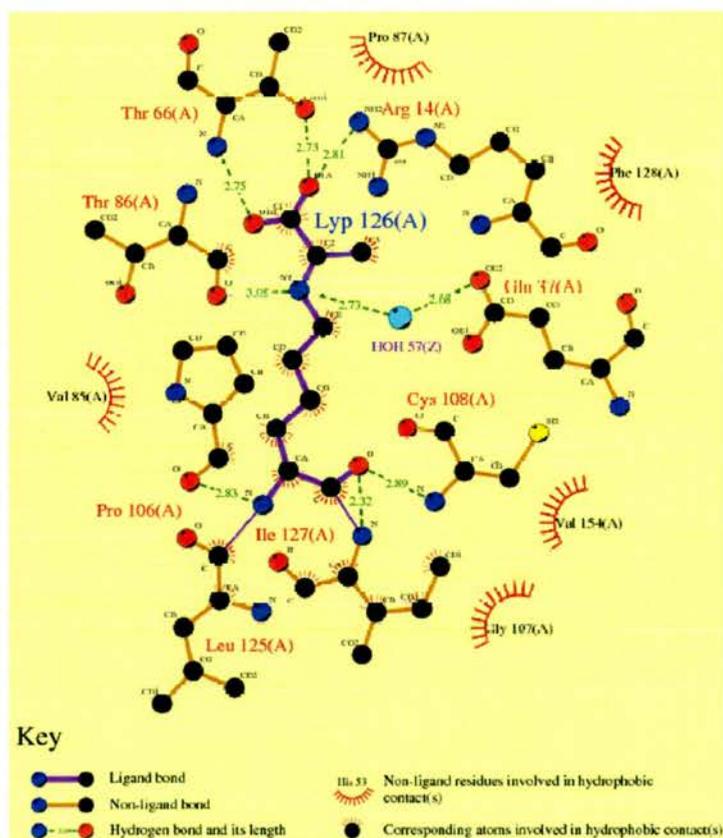


Figure 4.11 – LIGPLOT representation of interactions between the pyruvate modified lysine (Lyp-126) and KDPGal aldolase. (Wallace et al., 1995).

Superimposition of the Schiff base complexes of KDPGal aldolase and KDPG aldolase (*Thermotoga maritima*) shows little difference (Figure 4.12). The relative positions of protein and pyruvate residues are identical. Thr-156 and Val-15 in the *Thermotoga maritima* KDPG aldolase are replaced by Val-154 and Ile-20 in *E. coli* KDPGal aldolase, respectively.

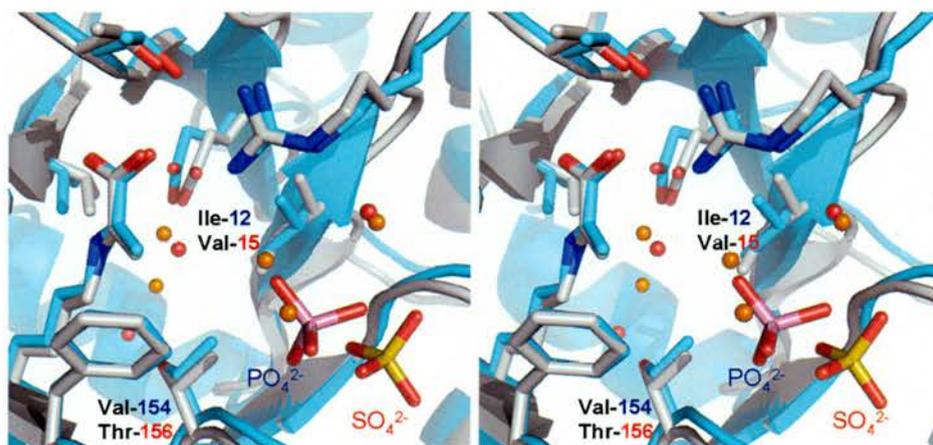


Figure 4.12 – Active site of KDPGal aldolase (C = white; N = blue; O = red; HOH = red spheres) superimposed with the active site of KDPG aldolase of *Thermotoga maritima* (C = cyan; HOH = orange spheres).

The positions of the sulfate and phosphate ions between these models differ by ~4 Å. The phosphate binding site seen here may not be a mimic for the substrate binding site. It was not possible to construct a simple substrate prediction model (as was done for KDPG) based on the pyruvate and phosphate. We also investigated the effect of soaking phenyl pyruvate into the crystals as we expected that it will result in the covalently altered lysine, however neither soaking nor co-crystallization yielded diffracting crystals.

Data collection	pyruvate soak
Wavelength (Å)	1.541
Resolution (Highest Shell) (Å)	28.0 – 1.8 (1.9 – 1.8)
Space Group	H3
Cell constants (Å)	104.20, 104.20, 73.24
Unique reflections	26836 (3475)
Average redundancy	3.1 (1.9)
I/σ	16.4 (3.0)
Completeness (%)	97.7 (86.4)
R_{merge} (%)	5.0 (38.3)
Refinement	
R	0.192
R_{free}	0.228
Rmsd bond (Å)/angle(°)	0.016 / 1.586
Residues in Ramachandran core (%)	100.0
PDB accession code	- - - -

Table 4.3 – Summary of data collection and refinement statistics

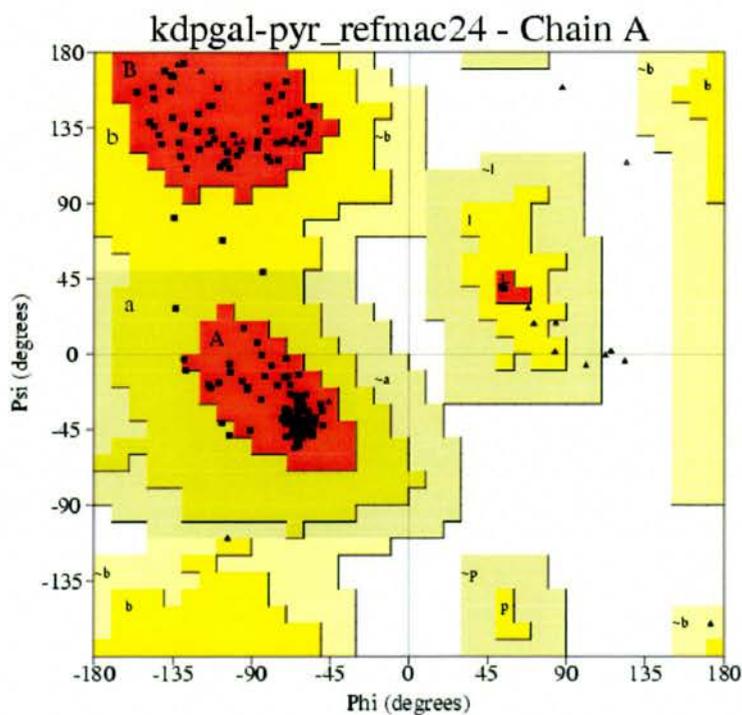


Figure 4.13 – Ramachandran plot of the KDPGal aldolase with pyruvate model.

4.5.4 Co-crystal structure of KDPGal aldolase and glyceraldehyde-3-phosphate

Native crystals were soaked with a solution of glyceraldehyde-3-phosphate (Sigma Chem. Co. Ltd) (10 mM) prepared in stabilization buffer for 10 min at 298 K immediately prior to flash freezing. A dataset was collected, measured to 2.6 Å resolution using a single crystal. The data were scaled and processed and the structure determined by molecular replacement from the existing model. The initial density obtained indicated the presence of two entities located in the active site. One was consistent with a sulfate or phosphate close to the active site lysine, and since no sulfate (or even pyruvate) was expected in the crystallization / soaking buffers, phosphate (from the G3P solution) was placed into the density. G3P was fitted into the second density and refinement continued. The final density obtained for the full refined model is shown in Figure 4.14.

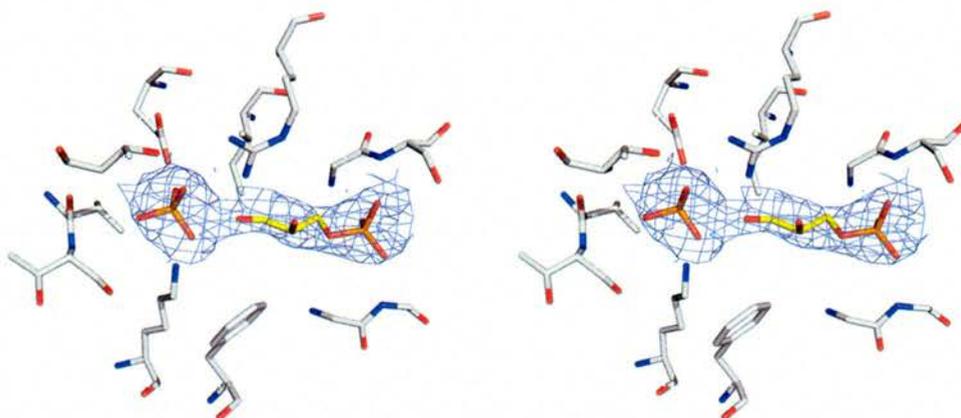


Figure 4.14 – Fully refined active site of KDPGal aldolase crystals soaked with G3P. The $2F_o-F_c$ density for obtained (steel blue mesh and 1.0σ) for the ligands is shown. There was no extra density obtained in the final F_o-F_c maps (3.0 or -3.0σ) which would suggest problems with the ligands. (C = white & yellow, N = blue, O = red, P = orange).

The G3P is orientated such that the hydroxyl and aldehyde atoms point towards Arg-14, the model suggests that the arginine is oriented such that interactions are shared between G3P, the phosphate anion and the protein. The B-factors for the amino acid residues in the active site are $\sim 50 \text{ \AA}^2$; however the B-factor for G3P atoms refine $\sim 65 \text{ \AA}^2$. This may be due to less than full occupancy of the ligand, possibly due to incomplete soaking or because of competition of phosphate for the phosphate binding site of G3P.

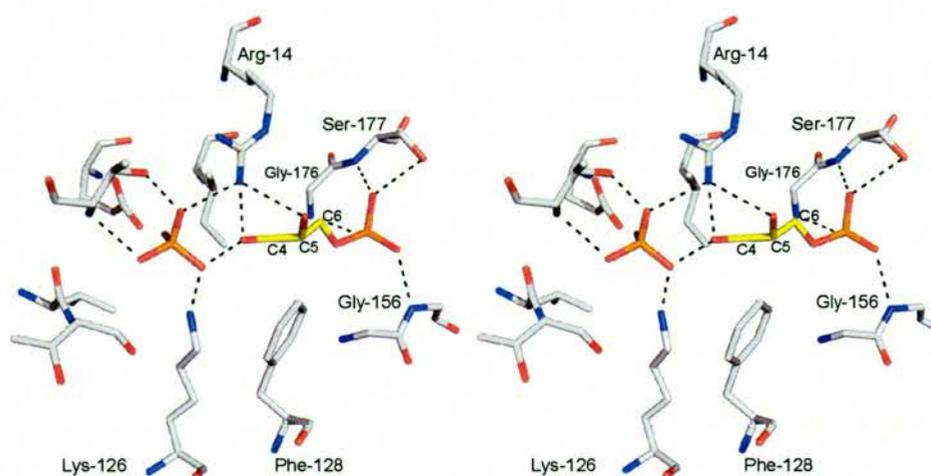


Figure 4.15 – Interactions between the G3P and phosphate to key residues in the active site of KDPGal aldolase (black dashes).

The data collection and refinement statistics are shown in Table 4.4 and the Ramachandran plot in Figure 4.16 suggests that the model quality is excellent.

Data collection	G3P
Wavelength (Å)	1.54178
Resolution (Highest Shell) (Å)	30.00 – 2.60 (2.71 – 2.60)
Space Group	H3
Cell constants (Å)	a = 105.50, b = 105.50, c = 74.30
Unique reflections	9411 (1341)
Average redundancy	2.3 (2.2)
I/σ	16.2 (4.6)
Completeness (%)	99.1 (95.8)
R _{merge} (%)	0.037 (0.255)
Refinement	
R	19.9
R _{free}	24.8
Rmsd bond (Å)/angle(°)	0.014 / 1.562
Residues in Ramachandran core (%)	100
PDB accession code	----

Table 4.4 – Summary of data collection and refinement statistics

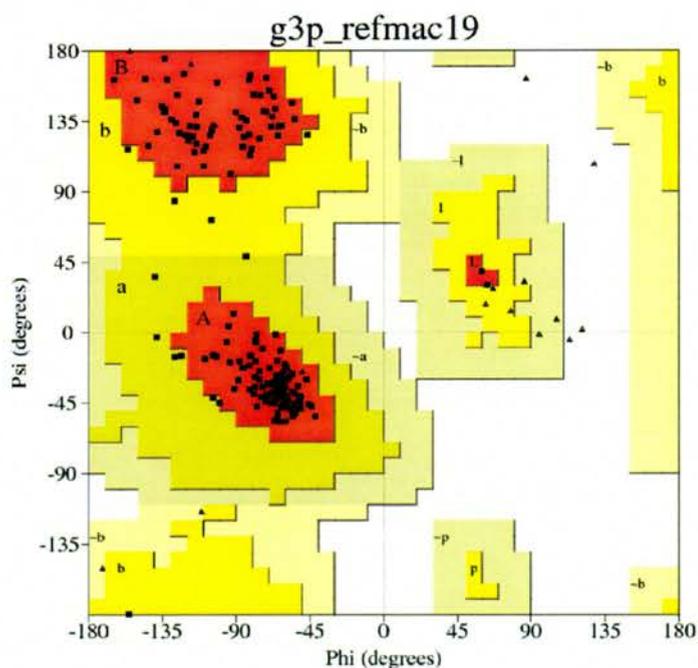


Figure 4.16 – Ramachandran plot of the G3P soaked KDPGal aldolase model

4.5.5 Co-crystal structure of KDPGal aldolase and KDPGal

KDPGal obtained from our collaborators (Eric Toone – Duke University) was soaked into native crystals using a solution of KDPGal (10 mM) prepared in stabilization buffer for 10 min at 298 K immediately prior to flash freezing. A dataset was collected measured to 2.5 Å resolution using a single crystal. The data were scaled and processed and the structure determined by molecular replacement from the existing pyruvate model coordinates. The unbiased F_o-F_c maps indicated a long blob of density into which KDPGal was modelled (shown in Figure 4.17) and refinement continued.

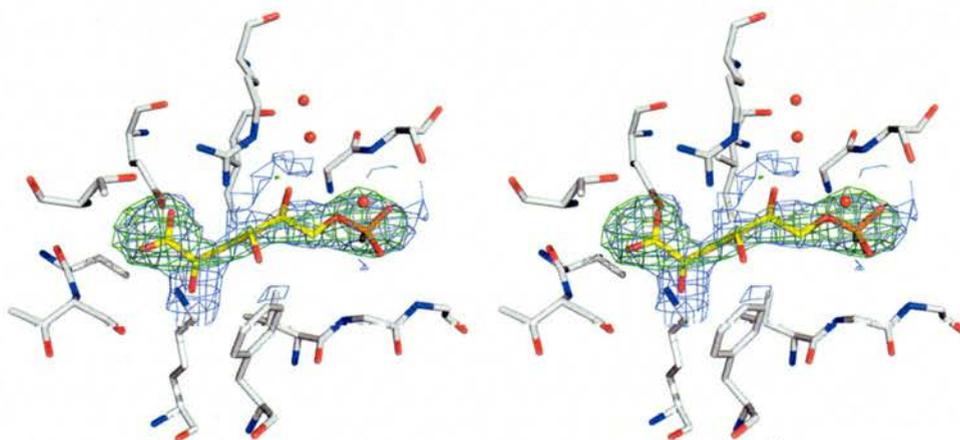


Figure 4.17 – KDPGal modelled into the unbiased F_o-F_c (green mesh and $+3.0 \sigma$) electron density map after molecular replacement with the existing protein model in the active site of KDPGal aldolase soaked with KDPGal. The $2F_o-F_c$ is shown also (steel blue mesh and 1.0σ).

As refinement continued the lysine moved out of position and discrepancies in the density suggested the production of the breakdown products G3P and pyruvate. There also appeared to be connectivity between the density at Lys-14 and the density above. As such, pyruvate and G3P were placed in the density and refinement resumed

indicating a much better fit for these ligands. The final density obtained is shown in Figure 4.18. There remains some connectivity in the density between pyruvate and G3P, however this may be explained by mixed species occupying the active sites. The data collection and refinement statistics are shown in Table 4.5 and the Ramachandran plot in Figure 4.23

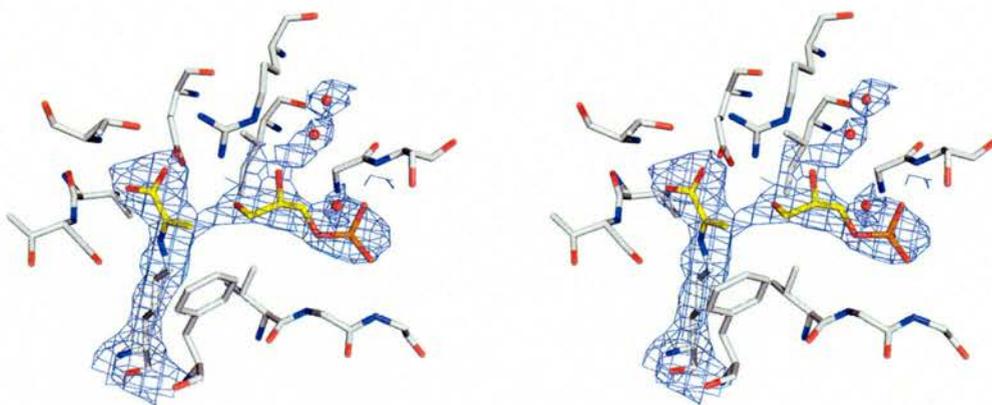


Figure 4.18 – Pyruvate and G3P fully refined into the active site of KDPGal aldolase, showing the final 2Fo-Fc electron density map (steel blue mesh and 1.0 σ).

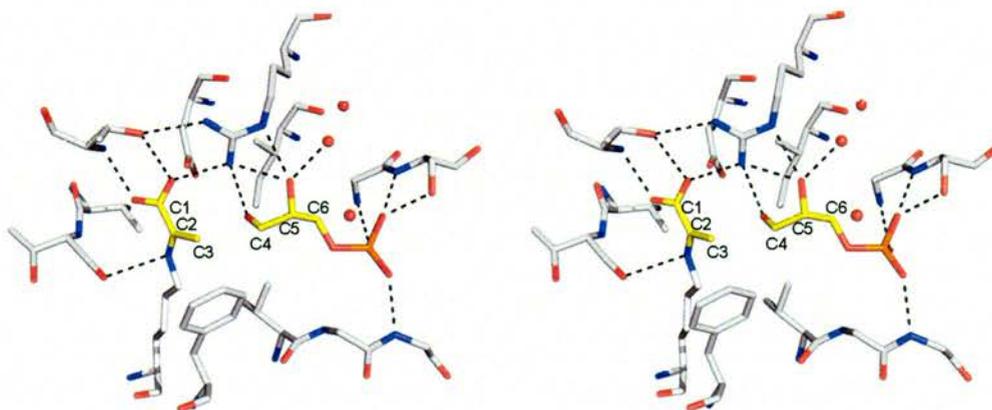


Figure 4.19 – Interactions between the G3P and phosphate to key residues in the active site of KDPGal aldolase (black dashes) in the model from crystals soaked with KDPGal. (Carbons are numbered according to their final numbering scheme in KDPGal).

Assuming that the active site contains only lysis products; Figure 4.19 shows the interactions between the ligands refined into the final density, the interactions are roughly equivalent to the G3P and pyruvate soak models. The interactions are also described by LIGPLOT diagram (Figure 4.20)

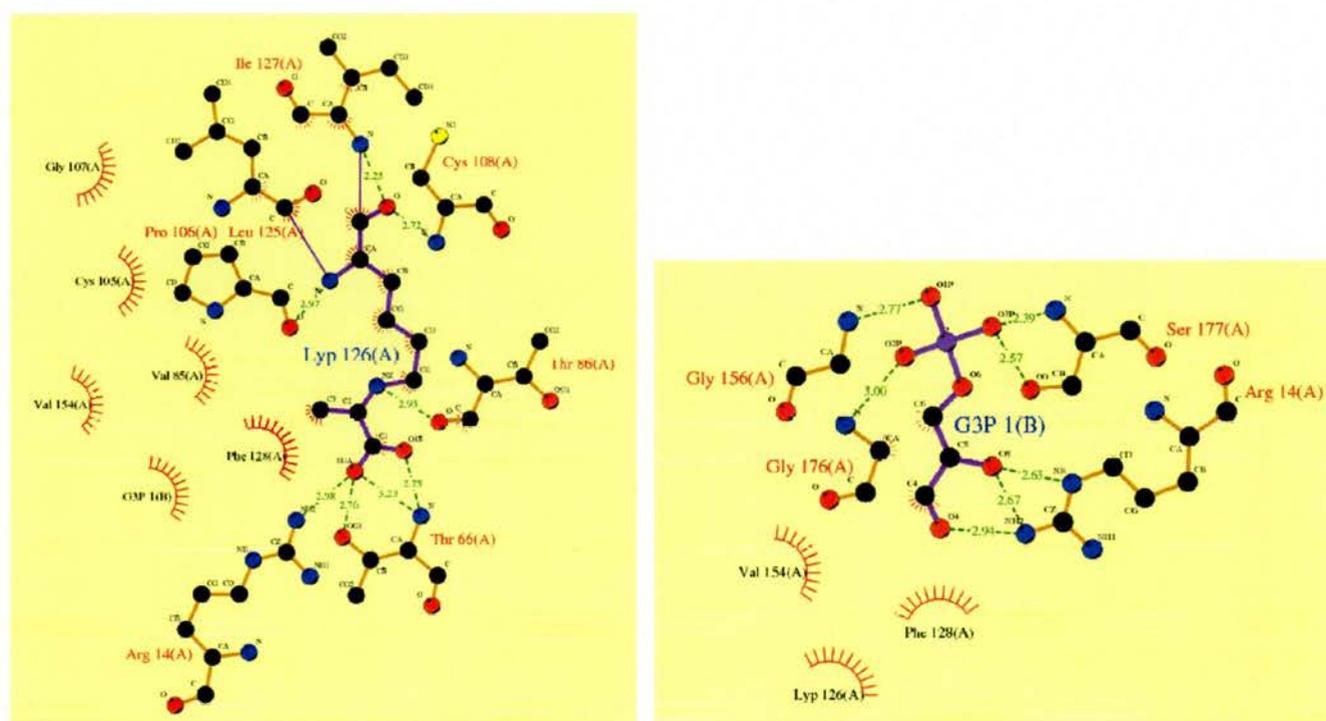


Figure 4.20 – LIGPLOT diagram of the interactions with respect to G3P and the pyruvate modified lysine (LYP).

However there is an apparent change between the conformation in the carbonyl of G3P (Figure 4.21). In the G3P soak model, the carbonyl appears to point towards the phosphate ion and away from Ile-12. The alternative G3P (KDPGal soak) points towards the active site arginine away from the pyruvate, to a position which may appear to favour attack from the Schiff base at the C4 position. The distance between C3 and C4 is ~ 2.8 Å at an angle of 40° (between C5-C4-C3), and a torsion of $\sim 120^\circ$ (C5-C4-C3-C2). This

would likely result in the production of KDPG as we would expect the carbonyl to be in the opposite direction. In the G3P soak model we find that the carbonyl is oriented as we may expect for attack by the anion for formation of KDPGal. Figure 4.21 shows the superimpositions of the KDPGal aldolase models and which highlights the difference between the G3P conformations.

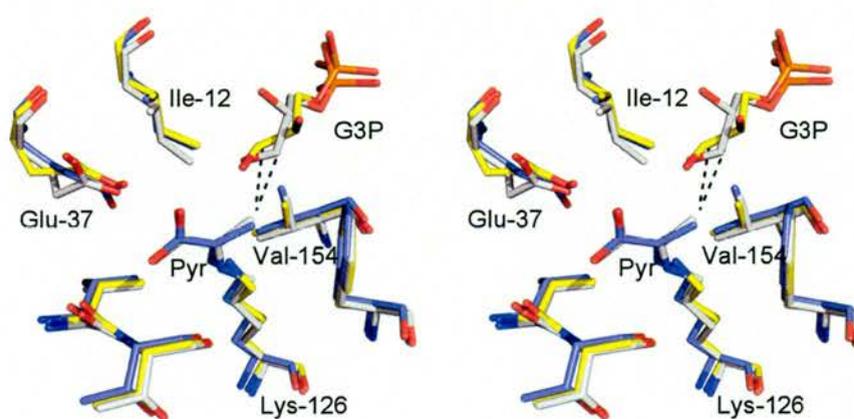


Figure 4.21 – Model superimposition of G3P soak (yellow), pyruvate (blue) and KDPGal soak (white). The possible attack trajectories of the pyruvate at C4 are shown (black dash).

The model is consistent with the ultimate stereochemistry of the resulting product (KDPGal). Figure 4.22 is the superimposition of *E. coli* KDPGal aldolase KDPGal soak model and *Thermotoga maritima* KDPG aldolase whereby the covalently bound product precursor is modelled into the active site. This confirms that the Val-154 and Ile-12 in the KDPGal aldolase (replaced by Thr-156 and Val-15 in KDPG aldolase) are important in conferring stereo-control as they will promote positioning of the carbonyl allowing stereospecific attack by the pyruvate. The complex model showing both products has likely been produced by re-positioning the G3P after lysis to produce an unreactive

complex. We would not expect to observe the Michaelis intermediate complex as this would be unstable. The electrostatic potential surface also identifies an extra pocket situated beside the G3P binding site in both proteins (Figure 4.22). This may be functionally significant or a recognition site for more complicated substrates. KDPG was also used in soaking experiments however no crystals have been obtained which survived soaking and experiments are ongoing to obtain the complex.

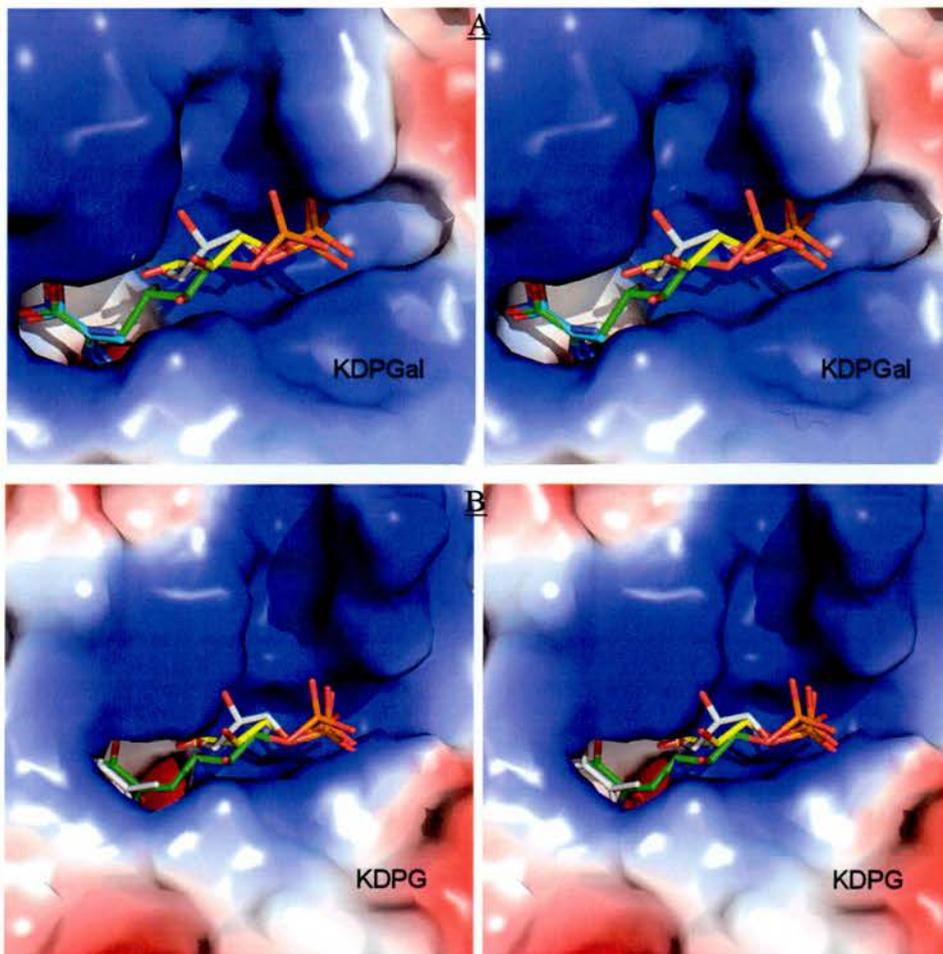


Figure 4.22 – The electron surface potential of A) KDPGal aldolase and B) KDPG aldolase showing G3P soak model (yellow), KDPGal soak model (white) and *Thermotoga maritima* KDPG aldolase (green) with the covalent product precursor modelled into the active site.

Data collection	KDPGal - short soak
Wavelength (Å)	1.541
Resolution (Highest Shell) (Å)	34.22 – 2.50 (2.64 – 2.50)
Space Group	H3
Cell constants (Å)	104.39, 104.39, 73.92
Unique reflections	10111 (2688)
Average redundancy	1.9 (1.8)
I/σ	11.0 (4.6)
Completeness (%)	97.4 (98.2)
R_{merge} (%)	7.2 (16.9)
Refinement	
R	0.177
R_{free}	0.243
Rmsd bond (Å)/angle(°)	0.016 / 1.683
Residues in Ramachandran core (%)	100.0
PDB accession code	----

Table 4.5 – Summary of data collection and refinement statistics

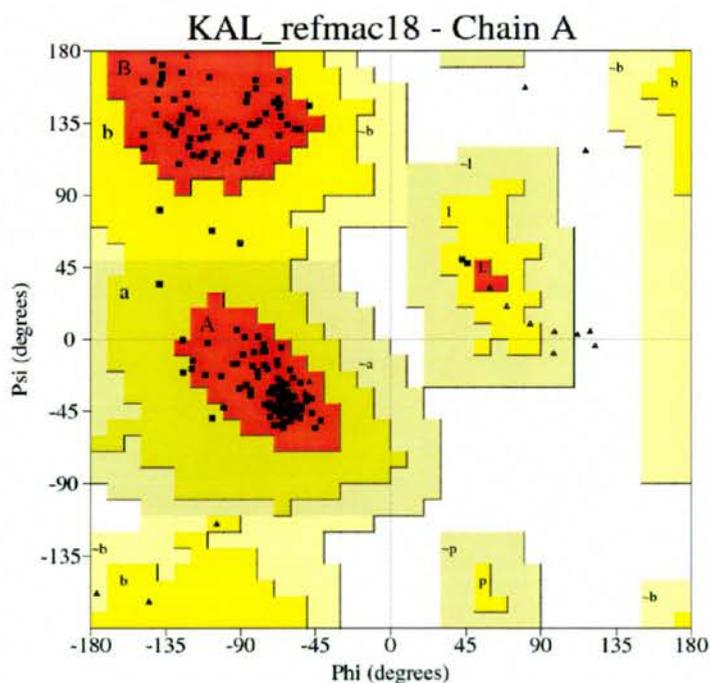


Figure 4.23 – Ramachandran plot of the KDPGal aldolase soaked with KDPGal model

4.6 CONCLUSIONS

This study shows some complex structures of KDPGal aldolase. Including a Schiff base intermediate which forms by covalent addition of pyruvate to the active site lysine. A G3P complex was obtained which shows a possible conformation which would agree with the stereo-specificity of the reaction assuming correct binding of the pyruvate. Finally a crystal soaked with KDPGal yielded the complex structure of the lytic products, showing the covalently modified lysine and G3P in the active site, as a product of lysis. The G3P appears to be in a conformation which would not favour formation of the KDPGal sugar. The enzyme is structurally similar to KDPG aldolase and the active site is almost identical. There is a conserved change in two residues from a valine to an isoleucine and a threonine to a valine. It is thought that these modifications will confer stereo selectivity by positioning the carbonyl of G3P.

4.7 FUTURE WORK

It is important to obtain both KDPGal or KDPG in the active site of crystal. This may only be possible by soaking conditions which do not favour turnover, for example lowering of the temperature. It has been shown in previous studies that mutation of active site residues may result in a change in substrate specificity and stereo-selectivity of the products. Alteration of such residues may confer new synthetic utilities for this class of enzyme. The Val-154-Thr, Ile-12-Val double mutant should be investigated as they represent the most obvious differences in the catalytic site of KDPGal and KDPG aldolase. Based on previous models of the active site of KDPG aldolases the mutant should revert the KDPGal aldolase to a KDPG aldolase, or else help better understand the reason for stereo-specificity.

Bibliography

Allard, J., Grochulski, P., and Sygusch, J. (2001). Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-Å resolution. *Proc Natl Acad Sci U S A* **98**, 3679-3684.

Almeida, M., Filipe, S., Humanes, M., Maia, M. F., Melo, R., Severino, N., da Silva, J. A., Frausto da Silva, J. J., and Wever, R. (2001). Vanadium haloperoxidases from brown algae of the Laminariaceae family. *Phytochemistry* **57**, 633-642.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Attieh, J. M., Hanson, A. D., and Saini, H. S. (1995). Purification and characterization of a novel methyltransferase responsible for biosynthesis of halomethanes and methanethiol in *Brassica oleracea*. *J Biol Chem* **270**, 9250-9257.

Beckwith, J. R., Clark, R., and Hager, L. P. (1963). Biological Chlorination. Vii. Studies on the Biosynthesis of Caldariomycin. *J Biol Chem* **238**, 3086-3090.

Bohm, H. J., Banner, D., Bendels, S., Kansy, M., Kuhn, B., Muller, K., Obst-Sander, U., and Stahl, M. (2004). Fluorine in medicinal chemistry. *Chembiochem* **5**, 637-643.

Bowman, J. P., McCammon, S. A., Nichols, D. S., Skerratt, J. H., Rea, S. M., Nichols, P. D., and McMeekin, T. A. (1997). *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5 omega 3) and grow anaerobically by dissimilatory Fe(III) reduction. *Int J Syst Bacteriol* **47**, 1040-1047.

Burd, W., Yourkevich, O., Voskoboev, A. J., and van Pee, K. H. (1995). Purification and properties of a non-haem chloroperoxidase from *Serratia marcescens*. *FEMS Microbiol Lett* **129**, 255-260.

Cadicamo, C. D., Courtieu, J., Deng, H., Meddour, A., and O'Hagan, D. (2004). Enzymatic fluorination in *Streptomyces cattleya* takes place with an inversion of configuration consistent with an S_N2 reaction mechanism. *Chembiochem* **5**, 685-690.

Canaves, J. M., Page, R., Wilson, I. A., and Stevens, R. C. (2004). Protein biophysical properties that correlate with crystallization success in *Thermotoga maritima*: maximum clustering strategy for structural genomics. *J Mol Biol* **344**, 977-991.

Christie, W. W., Hamilton, J. T. G., and Harper, D. B. (1998). Mass spectrometry of fluorinated fatty acids in the seed oil of *Dichapetalum toxicarium*. *Chemistry and Physics of Lipids* **97**, 41-47.

Cobb, S. L., Deng, H., Hamilton, J. T., McGlinchey, R. P., and O'Hagan, D. (2004). Identification of 5-fluoro-5-deoxy-D-ribose-1-phosphate as an intermediate in fluorometabolite biosynthesis in *Streptomyces cattleya*. *Chem Commun (Camb)*, 592-593.

Cobb, S. L., Deng, H., McEwan, A. R., Naismith, J. H., O'Hagan, D., and Robinson, D. A. (2006). Substrate specificity in enzymatic fluorination. The fluorinase from *Streptomyces cattleya* accepts 2'-deoxyadenosine substrates. *Org Biomol Chem* **4**, 1458-1460.

Conway, T., Fliege, R., Jones-Kilpatrick, D., Liu, J., Barnell, W. O., and Egan, S. E. (1991). Cloning, characterization and expression of the *Zymomonas mobilis* eda gene that encodes 2-keto-3-deoxy-6-phosphogluconate aldolase of the Entner-Doudoroff pathway. *Mol Microbiol* **5**, 2901-2911.

Cotterill, I. C., Henderson, D. P., Shelton, M. C., and Toone, E. J. (1998). The synthetic utility of KDPGal aldolase. *J Mol Catal B-Enzym* **5**, 103-111.

Dairi, T., Nakano, T., Aisaka, K., Katsumata, R., and Hasegawa, M. (1995). Cloning and nucleotide sequence of the gene responsible for chlorination of tetracycline. *Biosci Biotechnol Biochem* **59**, 1099-1106.

Deng, H., Cobb, S. L., McEwan, A. R., McGlinchey, R. P., Naismith, J. H., O'Hagan, D., Robinson, D. A., and Spencer, J. B. (2006). The fluorinase from *Streptomyces cattleya* is also a chlorinase. *Angew Chem Int Ed Engl* **45**, 759-762.

Deng, H., O'Hagan, D., and Schaffrath, C. (2004). Fluorometabolite biosynthesis and the fluorinase from *Streptomyces cattleya*. *Nat Prod Rep* **21**, 773-784.

Dong, C., Deng, H., Dorward, M., Schaffrath, C., O'Hagan, D., and Naismith, J. H. (2003). Crystallization and X-ray diffraction of 5'-fluoro-5'-deoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *Acta Crystallogr D Biol Crystallogr* **59**, 2292-2293.

Dong, C., Flecks, S., Unversucht, S., Haupt, C., van Pee, K. H., and Naismith, J. H. (2005). Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309**, 2216-2219.

Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D., and Naismith, J. H. (2004). Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **427**, 561-565.

Dreschel, E. (1886). Contribution to the chemistry of sea animal. *Zeitschrift fur Biologie* **33**, 85-107.

Espelt, L., Parella, T., Bujons, J., Solans, C., Joglar, J., Delgado, A., and Clapes, P. (2003). Stereoselective aldol additions catalyzed by dihydroxyacetone phosphate-dependent aldolases in emulsion systems: preparation and structural characterization of linear and cyclic iminopolyols from aminoaldehydes. *Chemistry* **9**, 4887-4899.

Evans, P. (2004). POINTLESS: A program for determining Laue groups (CCP4).

Evans, P. R. (1997). SCALA. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* **33**, 22-24.

Fullerton, S. W., Griffiths, J. S., Merkel, A. B., Cheriyan, M., Wymer, N. J., Hutchins, M. J., Fierke, C. A., Toone, E. J., and Naismith, J. H. (2006). Mechanism of the Class I KDPG aldolase. *Bioorg Med Chem* **14**, 3002-3010.

Garman, E. (1999). Cool data: quantity AND quality. *Acta Crystallogr D Biol Crystallogr* **55 (Pt 10)**, 1641-1653.

Garman, E., and Murray, J. W. (2003). Heavy-atom derivatization. *Acta Crystallogr D Biol Crystallogr* **59**, 1903-1913.

Gefflaut, T., Blonski, C., Perie, J., and Willson, M. (1995). Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. *Prog Biophys Mol Biol* **63**, 301-340.

Glukhovtsev, M. N., Pross, A., Schlegel, H. B., Bach, R. D., and Radom, L. (1996). Gas-Phase Identity S_N2 Reactions of Halides Anions and Methyl Halides with Retention of Configuration. *J Am Chem Soc* **118**, 11258-11264.

Griffiths, J. S., Wymer, N. J., Njolito, E., Niranjanakumari, S., Fierke, C. A., and Toone, E. J. (2002). Cloning, isolation and characterization of the *Thermotoga maritima* KDPG aldolase. *Bioorg Med Chem* **10**, 545-550.

Hall, D. R., Leonard, G. A., Reed, C. D., Watt, C. I., Berry, A., and Hunter, W. N. (1999). The crystal structure of Escherichia coli class II fructose-1, 6-bisphosphate aldolase in complex with phosphoglycolohydroxamate reveals details of mechanism and specificity. *J Mol Biol* **287**, 383-394.

Hamilton, J. T. G., Amin, M. R., Harper, D. H., and O'Hagan, D. (1997). Biosynthesis of fluoroacetate and 4-fluorothreonine by *Streptomyces cattleya*. Glycine and pyruvate as precursors. *Chem Commun (Camb)*, 797-798.

Hamilton, J. T. G., and Harper, D. B. (1997). Fluoro fatty acids in seed oil of *Dichapetalum toxicarium*. *Phytochemistry* **44**, 1129-1132.

Hamilton, J. T. G., Murphy, C. D., Amin, M. R., O'Hagan, D., and Harper, D. B. (1998). Exploring the biosynthetic origin of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *J Chem Soc, Perkin Trans 1*, 759-767.

Hammer, P. E., Hill, D. S., Lam, S. T., Van Pee, K. H., and Ligon, J. M. (1997). Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl Environ Microbiol* **63**, 2147-2154.

Harper, D. B., and O'Hagan, D. (1994). The fluorinated natural products. *Nat Prod Rep* **11**, 123-133.

Harper, D. B., O'Hagan, D., and Murphy, C. D. (2003). Fluorinated Natural Products: Occurrence and Biosynthesis. In *The Handbook of Environmental Chemistry* (Berlin Heidelberg, Springer-Verlag), pp. 141-169.

Heine, A., DeSantis, G., Luz, J. G., Mitchell, M., Wong, C. H., and Wilson, I. A. (2001). Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science* **294**, 369-374.

Henderson, D. P., Cotterill, I. C., Shelton, M. C., and Toone, E. J. (1998). 2-Keto-3-deoxy-6-phosphogalactonate Aldolase Catalyst for Stereocontrolled Carbon-Carbon Bond Formation. *J Org Chem* **63**, 906-907.

Horecker, B. L. (1972). Structure and possible regulatory properties of plant aldolases. *Biochem J* **128**, 28P.

Howell, P. L., Blessing, R. H., Smith, G. D., and Weeks, C. M. (2000). Optimizing DREAR and SnB parameters for determining Se-atom substructures. *Acta Crystallogr D Biol Crystallogr* **56**, 604-617.

Huang, F., Haydock, S. F., Spitteller, D., Mironenko, T., Li, T. L., O'Hagan, D., Leadlay, P. F., and Spencer, J. B. (2006). The gene cluster for fluorometabolite biosynthesis in *Streptomyces cattleya*: a thioesterase confers resistance to fluoroacetyl-coenzyme A. *Chem Biol* **13**, 475-484.

Kelly, S. M., Jess, T. J., and Price, N. C. (2005). How to study proteins by circular dichroism. *Biochim Biophys Acta* **1751**, 119-139.

Kelly, S. M., and Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci* **1**, 349-384.

Kirner, S., Hammer, P. E., Hill, D. S., Altmann, A., Fischer, I., Weislo, L. J., Lanahan, M., van Pee, K. H., and Ligon, J. M. (1998). Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J Bacteriol* **180**, 1939-1943.

Knoblauch, C., Sahm, K., and Jorgensen, B. B. (1999). Psychrophilic sulfate-reducing bacteria isolated from permanently cold arctic marine sediments: description of *Desulfofrigus oceanense* gen. nov., sp. nov., *Desulfofrigus fragile* sp. nov., *Desulfofaba gelida* gen. nov., sp. nov., *Desulfotalea psychrophila* gen. nov., sp. nov. and *Desulfotalea arctica* sp. nov. *Int J Syst Bacteriol* **49 Pt 4**, 1631-1643.

Kozbial, P. Z., and Mushegian, A. R. (2005). Natural history of S-adenosylmethionine-binding proteins. *BMC Struct Biol* **5**, 19.

Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-486.

Leslie, A. G. (1999). Integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **55 (Pt 10)**, 1696-1702.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* **26**.

Littlechild, J. (1999). Haloperoxidases and their role in biotransformation reactions. *Curr Opin Chem Biol* **3**, 28-34.

Liu, J. Q., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S., and Yamada, H. (1998). A novel metal-activated pyridoxal enzyme with a unique primary structure, low specificity D-threonine aldolase from *Arthrobacter* sp. Strain DK-38. Molecular cloning and cofactor characterization. *J Biol Chem* **273**, 16678-16685.

Machajewski, T. D., and Wong, C. H. (2000). The Catalytic Asymmetric Aldol Reaction. *Angew Chem Int Ed Engl* **39**, 1352-1375.

Marais, J. S. C. (1944). Monofluoroacetic acid, the toxic principle of "griffblaar" *Dichapetabum cymosum*. *Onderstepoort J Vet Sci Anim Indust* **20**, 606-613.

March, J., and Smith, M. B. (2001). *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 5 edn (New York, Wiley-Interscience).

- Marsh, J. J., and Lebherz, H. G. (1992). Fructose-bisphosphate aldolases: an evolutionary history. *Trends Biochem Sci* **17**, 110-113.
- Matthies, A., Ezziddin, S., Ulrich, E. M., Palmedo, H., Biersack, H. J., Bender, H., and Gohlke, S. (2004). Imaging of prostate cancer metastases with ^{18}F -fluoroacetate using PET/CT. *Eur J Nucl Med Mol Imaging* **31**, 797.
- Mavridis, I. M., Hatada, M. H., Tulinsky, A., and Lebiada, L. (1982). Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8 Å resolution. *J Mol Biol* **162**, 419-444.
- Mavridis, I. M., and Tulinsky, A. (1976). The folding and quaternary structure of trimeric 2-keto-3-deoxy-6-phosphogluconic aldolase at 3.5-Å resolution. *Biochemistry* **15**, 4410-4417.
- McClelland, M., and Nelson, M. (1992). Effect of site-specific methylation on DNA modification, methyltransferases and restriction endonucleases. *Nucleic Acids Res* **20**, 2145-2157.
- McEwan, A. R., Liu, H., Oke, M., Carter, L., Powers, H., Dorward, M., McMahon, S. A., White, M. F., and Naismith, J. H. (2006). Overexpression, purification, crystallization and data collection of *Sulfolobus solfataricus* Sso6206, a novel highly conserved protein. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **62**, 228-230.
- McGlinchey, R. P. (2006) Intermediates and enzymes involved in fluorometabolite biosynthesis *Streptomyces cattleya.*, PhD Thesis, University of St. Andrews, St. Andrews.
- Morris, D. R., and Hager, L. P. (1966). Chloroperoxidase. I. Isolation and properties of the crystalline glycoprotein. *J Biol Chem* **241**, 1763-1768.
- Murphy, C. D. (2003). New frontiers in biological halogenation. *J Appl Microbiol* **94**, 539-548.

Murphy, C. D., Moss, S. J., and O'Hagan, D. (2001a). Isolation of an aldehyde dehydrogenase involved in the oxidation of fluoroacetaldehyde to fluoroacetate in *Streptomyces cattleya*. *Appl Environ Microbiol* **67**, 4919-4921.

Murphy, C. D., O'Hagan, D., and Schaffrath, C. (2001b). Identification of a PLP-Dependent Threonine Transaldolase: A Novel Enzyme Involved in 4-Fluorothreonine Biosynthesis in *Streptomyces cattleya*. *Angew Chem Int Ed Engl* **40**, 4479-4481.

Murphy, C. D., Schaffrath, C., and O'Hagan, D. (2003). Fluorinated natural products: the biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Chemosphere* **52**, 455-461.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.

Naismith, J. H., and Liu, H. (2002). Unpublished work.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., *et al.* (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.

Ogawa, H., Gomi, T., and Fujioka, M. (2000). Serine hydroxymethyltransferase and threonine aldolase: are they identical? *Int J Biochem Cell Biol* **32**, 289-301.

O'Hagan, D., and Harper, D. B. (1999). Fluorine-containing natural products. *Journal of Fluorine Chemistry* **100**, 127-133.

O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T., and Murphy, C. D. (2002). Biochemistry: biosynthesis of an organofluorine molecule. *Nature* **416**, 279.

Oldfield, C. J., Ulrich, E. L., Cheng, Y., Dunker, A. K., and Markley, J. L. (2005). Addressing the intrinsic disorder bottleneck in structural proteomics. *Proteins* **59**, 444-453.

Paul, E. A., and Huang, P. M. (1980). Chemical aspects of soil. In *Handb. Environ. Chem.*, O. Hutzinger, ed. (Berlin, Springer), pp. 69-86.

Perrakis, A., Morris, R., and Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458-463.

Peters, R. A., and Shorthouse, M. (1967). Observations on the metabolism of fluoride in *Acacia georginae* and some other plants. *Nature* **216**, 80-81.

Peters, R. A., Wakelin, R. W., Martin, A. J., Webb, J., and Birks, F. T. (1959). Observations upon the toxic principle in the seeds of *Dichapetalum toxicarium*; separation of a long-chain fatty acid containing fluorine. *Biochem J* **71**, 245-248.

Pierce, M. M., Raman, C. S., and Nall, B. T. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods* **19**, 213-221.

Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **59**, 1131-1137.

Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002). The CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **58**, 1955-1957.

Ramagopal, U. A., Dauter, M., and Dauter, Z. (2003). Phasing on anomalous signal of sulfurs: what is the limit? *Acta Crystallogr D Biol Crystallogr* **59**, 1020-1027.

Ran, N., Draths, K. M., and Frost, J. W. (2004). Creation of a shikimate pathway variant. *J Am Chem Soc* **126**, 6856-6857.

Rodriguez-Aparicio, L. B., Ferrero, M. A., and Reglero, A. (1995). *N*-acetyl-D-neuraminic acid synthesis in *Escherichia coli* K1 occurs through condensation of *N*-acetyl-D-mannosamine and pyruvate. *Biochem J* **308** (Pt 2), 501-505.

Rutter, W. J. (1964). Evolution of Aldolase. *Fed Proc* **23**, 1248-1257.

Samland, A. K., and Sprenger, G. A. (2006). Microbial aldolases as C-C bonding enzymes-unknown treasures and new developments. *Appl Microbiol Biotechnol* **71**, 253-264.

Sanada, M., Miyano, T., Iwadare, S., Williamson, J. M., Arison, B. H., Smith, J. L., Douglas, A. W., Liesch, J. M., and Inamine, E. (1986). Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya*. *J Antibiot (Tokyo)* **39**, 259-265.

Schaffrath, C., Cobb, S. L., and O'Hagan, D. (2002). Cell-free biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Angew Chem Int Ed Engl* **41**, 3913-3915.

Schaffrath, C., Deng, H., and O'Hagan, D. (2003). Isolation and characterisation of 5'-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Lett* **547**, 111-114.

Schneider, T. R., and Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779.

Schoevaart, R., van Rantwijk, F., and Sheldon, R. A. (2000). Stereochemistry of nonnatural aldol reactions catalyzed by DHAP aldolases. *Biotechnol Bioeng* **70**, 349-352.

Schuttelkopf, A. W., and van Aalten, D. M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-1363.

Senn, H. M., O'Hagan, D., and Thiel, W. (2005). Insight into enzymatic C-F bond formation from QM and QM/MM calculations. *J Am Chem Soc* **127**, 13643-13655.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., *et al.* (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98**, 7835-7840.

Srikannathasan, V., and Naismith, J. H. (2005). Unpublished work.

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234.

Sun, P. D., Radaev, S., and Kattah, M. (2002). Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr D Biol Crystallogr* **58**, 1092-1098.

Sundaram, A. K., and Woodard, R. W. (2000). Probing the stereochemistry of *E. coli* 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (phenylalanine-sensitive)-catalyzed synthesis of KDO 8-P analogues. *J Org Chem* **65**, 5891-5897.

Sundaramoorthy, M., Turner, J., and Poulos, T. L. (1998). Stereochemistry of the chloroperoxidase active site: crystallographic and molecular-modeling studies. *Chem Biol* **5**, 461-473.

Taylor Ringia, E. A., and Schramm, V. L. (2005). Transition states and inhibitors of the purine nucleoside phosphorylase family. *Curr Top Med Chem* **5**, 1237-1258.

Terwilliger, T. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* **11**, 49-52.

Thomas, S. O., Singleton, V. L., Lowery, J. A., Sharpe, R. W., Pruess, L. M., Porter, J. N., Mowat, J. H., and Bohonos, N. (1956). Nucleocidin, a new antibiotic with activity against Trypanosomes. *Antibiot Annu*, 716-721.

Thompson, M. J., Mekhalfia, A., Hornby, D. P., and Blackburn, G. M. (1999). Synthesis of Two Stable Nitrogen Analogues of S-Adenosyl-L-methionine. *J Org Chem* **64**, 7467-7473.

Vagin, A., and Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56 Pt 12**, 1622-1624.

Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E., and Walsh, C. T. (2005a). Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191-1194.

Vaillancourt, F. H., Yin, J., and Walsh, C. T. (2005b). SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) alpha-ketoglutarate- and O₂-dependent halogenase. *Proc Natl Acad Sci U S A* **102**, 10111-10116.

van Pee, K. H. (1996). Biosynthesis of halogenated metabolites by bacteria. *Annu Rev Microbiol* **50**, 375-399.

van Pee, K. H., and Unversucht, S. (2003). Biological dehalogenation and halogenation reactions. *Chemosphere* **52**, 299-312.

Verlinde, C. L. M. J., and Quigley, P. M. (1999). Structure-based Re-evaluation of the Mechanism of Class I Fructose-1,6-bisphosphate Aldolase. *J Mol Model* **5**, 37-45.

Vidal, L., Ferrer, P., Alvaro, G., Benaiges, M. D., and Caminal, G. (2005). Influence of induction and operation mode on recombinant rhamnulose 1-phosphate aldolase production by *Escherichia coli* using the T5 promoter. *J Biotechnol* **118**, 75-87.

Vilter, H. (1995). Vanadium-dependent haloperoxidases. *Met Ions Biol Syst* **31**, 325-362.

Vincent, M. A., and Hillier, I. H. (2005). The solvated fluoride anion can be a good nucleophile. *Chem Commun (Camb)*, 5902-5903.

Voet, D., and Voet, J. G. (2004). *Biochemistry* (New York, John Wiley & Sons, Inc.).

Volkow, N. D., Mullani, N. A., and Bendriem, B. (1988). Positron emission tomography instrumentation: an overview. *Am J Physiol Imaging* **3**, 142-153.

von der Osten, C. H., Gioannetti, C., Barbas, C., Pederson, R. L., Wang, Y. F., Wong, C. H., Ozaki, A., Toone, E., Whitesides, G. M., and Sinskey, A. J. (1990). Molecular cloning of aldolases for synthetic applications. *Ann N Y Acad Sci* **613**, 771-775.

Wagenknecht, H. A., and Woggon, W. D. (1997). Identification of intermediates in the catalytic cycle of chloroperoxidase. *Chem Biol* **4**, 367-372.

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot Eng* **8**, 127-134.

Wiesner, W., van Pee, K. H., and Lingens, F. (1986). Detection of a new chloroperoxidase in *Pseudomonas pyrrocinia*. *FEBS Lett* **209**, 321-324.

Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374**, 300-321.

Witke, C., and Gotz, F. (1993). Cloning, sequencing, and characterization of the gene encoding the class I fructose-1,6-bisphosphate aldolase of *Staphylococcus carnosus*. *J Bacteriol* **175**, 7495-7499.

Wong, C. H., Liu, K. K., Kajimoto, T., Chen, L., Zhong, Z., Ichikawa, Y., and Shen, G. J. (1992). Developing better enzymes for organic synthesis. *Ann N Y Acad Sci* **672**, 343-351.

Wuosmaa, A. M., and Hager, L. P. (1990). Methyl chloride transferase: a carbocation route for biosynthesis of halometabolites. *Science* **249**, 160-162.

Xu, X. H., Yao, G. M., Li, Y. M., Lu, J. H., Lin, C. J., Wang, X., and Kong, C. H. (2003). 5-Fluorouracil derivatives from the sponge *Phakellia fusca*. *J Nat Prod* **66**, 285-288.

Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369-3376.

Zechel, D. L., Reid, S. P., Nashiru, O., Mayer, C., Stoll, D., Jakeman, D. L., Warren, R. A., and Withers, S. G. (2001). Enzymatic synthesis of carbon-fluorine bonds. *J Am Chem Soc* **123**, 4350-4351.

Zechel, D. L., and Withers, S. G. (2001). Dissection of nucleophilic and acid-base catalysis in glycosidases. *Curr Opin Chem Biol* **5**, 643-649.

Crystal screen 1+2

- A1. 30% MPD, 0.1 M Na Acetate pH 4.6, 0.02 M Calcium Chloride
- A2. 0.4 M K, Na Tartrate
- A3. 0.4 M Ammonium Phosphate
- A4. 2.0 M Ammonium Sulfate, 0.1 M Tris HCl pH 8.5
- A5. 30% MPD, 0.1 M Na Hepes pH 7.5, 0.2 M Sodium Citrate
- A6. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Magnesium Chloride
- A7. 1.4 M Sodium Acetate, 0.1 M Na Cacodylate pH 6.5
- A8. 30% iso-Propanol, 0.1 M Na Cacodylate pH 6.5, 0.2 M Sodium Citrate
- A9. 30% PEG 4000, 0.1 M Na Citrate pH 5.6, 0.2 M Ammonium Acetate
- A10. 30% PEG 4000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium Acetate
- A11. 1.0 M Ammonium Phosphate, 0.1 M Na Citrate pH 5.6
- A12. 30% iso-Propanol, 0.1 M Na Hepes pH 7.5, 0.2 M Magnesium Chloride
- B1. 30% PEG 400, 0.1 M Tris HCl pH 8.5, 0.2 M Sodium Citrate
- B2. 28% PEG 400, 0.1 M Na Hepes pH 7.5, 0.2 M Calcium Chloride
- B3. 30% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Ammonium Sulfate
- B4. 1.5 M Lithium Sulfate, 0.1 M Na Hepes pH 7.5
- B5. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Lithium Sulfate
- B6. 20% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Magnesium Acetate
- B7. 30% iso-Propanol, 0.1 M Tris HCl pH 8.5, 0.2 M Ammonium Acetate
- B8. 25% PEG 4000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium Sulfate
- B9. 30% MPD, 0.1 M Na Cacodylate pH 6.5, 0.2 M Magnesium Acetate
- B10. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Sodium Acetate
- B11. 30% PEG 400, 0.1 M Na Hepes pH 7.5, 0.2 M Magnesium Chloride
- B12. 20% iso-Propanol, 0.1 M Na Acetate pH 4.6, 0.2 M Calcium Chloride
- C1. 1.0 M Sodium Acetate, 0.1 M Imidazole pH 6.5
- C2. 30% MPD, 0.1 M Na Citrate pH 5.6, 0.2 M Ammonium Acetate
- C3. 20% iso-Propanol, 0.1 M Na Hepes pH 7.5, 0.2 M Sodium Citrate
- C4. 30% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Sodium Acetate
- C5. 0.8 M K, Na Tartrate, 0.1 M Na Hepes pH 7.5
- C6. 30% PEG 8000, 0.2 M Ammonium Sulfate
- C7. 30% PEG 4000, 0.2 M Ammonium Sulfate
- C8. 2.0 M Ammonium Sulfate
- C9. 4.0 M Sodium Formate
- C10. 2.0 M Sodium Formate, 0.1 M Na Acetate pH 4.6
- C11. 1.6 M Na, K Phosphate, 0.1 M Na Hepes pH 7.5
- C12. 8% PEG 8000, 0.1 M Tris HCl pH 8.5
- D1. 8% PEG 4000, 0.1 M Na Acetate pH 4.6
- D2. 1.4 M Sodium Citrate, 0.1 M Na Hepes pH 7.5
- D3. 2% PEG 400, 0.1 M Na Hepes pH 7.5, 2.0 M Ammonium Sulfate
- D4. 20% iso-Propanol, 0.1 M Na Citrate pH 5.6, 20% PEG 4000
- D5. 10% iso-Propanol, 0.1 M Na Hepes pH 7.5, 20% PEG 4000
- D6. 20% PEG 8000, 0.05 M Potassium Phosphate
- D7. 30% PEG 1500
- D8. 0.2 M Magnesium Formate
- D9. 18% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Zinc Acetate
- D10. 18% PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Calcium Acetate
- D11. 2.0 M Ammonium Sulfate, 0.1 M Sodium Acetate pH 4.6
- D12. 2.0 M Ammonium Phosphate, 0.1 M Tris HCl pH 8.5

Appendix

- E1. 10% PEG 6000, 2.0 M Sodium Chloride
- E2. 0.5 M Sodium Chloride, 0.01 M CTAB, 0.01 M Magnesium Chloride
- E3. 25% Ethylene Glycol
- E4. 35% Dioxane
- E5. 5% iso-Propanol, 2.0 M Ammonium Sulfate
- E6. 1.0 M Imidazole pH 7.0
- E7. 10% PEG 1000, 10% PEG 8000
- E8. 10% Ethanol, 1.5 M Sodium Chloride
- E9. 2.0 M Sodium Chloride, 0.1 M Na Acetate pH 4.6
- E10. 30% MPD, 0.1 M Na Acetate pH 4.6, 0.2 M Sodium Chloride
- E11. 1.0 M 1,6 Hexanediol, 0.1 M Na Acetate pH 4.6, 0.01 M Cobalt Chloride
- E12. 30% PEG 400, 0.1 M Na Acetate pH 4.6, 0.1 M Cadmium Chloride
- F1. 30% PEG MME 2000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium Sulfate
- F2. 2.0 M Ammonium Sulfate, 0.1 M Na Citrate pH 5.6, 0.2 M K/Na Tartrate
- F3. 1.0 M Lithium Sulfate, 0.1 M Na Citrate pH 5.6, 0.5 M Ammonium Sulfate
- F4. 2% Polyethyleneimine, 0.1 M Na Citrate pH 5.6, 0.5 M Sodium Chloride
- F5. 35% tert-Butanol, 0.1 M Na Citrate pH 5.6
- F6. 10% Jeffamine M-600, 0.1 M Na Citrate pH 5.6, 0.01 M Ferric Chloride
- F7. 2.5 M 1,6 Hexanediol, 0.1 M Na Citrate pH 5.6
- F8. 1.6 M Magnesium Sulfate, 0.1 M MES pH 6.5
- F9. 2.0 M Sodium Chloride, 0.1 M MES pH 6.5, 0.2 M Na/K Phosphate
- F10. 12% PEG 20,000, 0.1 M MES pH 6.5
- F11. 10% Dioxane, 0.1 M MES pH 6.5, 1.6 M Ammonium Sulfate
- F12. 30% Jeffamine M-600, 0.1 M MES pH 6.5, 0.05 M Cesium Chloride
- G1. 1.8 M Ammonium Sulfate, 0.1 M MES pH 6.5, 0.01 M Cobalt Chloride
- G2. 30% PEG MME 5000, 0.1 M MES pH 6.5, 0.2 M Ammonium Sulfate
- G3. 25% PEG MME 550, 0.1 M MES pH 6.5, 0.01 M Zinc Sulfate
- G4. 1.6 M Sodium Citrate pH 6.5
- G5. 30% MPD, 0.1 M HEPES pH 7.5, 0.5 M Ammonium Sulfate
- G6. 10% PEG 6000, 0.1 M HEPES pH 7.5, 5% MPD
- G7. 20% Jeffamine M-600, 0.1 M HEPES pH 7.5
- G8. 1.6 M Ammonium Sulfate, 0.1 M HEPES pH 7.5, 0.1 M Sodium Chloride
- G9. 2.0 M Ammonium Formate, 0.1 M HEPES pH 7.5
- G10. 1.0 M Sodium Acetate, 0.1 M HEPES pH 7.5, 0.05 M Cadmium Sulfate
- G11. 70% MPD, 0.1 M HEPES pH 7.5
- G12. 4.3 M Sodium Chloride, 0.1 M HEPES pH 7.5
- H1. 10% PEG 8000, 0.1 M HEPES pH 7.5, 8% Ethylene Glycol
- H2. 20% PEG 10,000, 0.1 M HEPES pH 7.5
- H3. 3.4 M 1,6 Hexanediol, 0.1 M Tris pH 8.5, 0.2 M Magnesium Chloride
- H4. 25% tert-Butanol, 0.1 M Tris pH 8.5
- H5. 1.0 M Lithium Sulfate, 0.1 M Tris pH 8.5, 0.01 M Nickel (II) Chloride
- H6. 12% Glycerol, 0.1 M Tris pH 8.5, 1.5 M Ammonium Sulfate
- H7. 50% MPD, 0.1 M Tris pH 8.5, 0.2 M Ammonium Phosphate
- H8. 20% Ethanol, 0.1 M Tris pH 8.5
- H9. 20% PEG MME 2000, 0.1 M Tris pH 8.5, 0.01 M Nickel (II) Chloride
- H10. 20% PEG MME 550, 0.1 M Bicine pH 9.0, 0.1 M Sodium Chloride
- H11. 2.0 M Magnesium Chloride, 0.1 M Bicine pH 9.0
- H12. 10% PEG 20,000, 0.1 M Bicine pH 9.0, 2% Dioxane

Wizard I

	Crystallant	buffer (0.1 M)	salt (0.2 M)
1	20% (w/v) PEG-8000	CHES pH 9.5	none
2	10% (v/v) 2-propanol	HEPES pH 7.5	NaCl
3	15% (v/v) ethanol	CHES pH 9.5	none
4	35% (v/v) 2-methyl-2,4-pentanediol	imidazole pH 8.0	MgCl ₂
5	30% (v/v) PEG-400	CAPS pH 10.5	none
6	20% (w/v) PEG-3000	citrate pH 5.5	none
7	10% (w/v) PEG-8000	MES pH 6.0	Zn(OAc) ₂
8	2.0 M (NH ₄) ₂ SO ₄	citrate pH 5.5	none
9	1.0 M (NH ₄) ₂ HPO ₄	acetate pH 4.5	none
10	20% (w/v) PEG-2000 MME	Tris pH 7.0	none
11	20% (v/v) 1,4-butanediol	MES pH 6.0	Li ₂ SO ₄
12	20% (w/v) PEG-1000	imidazole pH 8.0	Ca(OAc) ₂
13	1.26 M (NH ₄) ₂ SO ₄	cacodylate pH 6.5	none
14	1.0 M sodium citrate	cacodylate pH 6.5	none
15	10% (w/v) PEG-3000	imidazole pH 8.0	Li ₂ SO ₄
16	2.5 M NaCl	Na/K phosphate pH 6.2	none
17	30% (w/v) PEG-8000	acetate pH 4.5	Li ₂ SO ₄
18	1.0 M K/Na tartrate	imidazole pH 8.0	NaCl
19	20% (w/v) PEG-1000	Tris pH 7.0	none
20	0.4 M NaH ₂ PO ₄ /1.6 M K ₂ HPO ₄	imidazole pH 8.0	NaCl
21	20% (w/v) PEG-8000	HEPES pH 7.5	none
22	10% (v/v) 2-propanol	Tris pH 8.5	none
23	15% (v/v) ethanol	imidazole pH 8.0	MgCl ₂
24	35% (v/v) 2-methyl-2,4-pentanediol	Tris pH 7.0	NaCl
25	30% (v/v) PEG-400	Tris pH 8.5	MgCl ₂
26	10% (w/v) PEG-3000	CHES pH 9.5	none
27	1.2 M NaH ₂ PO ₄ /0.8 M K ₂ HPO ₄	CAPS pH 10.5	Li ₂ SO ₄
28	20% (w/v) PEG-3000	HEPES pH 7.5	NaCl
29	10% (w/v) PEG-8000	CHES pH 9.5	NaCl
30	1.26 M (NH ₄) ₂ SO ₄	acetate pH 4.5	NaCl
31	20% (w/v) PEG-8000	phosphate-citrate pH 4.2	NaCl
32	10% (w/v) PEG-3000	Na/K phosphate pH 6.2	none
33	2.0 M (NH ₄) ₂ SO ₄	CAPS pH 10.5	Li ₂ SO ₄
34	1.0 M (NH ₄) ₂ HPO ₄	imidazole pH 8.0	none
35	20% (v/v) 1,4-butanediol	acetate pH 4.5	none
36	1.0 M sodium citrate	imidazole pH 8.0	none
37	2.5 M NaCl	imidazole pH 8.0	none
38	1.0 M K/Na tartrate	CHES pH 9.5	Li ₂ SO ₄
39	20% (w/v) PEG-1000	phosphate-citrate pH 4.2	Li ₂ SO ₄
40	10% (v/v) 2-propanol	MES pH 6.0	Ca(OAc) ₂
41	30% (w/v) PEG-3000	CHES pH 9.5	none
42	15% (v/v) ethanol	Tris pH 7.0	none
43	35% (v/v) 2-methyl-2,4-pentanediol	Na/K phosphate pH 6.2	none
44	30% (v/v) PEG-400	acetate pH 4.5	Ca(OAc) ₂
45	20% (w/v) PEG-3000	acetate pH 4.5	none
46	10% (w/v) PEG-8000	imidazole pH 8.0	Ca(OAc) ₂
47	1.26 M (NH ₄) ₂ SO ₄	Tris pH 8.5	Li ₂ SO ₄
48	20% (w/v) PEG-1000	acetate pH 4.5	Zn(OAc) ₂

Wizard II

	Crystallant	buffer (0.1 M)	salt (0.2 M)
1	10% (w/v) PEG-3000	acetate pH 4.5	Zn(OAc) ₂
2	35% (v/v) 2-methyl-2,4-pentanediol	MES pH 6.0	Li ₂ SO ₄
3	20% (w/v) PEG-8000	Tris pH 8.5	MgCl ₂
4	2.0 M (NH ₄) ₂ SO ₄	cacodylate pH 6.5	NaCl
5	20% (v/v) 1,4-butanediol	HEPES pH 7.5	NaCl
6	10% (v/v) 2-propanol	phosphate-citrate pH 4.2	Li ₂ SO ₄
7	30% (w/v) PEG-3000	Tris pH 7.0	NaCl
8	10% (w/v) PEG-8000	Na/K phosphate pH 6.2	NaCl
9	2.0 M (NH ₄) ₂ SO ₄	phosphate-citrate pH 4.2	none
10	1.0 M (NH ₄) ₂ HPO ₄	Tris pH 8.5	none
11	10% (v/v) 2-propanol	cacodylate pH 6.5	Zn(OAc) ₂
12	30% (v/v) PEG-400	cacodylate pH 6.5	Li ₂ SO ₄
13	15% (v/v) ethanol	citrate pH 5.5	Li ₂ SO ₄
14	20% (w/v) PEG-1000	Na/K phosphate pH 6.2	NaCl
15	1.26 M (NH ₄) ₂ SO ₄	HEPES pH 7.5	none
16	1.0 M sodium citrate	CHES pH 9.5	none
17	2.5 M NaCl	Tris pH 7.0	MgCl ₂
18	20% (w/v) PEG-3000	Tris pH 7.0	Ca(OAc) ₂
19	1.6 M NaH ₂ PO ₄ /0.4 M K ₂ HPO ₄	phosphate-citrate pH 4.2	none
20	15% (v/v) ethanol	MES pH 6.0	Zn(OAc) ₂
21	35% (v/v) 2-methyl-2,4-pentanediol	acetate pH 4.5	none
22	10% (v/v) 2-propanol	imidazole pH 8.0	none
23	15% (v/v) ethanol	HEPES pH 7.5	MgCl ₂
24	30% (w/v) PEG-8000	imidazole pH 8.0	NaCl
25	35% (v/v) 2-methyl-2,4-pentanediol	HEPES pH 7.5	NaCl
26	30% (v/v) PEG-400	CHES pH 9.5	none
27	10% (w/v) PEG-3000	cacodylate pH 6.5	MgCl ₂
28	20% (w/v) PEG-8000	MES pH 6.0	Ca(OAc) ₂
29	1.26 M (NH ₄) ₂ SO ₄	CHES pH 9.5	NaCl
30	20% (v/v) 1,4-butanediol	imidazole pH 8.0	Zn(OAc) ₂
31	1.0 M sodium citrate	Tris pH 7.0	NaCl
32	20% (w/v) PEG-1000	Tris pH 8.5	none
33	1.0 M (NH ₄) ₂ HPO ₄	citrate pH 5.5	NaCl
34	10% (w/v) PEG-8000	imidazole pH 8.0	none
35	0.8 M NaH ₂ PO ₄ /1.2 M K ₂ HPO ₄	acetate pH 4.5	none
36	10% (w/v) PEG-3000	phosphate-citrate pH 4.2	NaCl
37	1.0 M K/Na tartrate	Tris pH 7.0	Li ₂ SO ₄
38	2.5 M NaCl	acetate pH 4.5	Li ₂ SO ₄
39	20% (w/v) PEG-8000	CAPS pH 10.5	NaCl
40	20% (w/v) PEG-3000	imidazole pH 8.0	Zn(OAc) ₂
41	2.0 M (NH ₄) ₂ SO ₄	Tris pH 7.0	Li ₂ SO ₄
42	30% (v/v) PEG-400	HEPES pH 7.5	NaCl
43	10% (w/v) PEG-8000	Tris pH 7.0	MgCl ₂
44	20% (w/v) PEG-1000	cacodylate pH 6.5	MgCl ₂
45	1.26 M (NH ₄) ₂ SO ₄	MES pH 6.0	none
46	1.0 M (NH ₄) ₂ HPO ₄	imidazole pH 8.0	NaCl
47	2.5 M NaCl	imidazole pH 8.0	Zn(OAc) ₂
48	1.0 M K/Na tartrate	MES pH 6.0	none

The Peg's Suite

No.	Salt	Buffer	Precipitant
1		0.1 M Sodium acetate pH 4.6	40 % (v/v) PEG 200
2		0.1 M Sodium acetate pH 4.6	30 % (v/v) PEG 300
3		0.1 M Sodium acetate pH 4.6	30 % (v/v) PEG 400
4		0.1 M Sodium acetate pH 4.6	25 % (v/v) PEG 550 MME
5		0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 1000
6		0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 2000 MME
7		0.1 M MES pH 6.5	40 % (v/v) PEG 200
8		0.1 M MES pH 6.5	30 % (v/v) PEG 300
9		0.1 M MES pH 6.5	30 % (v/v) PEG 400
10		0.1 M MES pH 6.5	25 % (v/v) PEG 550 MME
11		0.1 M MES pH 6.5	25 % (w/v) PEG 1000
12		0.1 M MES pH 6.5	25 % (w/v) PEG 2000 MME
13		0.1 M HEPES pH 7.5	40 % (v/v) PEG 200
14		0.1 M HEPES pH 7.5	30 % (v/v) PEG 300
15		0.1 M HEPES pH 7.5	30 % (v/v) PEG 400
16		0.1 M HEPES pH 7.5	25 % (v/v) PEG 550 MME
17		0.1 M HEPES pH 7.5	25 % (w/v) PEG 1000
18		0.1 M HEPES pH 7.5	25 % (w/v) PEG 2000 MME
19		0.1 M TRIS.HCl pH 8.5	40 % (v/v) PEG 200
20		0.1 M TRIS.HCl pH 8.5	30 % (v/v) PEG 300
21		0.1 M TRIS.HCl pH 8.5	30 % (v/v) PEG 400
22		0.1 M TRIS.HCl pH 8.5	25 % (v/v) PEG 550 MME
23		0.1 M TRIS.HCl pH 8.5	25 % (w/v) PEG 1000
24		0.1 M TRIS.HCl pH 8.5	25 % (w/v) PEG 2000 MME
25		0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 3000
26		0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 4000
27		0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 6000
28		0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 8000
29		0.1 M Sodium acetate pH 4.6	20 % (w/v) PEG 10000
30		0.1 M Sodium acetate pH 4.6	15 % (w/v) PEG 20000
31		0.1 M MES pH 6.5	25 % (w/v) PEG 3000
32		0.1 M MES pH 6.5	25 % (w/v) PEG 4000
33		0.1 M MES pH 6.5	25 % (w/v) PEG 6000
34		0.1 M MES pH 6.5	25 % (w/v) PEG 8000
35		0.1 M MES pH 6.5	20 % (w/v) PEG 10000
36		0.1 M MES pH 6.5	15 % (w/v) PEG 20000
37		0.1 M HEPES pH 7.5	25 % (w/v) PEG 3000
38		0.1 M HEPES pH 7.5	25 % (w/v) PEG 4000
39		0.1 M HEPES pH 7.5	25 % (w/v) PEG 6000
40		0.1 M HEPES pH 7.5	25 % (w/v) PEG 8000
41		0.1 M HEPES pH 7.5	20 % (w/v) PEG 10000
42		0.1 M HEPES pH 7.5	15 % (w/v) PEG 20000
43		0.1 M TRIS.HCl pH 8.5	25 % (w/v) PEG 3000
44		0.1 M TRIS.HCl pH 8.5	25 % (w/v) PEG 4000
45		0.1 M TRIS.HCl pH 8.5	25 % (w/v) PEG 6000
46		0.1 M TRIS.HCl pH 8.5	25 % (w/v) PEG 8000
47		0.1 M TRIS.HCl pH 8.5	20 % (w/v) PEG 10000

48		0.1 M TRIS.HCl pH 8.5	15 %(w/v) PEG 20000
49	0.2 M Sodium fluoride		20 %(w/v) PEG 3350
50	0.2 M Potassium fluoride		20 %(w/v) PEG 3350
51	0.2 M Ammonium fluoride		20 %(w/v) PEG 3350
52	0.2 M Lithium chloride		20 %(w/v) PEG 3350
53	0.2 M Magnesium chloride		20 %(w/v) PEG 3350
54	0.2 M Sodium chloride		20 %(w/v) PEG 3350
55	0.2 M Calcium chloride		20 %(w/v) PEG 3350
56	0.2 M Potassium chloride		20 %(w/v) PEG 3350
57	0.2 M Ammonium chloride		20 %(w/v) PEG 3350
58	0.2 M Sodium iodide		20 %(w/v) PEG 3350
59	0.2 M Potassium iodide		20 %(w/v) PEG 3350
60	0.2 M Ammonium iodide		20 %(w/v) PEG 3350
61	0.2 M Sodium thiocyanate		20 %(w/v) PEG 3350
62	0.2 M Potassium thiocyanate		20 %(w/v) PEG 3350
63	0.2 M Lithium nitrate		20 %(w/v) PEG 3350
64	0.2 M Magnesium nitrate		20 %(w/v) PEG 3350
65	0.2 M Sodium nitrate		20 %(w/v) PEG 3350
66	0.2 M Potassium nitrate		20 %(w/v) PEG 3350
67	0.2 M Ammonium nitrate		20 %(w/v) PEG 3350
68	0.2 M Magnesium formate		20 %(w/v) PEG 3350
69	0.2 M Sodium formate		20 %(w/v) PEG 3350
70	0.2 M Potassium formate		20 %(w/v) PEG 3350
71	0.2 M Ammonium formate		20 %(w/v) PEG 3350
72	0.2 M Lithium acetate		20 %(w/v) PEG 3350
73	0.2 M Magnesium acetate		20 %(w/v) PEG 3350
74	0.2 M Zinc acetate		20 %(w/v) PEG 3350
75	0.2 M Sodium acetate		20 %(w/v) PEG 3350
76	0.2 M Calcium acetate		20 %(w/v) PEG 3350
77	0.2 M Potassium acetate		20 %(w/v) PEG 3350
78	0.2 M Ammonium acetate		20 %(w/v) PEG 3350
79	0.2 M Lithium sulfate		20 %(w/v) PEG 3350
80	0.2 M Magnesium sulfate		20 %(w/v) PEG 3350
81	0.2 M Sodium sulfate		20 %(w/v) PEG 3350
82	0.2 M Potassium sulfate		20 %(w/v) PEG 3350
83	0.2 M Ammonium sulfate		20 %(w/v) PEG 3350
84	0.2 M di-Sodium tartate		20 %(w/v) PEG 3350
85	0.2 M K/Na tartrate		20 %(w/v) PEG 3350
86	0.2 M di-Ammonium tartrate		20 %(w/v) PEG 3350
87	0.2 M Sodium phosphate		20 %(w/v) PEG 3350
88	0.2 M di-Sodium phosphate		20 %(w/v) PEG 3350
89	0.2 M Potassium phosphate		20 %(w/v) PEG 3350
90	0.2 M di-Potassium phosphate		20 %(w/v) PEG 3350
91	0.2 M Ammonium phosphate		20 %(w/v) PEG 3350
92	0.2 M di-Ammonium phosphate		20 %(w/v) PEG 3350
93	0.2 M tri-Lithium citrate		20 %(w/v) PEG 3350
94	0.2 M tri-Sodium citrate		20 %(w/v) PEG 3350
95	0.2 M tri-Potassium citrate		20 %(w/v) PEG 3350
96	0.2 M di-Ammonium citrate		20 %(w/v) PEG 3350

Premier (In house)

P1-1	0.1M SPG buffer pH 4	-	25% w/v PEG 1500
P1-2	0.1M SPG buffer pH 5	-	25% w/v PEG 1500
P1-3	0.1M SPG buffer pH 6	-	25% w/v PEG 1500
P1-4	0.1M SPG buffer pH 7	-	25% w/v PEG 1500
P1-5	0.1M SPG buffer pH 8	-	25% w/v PEG 1500
P1-6	0.1M SPG buffer pH 9	-	25% w/v PEG 1500
P1-7	0.2M Sodium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
P1-8	0.2M Ammonium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
P1-9	0.2M Lithium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
P1-10	0.2M Magnesium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
P1-11	0.2M Calcium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
P1-12	0.01M Zinc chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
P1-13	0.1M MIB buffer pH 4	-	25% w/v PEG 1500
P1-14	0.1M MIB buffer pH 5	-	25% w/v PEG 1500
P1-15	0.1M MIB buffer pH 6	-	25% w/v PEG 1500
P1-16	0.1M MIB buffer pH 7	-	25% w/v PEG 1500
P1-17	0.1M MIB buffer pH 8	-	25% w/v PEG 1500
P1-18	0.1M MIB buffer pH 9	-	25% w/v PEG 1500
P1-19	0.2M Sodium chloride	0.1M MES pH 6	20% w/v PEG 6000
P1-20	0.2M Ammonium chloride	0.1M MES pH 6	20% w/v PEG 6000
P1-21	0.2M Lithium chloride	0.1M MES pH 6	20% w/v PEG 6000
P1-22	0.2M Magnesium chloride	0.1M MES pH 6	20% w/v PEG 6000
P1-23	0.2M Calcium chloride	0.1M MES pH 6	20% w/v PEG 6000
P1-24	0.01M Zinc chloride	0.1M MES pH 6	20% w/v PEG 6000
P1-25	0.1M PCB buffer pH 4	-	25% w/v PEG 1500
P1-26	0.1M PCB buffer pH 5	-	25% w/v PEG 1500
P1-27	0.1M PCB buffer pH 6	-	25% w/v PEG 1500
P1-28	0.1M PCB buffer pH 7	-	25% w/v PEG 1500
P1-29	0.1M PCB buffer pH 8	-	25% w/v PEG 1500
P1-30	0.1M PCB buffer pH 9	-	25% w/v PEG 1500
P1-31	0.2M Sodium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
P1-32	0.2M Ammonium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
P1-33	0.2M Lithium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
P1-34	0.2M Magnesium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
P1-35	0.2M Calcium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
P1-36	0.01M Zinc chloride	0.1M Hepes pH 7	20% w/v PEG 6000
P1-37	0.1M MMT buffer pH 4	-	25% w/v PEG 1500
P1-38	0.1M MMT buffer pH 5	-	25% w/v PEG 1500
P1-39	0.1M MMT buffer pH 6	-	25% w/v PEG 1500
P1-40	0.1M MMT buffer pH 7	-	25% w/v PEG 1500
P1-41	0.1M MMT buffer pH 8	-	25% w/v PEG 1500
P1-42	0.1M MMT buffer pH 9	-	25% w/v PEG 1500
P1-43	0.2M Sodium chloride	0.1M Tris pH 8	20% w/v PEG 6000
P1-44	0.2M Ammonium chloride	0.1M Tris pH 8	20% w/v PEG 6000
P1-45	0.2M Lithium chloride	0.1M Tris pH 8	20% w/v PEG 6000
P1-46	0.2M Magnesium chloride	0.1M Tris pH 8	20% w/v PEG 6000
P1-47	0.2M Calcium chloride	0.1M Tris pH 8	20% w/v PEG 6000
P1-48	0.01M Zinc chloride	0.1M Tris pH 8	20% w/v PEG 6000

Appendix

P2-1	0.2M Sodium fluoride	-	20% w/v PEG 3350
P2-2	0.2M Sodium bromide	-	20% w/v PEG 3350
P2-3	0.2M Sodium iodide	-	20% w/v PEG 3350
P2-4	0.2M Potassium thiocyanate	-	20% w/v PEG 3350
P2-5	0.2M Sodium nitrate	-	20% w/v PEG 3350
P2-6	0.2M Sodium formate	-	20% w/v PEG 3350
P2-7	0.2M Sodium acetate	-	20% w/v PEG 3350
P2-8	0.2M Sodium sulphate	-	20% w/v PEG 3350
P2-9	0.2M Potassium/sodium tartrate	-	20% w/v PEG 3350
P2-10	0.2M Sodium/potassium phosphate	-	20% w/v PEG 3350
P2-11	0.2M Sodium citrate	-	20% w/v PEG 3350
P2-12	0.2M Sodium malonate	-	20% w/v PEG 3350
P2-13	0.2M Sodium fluoride	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-14	0.2M Sodium bromide	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-15	0.2M Sodium iodide	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-16	0.2M Potassium thiocyanate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-17	0.2M Sodium nitrate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-18	0.2M Sodium formate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-19	0.2M Sodium acetate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-20	0.2M Sodium sulphate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-21	0.2M Potassium/sodium tartrate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-22	0.2M Sodium/potassium phosphate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-23	0.2M Sodium citrate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-24	0.2M Sodium malonate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
P2-25	0.2M Sodium fluoride	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-26	0.2M Sodium bromide	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-27	0.2M Sodium iodide	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-28	0.2M Potassium thiocyanate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-29	0.2M Sodium nitrate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-30	0.2M Sodium formate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-31	0.2M Sodium acetate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-32	0.2M Sodium sulphate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-33	0.2M Potassium/sodium tartarte	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-34	0.2M Sodium/potassium phosphate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-35	0.2M Sodium citrate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-36	0.2M Sodium malonate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
P2-37	0.2M Sodium fluoride	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-38	0.2M Sodium bromide	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-39	0.2M Sodium iodide	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-40	0.2M Potassium thiocyanate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-41	0.2M Sodium nitrate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-42	0.2M Sodium formate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-43	0.2M Sodium acetate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-44	0.2M Sodium sulphate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-45	0.2M Potassium/sodium tartrate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-46	0.2M Sodium/potassium phosphate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-47	0.2M Sodium citrate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
P2-48	0.2M Sodium malonate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350

AI a**FIA – ProtParam results**

Number of amino acids: 299 Molecular weight: 32369.5 Theoretical pI: 5.22

Amino acid composition:

Ala (A)	35	11.7%
Arg (R)	19	6.4%
Asn (N)	9	3.0%
Asp (D)	13	4.3%
Cys (C)	3	1.0%
Gln (Q)	5	1.7%
Glu (E)	22	7.4%
Gly (G)	27	9.0%
His (H)	7	2.3%
Ile (I)	17	5.7%
Leu (L)	21	7.0%
Lys (K)	6	2.0%
Met (M)	6	2.0%
Phe (F)	11	3.7%
Pro (P)	20	6.7%
Ser (S)	17	5.7%
Thr (T)	22	7.4%
Trp (W)	3	1.0%
Tyr (Y)	13	4.3%
Val (V)	23	7.7%

Total number of negatively charged residues (Asp + Glu): 35

Total number of positively charged residues (Arg + Lys): 25

Atomic composition:

Carbon	C	1444
Hydrogen	H	2239
Nitrogen	N	393
Oxygen	O	436
Sulfur	S	9

Formula: C₁₄₄₄H₂₂₃₉N₃₉₃O₄₃₆S₉ Total number of atoms: 4521

Extinction coefficients: Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm.

Ext. coefficient 35995

Abs 0.1% (=1 g/l) 1.112, assuming ALL Cys residues appear as half Cysteine

Ext. coefficient 35870

Abs 0.1% (=1 g/l) 1.108, assuming NO Cys residues appear as half Cysteine

Estimated half-life: The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 35.76

This classifies the protein as stable.

Aliphatic index: 83.58

Grand average of hydropathicity (GRAVY): -0.096

AII b**Sso6206 – ProtParam results**

Number of amino acids: 95	Molecular weight: 10483.1	Theoretical pI: 4.49
Amino acid composition:		
Ala (A) 3	3.2%	
Arg (R) 6	6.3%	
Asn (N) 3	3.2%	
Asp (D) 10	10.5%	
Cys (C) 1	1.1%	
Gln (Q) 0	0.0%	
Glu (E) 9	9.5%	
Gly (G) 8	8.4%	
His (H) 1	1.1%	
Ile (I) 14	14.7%	
Leu (L) 9	9.5%	
Lys (K) 5	5.3%	
Met (M) 5	5.3%	
Phe (F) 1	1.1%	
Pro (P) 1	1.1%	
Ser (S) 6	6.3%	
Thr (T) 4	4.2%	
Trp (W) 0	0.0%	
Tyr (Y) 0	0.0%	
Val (V) 9	9.5%	

Total number of negatively charged residues (Asp + Glu): 19

Total number of positively charged residues (Arg + Lys): 11

Atomic composition:

Carbon C	453
Hydrogen H	769
Nitrogen N	123
Oxygen O	147
Sulfur S	6

Formula: C₄₅₃H₇₆₉N₁₂₃O₁₄₇S₆ Total number of atoms: 1498

Extinction coefficients: This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm.

Ext. coefficient 0

Abs 0.1% (=1 g/l) 0.000, assuming ALL Cys residues appear as half cystines

Ext. coefficient 0

Abs 0.1% (=1 g/l) 0.000, assuming NO Cys residues appear as half cystines

Estimated half-life: The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 30.20

This classifies the protein as stable.

Aliphatic index: 125.05

Grand average of hydropathicity (GRAVY): 0.169

AII c**KDPGal aldolase – ProtParam results**

Number of amino acids: 205	Molecular weight: 21390.6	Theoretical pI: 4.99
Amino acid composition:		
Ala (A) 36	17.6%	
Arg (R) 7	3.4%	
Asn (N) 3	1.5%	
Asp (D) 9	4.4%	
Cys (C) 4	2.0%	
Gln (Q) 13	6.3%	
Glu (E) 10	4.9%	
Gly (G) 19	9.3%	
His (H) 2	1.0%	
Ile (I) 15	7.3%	
Leu (L) 16	7.8%	
Lys (K) 7	3.4%	
Met (M) 3	1.5%	
Phe (F) 6	2.9%	
Pro (P) 12	5.9%	
Ser (S) 8	3.9%	
Thr (T) 10	4.9%	
Trp (W) 3	1.5%	
Tyr (Y) 5	2.4%	
Val (V) 17	8.3%	

Total number of negatively charged residues (Asp + Glu): 19

Total number of positively charged residues (Arg + Lys): 14

Atomic composition:

Carbon C	959
Hydrogen H	1522
Nitrogen N	256
Oxygen O	283
Sulfur S	7

Formula: C₉₅₉H₁₅₂₂N₂₅₆O₂₈₃S₇ Total number of atoms: 3027

Extinction coefficients: Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm.

Ext. coefficient 24200

Abs 0.1% (=1 g/l) 1.131, assuming ALL Cys residues appear as half cystines

Ext. coefficient 23950

Abs 0.1% (=1 g/l) 1.120, assuming NO Cys residues appear as half cystines

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 25.49

This classifies the protein as stable.

Aliphatic index: 100.59

Grand average of hydropathicity (GRAVY): 0.292