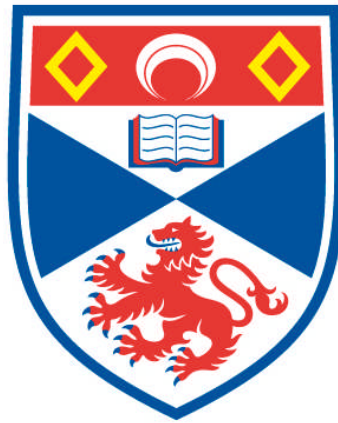# IDENTIFICATION AND ENZYME STUDIES OF RARE AMINO ACID BIOSYNTHESIS FROM STREPTOMYCES CATTLEYA

K. K. Jason Chan

A Thesis Submitted for the Degree of PhD
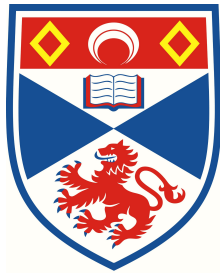## at the
University of St Andrews

2013

UNIVERSITY OF ST ANDREWS

IDENTIFICATION AND ENZYME STUDIES OF RARE AMINO ACID BIOSYNTHESIS FROM *STREPTOMYCES CATTLEYA*

K. K. JASON CHAN

A thesis presented to the University of St Andrews
for the degree of Doctor of Philosophy

JUNE 2013

Declarations

1. Candidate's declarations:

I, Kwan Kit Jason Chan, hereby certify that this thesis, which is approximately 40,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October 2008 and as a candidate for the degree of Doctor of Philosophy in October 2009; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2013.

Date : 25 November 2013   Signature of candidate :
                 (K. K. J. Chan) _____

2. Supervisor's declaration:

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

Date : 25 November 2013   Signature of supervisor :
                 (Prof. D. O'Hagan) _____

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below. The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis: Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date : 25 November 2013   Signature of candidate :
                 (K. K. J. Chan) _____
                 Signature of supervisor :
                 (Prof. D. O'Hagan) _____

BEATISSIMÆ VIRGINI MARIÆ

STELLÆ MARIS

MATRI MISERICORDIÆ

PROPTER AUXILIA STUDIIS MEIS ET

QUIA AUDIVIT DEPRECATIONES MEAS

ATQUE QUIA ME CONSOLATA EST

IN DESOLATIONIBUS

IN GRATI ANIMI TESTIMONIUM

HANC THESEM

EGO HUMILITER EI DEDICO

Acknowledgement

I would like to thank my supervisor, Professor David O'Hagan for accepting me as a research student in his group at the University of St Andrews. I am truly grateful for his guidance and patience. Despite my shortcomings, he always encouraged me to work towards my goals and not to give up easily. I thank him for the invaluable formation in these years.

I am fortunate to have been taught and to have worked alongside other talented scientists at the University. Coming from a background of Organic Chemistry, I am particularly indebted to those who introduced me to Molecular Biology and Genetics. I thank Drs Jesko Koehnke, Andrew Bent and Judith Reeks from the lab of Prof. James Naismith for teaching me the skills in Protein Crystallography. I thank Drs Chunhua Zhao, Margit Winkler, Mayca Onega, Hai Deng and Jason Schmidberger, biologists from the DOH group who taught me many skills in biological experiments. I also thank all members of the DOH group.

The research was made possible by the work of dedicated staff in the Department. I thank Mrs Caroline Horsburgh for mass spectrometry analyses, Drs Catherine Botting and Sally Shirran for protein mass spectrometry and LCT analyses, Dr Tomas Lebl and Mrs Melanja Smith for running the NMR facility, Prof. Alexandra Slawin for X-ray crystal structure determination and Ian Armitt for media preparation. I also thank Dr Colin McRoberts at Agri-Food and Biosciences Institute, Queen's University of Belfast for GC-MS analyses. I gratefully acknowledge an EPSRC studentship.

During my time in St Andrews, I have been blessed by the company of many friends from such varied backgrounds. I treasure their friendships and I thank them from my heart: Nawaf Al-Maharik, Nouchali Bandaranayaka, Davide Bello, Stefano Bresciani, Poh Wai Chai, Rodrigo Cormanich, Liren Dai, Christina Dreher, Alastair Durie, Tomoya Fujiwara, Deng Hai, Izzaty Hassan, Xiang-Guo Li, Xiaowei Lin, Zheng Liu, Long Ma, Małgorzata Marańska, Krzysztof Marański, Gavin Milne, Pitak Nasomjai, Mayca Onega, Carmela Privitera, Nikos Raheem, Shyam Reyal, Jason Schmidberger, Maciek Skibiński, Daniel Smith, Stephen Thompson, Yi Wang, Yuan Wang, Margit Winkler, Mengyang Xuan, Yangping Yu, Qingzhi Zhang, Chunhua Zhao and Linna Zhou.

My loving parents, brother, aunt Miranda and my grandmother have supported and encouraged me throughout my studies in the UK for over a decade. I thank them most specially.

My friends at Rosary Church in Hong Kong have continually assured me of their prayers. I am most grateful to them. Lastly, may my thanks be unto God, who watched over me in my studies and gave me his graces and blessings. Glory be to the Father, and to the Son and to the Holy Spirit: As it was in the beginning, is now, and ever shall be, world without end. Amen.

Jason Chan

Abstract

This thesis is focused on the biosynthesis of three toxins: fluoroacetate, 4-fluoro-L-theronine and β-ethynyl-L-serine which are biosynthesised by the soil bacteria *Streptomyces cattleya*. The two fluorinated metabolites originate from a common biosynthetic pathway and the thesis describes studies carried out on an aldose-ketose isomerase enzyme of the pathway. The biosynthetic origin of β-ethynyl-L-serine is not known. A total synthesis of this acetylenic amino acid is described along with the development of a new analytical method for identifying the metabolite and for future isotope-labelling based biosynthetic studies.

Chapter 1 presents the background of this research. It is focused on the biosynthesis of fluoroacetate and 4-fluoro-L-threonine by *S. cattleya* and it also introduces alkyne-containing natural products and their biosynthesis.

Chapter 2 describes the work carried out on crystallisation of the aldose-ketose isomerase of the fluorometabolite pathway in *S. cattleya*. Crystals of the isomerase were obtained and they were diffracted by X-ray, however a structure could not be solved.

Chapter 3 contains a site-directed mutagenesis studies of the isomerase from *S. cattleya*.

Chapter 4 describes an enantioselective total synthesis of β-ethynyl-L-serine. A robust analytical technique based on derivatisation using 'Click' chemistry and LC-MS was developed for the detection of this amino acid directly from the fermentation broth.

Chapter 5 details the experimental procedures for compounds synthesised in this thesis and the biological procedures for gene cloning and protein purification.

List of Abbreviations

5'-FDA	5'-deoxy-5'-fluoroadenosine
5'-FDI	5'-deoxy-5'-fluoroinosine
5-FDRP	5-deoxy-5-fluoro-α-D-ribose 1-phosphate
5-FDRPi	5-deoxy-5-fluoro-α-D-ribose 1-phosphate isomerase
5-FDRulP	5-deoxy-5-fluoro-D-ribulose 1-phosphate
aa	amino acids
Ac	acetyl
ACP	acyl carrier protein
Ad	adenine
ADP	adenosine 5'-diphosphate
AMP	adenosine 5'-monophosphate
Ar	any aromatic group
ATP	adenosine 5'-triphosphate
Aux	chiral auxiliary
BLAST	basic local alignment search tool
BME	β-mercaptoethanol
Bn	benzyl
Boc	*tert*-butoxycarbonyl
bp	base pairs
BPB	(S)-2-[N-(N-benzylprolyl) amino]benzophenone
Bu	butyl
c	concentration
ca.	*circa*, approximately
calc.	calculated
cat.	catalytic
cf.	*confer*, refer to
CFE	cell free extract
CIMS	chemical ionisation mass spectrometry
CoA	coenzyme A
conc.	concentrated
COSY	^{1}H-^{1}H correlation spectroscopy
CTP	cytidine 5'-triphosphate
Cy	cyclohexyl
Da	Dalton
DCM	dichloromethane
DEPT	distortionless enhancement by polarisation transfer
DHAP	dihydroxyacetone phosphate
DIBAL	diisobutylaluminium hydride
DMAP	4-dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	dimethylsulfoxide
DMSO-d_6	hexadeuteriodimethylsulfoxide

DNA	2'-deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eg.	*exempli gratia*, for example
EIMS	electron impact ionisation mass spectrometry
Enz	enzyme
eq.	equivalent
ES	electrospray
$E^{\ominus}$	standard electrode potential
ESI	electrospray ionisation
Et	ethyl
et al.	*et alii*, and others
etc.	*et cetera*, and so forth
FAS	fatty acid synthase
Fig	Figure
Fmoc	fluorenylmethyloxycarbonyl
g	Earth's gravitational acceleration
G3P	D-glyceraldehyde 3-phosphate
GC	gas chromatography
GTP	guanosine 5'-triphosphate
h	hour(s)
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	histidine
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	^{1}H-^{13}C heteronuclear single quantum coherence spectroscopy
i-	iso-
I	magnetic spin number
Im	imidazole
incl.	including
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infra-red
J	coupling constant
k	kilo, 10^{3}
kb	kilobase-pairs
KIE	kinetic isotope effect
K_{M}	Michaelis-Menten constant
LC-MS	liquid chromatography- mass spectrometry
LD$_{50}$	50 % lethal dose

m	mass
M	mol dm^{-3}
M	molecular ion
m/z	mass divided by charge
Mb	mega-base pairs
MBP	maltose binding protein
MCS	multiple cloning site
Me	methyl
min	minutes or minus
m.p.	melting point
MS	mass spectrometry
Ms	mesyl, methanesulfonyl
MS-MS	tandem mass spectrometry
MSTFA	*N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide
MTRP	5-methylthio-α-D-ribose 1-phosphate
MTRPi	5-methylthio- α-D- ribose 1-phosphate isomerase
MW	molecular weight
MWCO	molecular weight cut-off
n-	normal-
Nº	numbers
NAD⁺	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
NADP⁺	nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
NMR	nuclear magnetic resonance spectroscopy
NRPS	non-ribosomal peptide synthase
NTA	nitriloacetic acid
Nu	nucleophile
o-	ortho-
OAc	acetate
OD	optical density
ORF	open reading frame
ori	origin of replication
p-	para-
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer-saline
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol

PEGMME	polyethylene glycol monomethyl ether
Ph	phenyl
Pi	inorganic phosphate
PPi	inorganic pyrophosphate
PKS	polyketide synthase
PLP	pyridoxial 5'-phosphate
PNP	purine nucleoside phosphorylase
ppm	parts per million
Pr	propyl
pyr	pyridine
quant.	quantitative
r.t.	room temperature (20 °C)
rac-	racemic
R_f	retention factor
RP	reverse phase
rpm	revolutions per minute
SAM	*S*-adenosyl L-methionine
SDS	sodium dodecylsulfate
SEC	size exclusion chromatography
SM	starting material
S$_N$2	bimolecular nucleophilic substitution
t-, *tert*-	tertiary-
TBAF	tetra-*n*-butylammonium fluoride
TBDPS	*tert*-butyldiphenylsilyl
TBS, TBDMS	*tert*-butyldimethylsilyl
TBTA	tris-(benzyltriazolylmethyl) amine
TEV	tobacco etch virus
Tf	trifluoromethanesulfonyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TMS	trimethylsilyl
TOF	time of flight
Tr	trityl, triphenylmethyl
Tris	tris(hydroxymethyl)amino-methane
Ts	tosyl, *p*-toluenesulfonyl
UV	ultraviolet
v	volume
X	any halogen
X	times concentrated (for stock reagents)
δ	chemical shift

In addition to this list, standard one-letter and three-letter codes for common amino acids are also used throughout this thesis.

Table of Contents

Chapter 1

Introduction

Chapter 2

Crystallisation of the 5-FDRPi isomerase

Chapter 3

Studies explaining the isomerase mechanism

Chapter 4

The total synthesis of β-ethynyl-L-serine and its *in vivo* detection by the click reaction

Chapter 5

Experimental section

Appendix

1. Introduction

1.1 Cell metabolism

Metabolism describes the chemical reactions that take place in a living organism, and may be subdivided into primary and secondary metabolism.

Primary metabolism is the ordinary metabolic pathways which are essential to the survival and function of a living organism. These pathways are involved in the central roles of the cell for generation of energy, cellular reproduction, growth and development, the breakdown of waste products and synthesis of the cellular components, such as nucleic acids, amino acids, peptides, saccharides and lipids. The functional molecules generated from primary metabolism are known as *primary metabolites*; many biosynthetic paths of primary metabolites are universally found in lower and higher organisms.

Secondary metabolism, on the other hand, refers to the metabolic pathways that serve an alternative purpose other than the basic functions of the cell. Some bacteria produce many secondary metabolites which act as signalling molecules, antibiotics, toxins, pigments, *etc.* and they play important roles in their resource management and interspecies defences. The biosynthetic pathways of the secondary metabolites are highly differentiated among each species of bacteria and therefore give rise to a huge range of natural products with a diversity of structure and complexity. The major classes of secondary metabolites are the polyketides, terpenoids, non-ribosomal peptides, alkaloids and glycosides.

Despite the chemical complexity of secondary metabolites, they are all assembled from simple building blocks that are found in the primary metabolic pathways.

Acetate, shikimate and the 20 amino acids are the basic building blocks from which most secondary metabolites are assembled (Scheme 1.1).

CO_2 + H_2O

photosynthesis

Calvin cycle

carbon fixation

Sugars
Starch
Cellulose

Nucleic acids
Glycosides
Polysaccharides

HO — O — OH
HO — OH
OH

glucose

pentose phosphate pathway

glycolysis

^-OOC — OPO_3^{2-}

phosphoenolpyruvate

OH
OHC — OPO_3^{2-}
OH

erythrose 4-phosphate

COO^-
HO — OH
OH

shikimate

Aromatic compounds

Phenolics

^-OOC — O

pyruvate

non-mevalonate pathway

Aromatic amino acids

Isoprenoids

Alipthatic amino acids

Proteins
Alkaloids
Peptides

CoAS — O

Krebs cycle

acetyl CoA

OH
^-OOC — OH

mevalonate

Isoprenoids

Fatty acids
Polyketides

Scheme 1.1. Simplified scheme showing the relationships between major primary and secondary metabolites. (Scheme adapted from *Biochemistry*, David E. Metzler, Academic Press 1977)

1.2 Soil bacteria are a rich source of natural products

It has been estimated that there are 40 million bacterial cells of over 10, 000 different species in every gram of soil.[1] Each of these bacteria strives to out-compete each other for the limited nutrients available. This fierce natural selection process fuels the development of their secondary metabolism, making the soil bacteria by-far the best known microbial producers of antibiotics and other bioactive natural products; with over half of the microbial antibiotics discovered originating from Actinomycetes.[2]

Of all the soil-dwelling bacteria, the most notable genus is the Streptomycetes, the largest antibiotic-producing genus known.[3] *Streptomyces* bacteria are ubiquitous in nature; they are spore-forming with very limited mobility; after germination from a spore, they grow into mycelium fibres and finally differentiate into coiled threads of spore particles for dispersion (Fig. 1.1). Due to their ecology and physiology, the Streptomycetes protect their resources and colonise the soil by excretion of anti-microbial metabolites to eliminate competing microorganisms in their vicinity.[4]

Many antibiotics currently in clinical use were discovered from *Streptomyces* bacteria, some examples are given below in Fig 1.2. These antibiotics have contributed significantly towards the management of bacterial infections in modern medicine and have improved survival rates and helped save many lives.
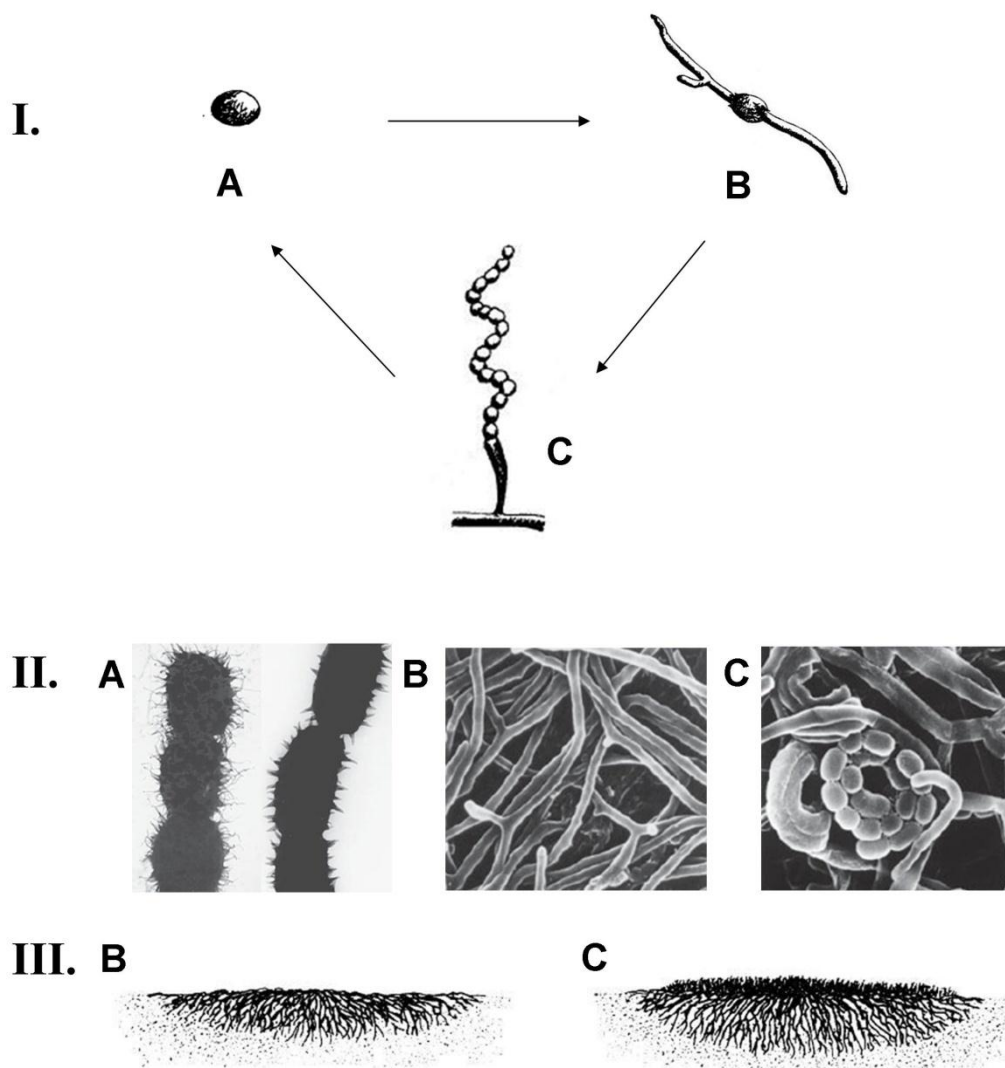
Fig. 1.1. *Streptomyces* at various stages of its life cycle: **A.** Spore **B.** Vegetative hyphae develop from spores germination **C.** Aerial hyphae with new spores formed as coiled threads at the ends.
I. Life cycle illustration in drawing. **II.** Scanning electron microscope images **III.** View of the colony. (Images adapted from *Practical Streptomyces Genetics*[4])

Oxytetracycline
(*Streptomyces rimosus*)

Chloramphenicol
(*Streptomyces venezuelae*)

Kanamycin
(*Streptomyces kanamyceticus*)
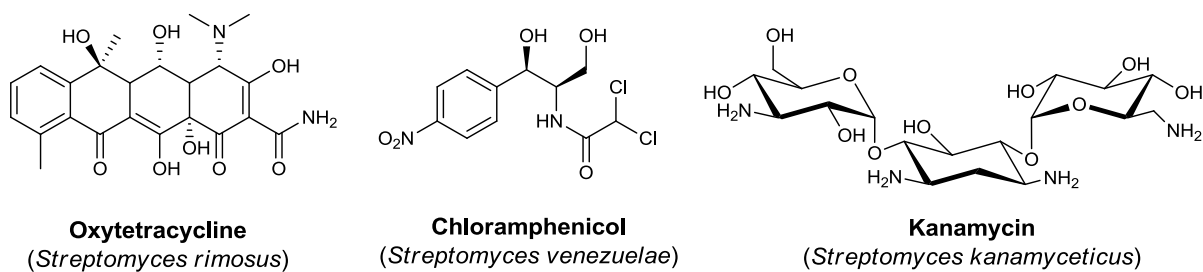
Fig. 1.2. Examples of important antibiotics produced by *Streptomyces* species.

1.2.1 *Streptomyces cattleya*

In 1976, a new β-lactam antibiotic thienamycin **1** was discovered by Kahan *et al.* during the screening of various soil bacteria cultures.[5] This thienamycin producing strain was then assigned the name *Streptomyces cattleya* (Fig. 1.3).
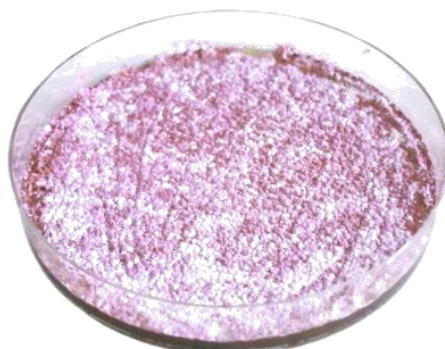
Fig. 1.3. *S. cattleya* maintained on a petri dish (30 days) in the laboratory environment. It produces characteristic purple-coloured spores.

A group at Merck had focused their research efforts on the optimisation of thienamycin production. This led to a chance discovery reported in 1986, of two fluorinated metabolites from the broth of a fermentation of *S. cattleya*.[6] These metabolites were shown to be fluoroacetate **2** and 4-fluoro-L-threonine **3**.
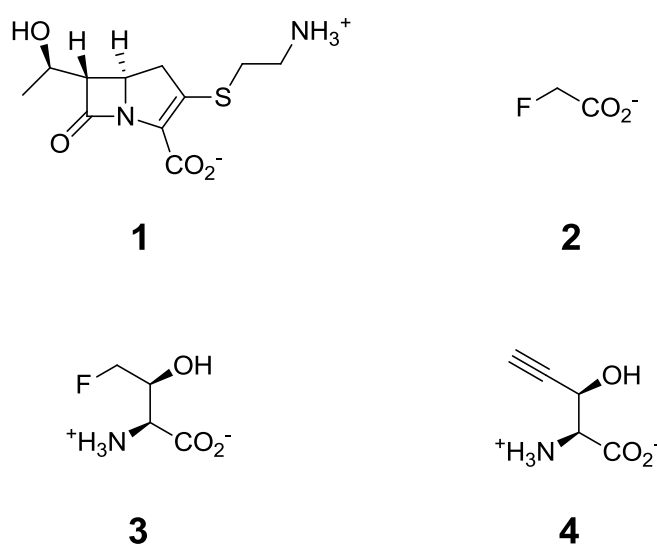
Fig. 1.4. Natural products of *Streptomyces cattleya*.

It was found that the soy bean flour from that experiment had an unusually high concentration of fluoride salts and it was shown that when this microorganism was grown in a media supplemented with fluoride, it produced these rare fluorometabolites. Fluoroacetate **2** is known to be a very toxic substance[7, 8] and 4-fluoro-L-threonine **3** was shown to be a broad spectrum antibiotic, although the latter is not a useful candidate as an anti-microbial agent due to its high toxicity.[6] However these metabolites have attracted interest as fluorinated natural products are extremely rare in biology.

The Merck group isolated another rare amino acid, β-ethynyl-L-serine **4** in the cultures of *S. cattleya*.[9] This acetylenic amino acid had previously been identified from the fungus *Sclerotium rolfsii*[10] and was already known to be a potent toxin.

The complete genome sequence of *S. cattleya* was published in 2011 by two groups independently.[11, 12] Its genome was found to consist of a linear chromosome of 6.28 Mb and a linear mega-plasmid of 1.81 Mb. The published genome has become a new and powerful tool for researchers to gain an insight into the complex secondary metabolism of this bacterium.
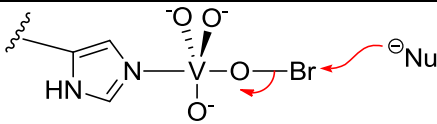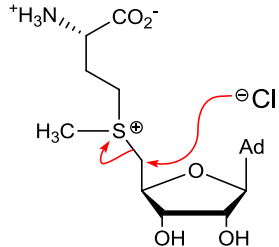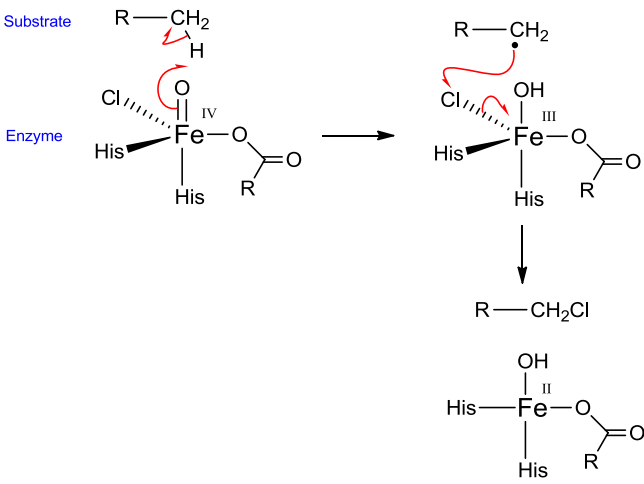
1.3 Halogenated natural products

The halogens are often utilised by living organisms and are incorporated into a myriad of natural products; no less than 4000 of natural organohalogens have been described so far.[13] The halogen is often crucial for their bioactivities.[14,15]

There are three general ways in which carbon-halogen bonds are generated biochemically: either *via* a halide anion (X^-) acting as a nucleophile, a cationic halonium ion species (X^+) acting as an electrophile or through a radical $(X^{\cdot})$ process.[16] Examples of each category of reaction are shown in Table 1.2.

Electrophilic halogenation involves oxidation of the halide by hydrogen peroxide. This type of reaction is well known in marine plants and microorganisms, as many haloperoxidases have been characterised.[17, 18] The active halogen is a hypohalite species stabilised by coordination to a transition metal centre (heme-Fe dependent haloperoxidases,[17] vanadium-dependent haloperoxidases[18, 19]) or an aromatic ring (flavin-dependent haloperoxidases[20]). Nucleophilic halogenation uses the halide ion directly as a nucleophile to attack an electrophilic carbon of the substrate.[21, 22] Enzymatic radical halogenation processes are known to modify inert sp^3 carbons, such as a methyl group. Several $Fe(II)$, α-ketoglutarate and O_2 dependent chlorinases have been studied.[23-25] They were observed to generate a Cl-$Fe(IV)$-oxo species that is capable of abstracting a hydrogen atom to generate a substrate radical and an intermediate Cl-$Fe(III)$-OH species, followed by C-Cl bond formation and reduction to $Fe(II)$.

8

Table 1.2. Examples of enzymatic halogenations

Mode of halogenation	Example	Active halogen species
Electrophilic (X⁺)	**Vanadium-dependent bromoperoxidases**[19] (marine seaweeds)	vanadium-bound hypobromite
Nucleophilic (X⁻)	**SalL, chlorinase**[21] (Salinosporamide A biosynthesis, *Salinispora tropica*)	free chloride ion
Radical (X˙)	**BarB1, BarB2 Iron-dependent halogenases**[26] (Barbamide biosynthesis, *Lyngbya majuscula*)	Cl-Fe(IV)-oxo species

The chemistry for chlorination, bromination and iodination could all be achieved by nucleophilic reaction with the halide ion, electrophilic reaction with the halonium ion or generation of a halogen radical species in biological systems.

Biological fluorination, on the other hand, is only possible *via* a nucleophilic fluoride ion. In order to generate the fluorine radical or a cationic fluorine species, F(-1) must be oxidised; however the oxidation potential of F(-1) required to generate F(0) is very high ($E^{\ominus}_{ox}$ = -2.87 V) and unattainable by coupling to either the reduction of H_2O_2 ($E^{\ominus}$ = +1.77 V), the oxidant of haloperoxidase reactions or any other known reductions in the cell. As a result, both radical and cationic fluorine cannot be generated. Even the hypofluorite ion OF^- has F in a -1 oxidation state, F being more electronegative than O. Fluoride ion (F^-) is a very poor nucleophile in aqueous systems for two reasons, firstly the anion is heavily hydrated due to its high charge density, and the resulting hydration shell reduces the nucleophilicity of F^-. Secondly, the fluorine lone pair orbitals are contracted and good orbital interactions are difficult to establish (Table 1.3).

Table 1.3. Selected thermodynamic values for halogens.[27]

Halogen	Redox potential 0→-1 $\frac{1}{2}X_2 + e^- \rightleftharpoons X^-$ $E^{\ominus}/V$	Redox potential +1→0 $HOX + H^+ + e^- \rightleftharpoons \frac{1}{2}X_2 + H_2O$ $E^{\ominus}/V$	Hydration enthalpy X^- (g) $\rightleftharpoons$ X^- (aq) $\Delta H^{\ominus}_{hydr}/kJmol^{-1}$
F	+2.87	*not applicable*	506
Cl	+1.36	+1.64	364
Br	+1.07	+1.59	335
I	+0.54	+1.45	293

Chlorine is the most common halogen found in natural products, followed by bromine. Fluorinated and iodinated natural products are comparatively rare, although over 100 iodinated products are known.[13]

The Oceans are where the majority of bio-available halogens are to be found, where chlorine (1.94 % by mass) is significantly more abundant than bromine (67.3 ppm), fluorine (1.3 ppm) and iodine (0.06 ppm).[27] This order of halide abundances in the Oceans corresponds to the order of the numbers of discovered halogenated natural products, however the number of fluorinated natural products is fewer than excepted, and this is most likely due to the difficulties in biological fluorination chemistry described above.

Although the levels of fluoride are low in sea water, fluorine is still the most abundant of the halogens on the Earth's crust (0.058 % by mass[27]). However, most of the Earth's fluorine occur as highly insoluble metal fluoride minerals, such as fluorite (CaF_2) and cryolite (Na_3AlF_6). In this insoluble form, fluoride ion cannot be easily utilised by living organisms, a significant factor contributing to the under-development of a fluorine biochemistry.

1.3.1 Fluorinated natural products

Five fluorinated natural products have been unambiguously described from plants and bacteria (Table 1.4). Additionally two other fluorinated compounds (fluoroacetone[28, 29] and 5-fluorouracil[30]) were reported but it is not clear they are true natural products.

Table 1.4. The five fluorinated natural products.

Fluorinated Natural Product	Structure	First identified	Producing Organism(s)
Fluoroacetate	$F\diagup CO_2^-$ **2**	1943	*Dichapetalum cymosum*[7, 31] (South African shrub) *Streptomyces cattleya*[9] (soil bacterium)
$(2R, 3R)$- 2-Fluorocitrate	HOOC, COOH structure **5**	-	many plants[32]
ω-Fluorofatty acids	$F\diagdown\diagup_7 = _7 COOH$ **6** eg. ω-fluorooleic acid	1959	*Dichapetalum toxicarium* (West African shrub)[33-35]
Nucleocidin	structure **7**	1957	*Streptomyces calvus*[36-38] (soil bacterium)
4-Fluoro-L-threonine	$F\diagup \diagdown OH$ $^+H_3N \diagdown CO_2^-$ **3**	1986	*Streptomyces cattleya*[9] (soil bacterium)

Fluoroacetate **2** is the most common natural product of this class, first identified in 1943 as the toxic component of a South African plant *Dichapetalum cymosum.*[7, 31] Later, it was found that many other species of poisonous plants across Africa,[39, 40] Australia[41] and Brazil[42] accumulated fluoroacetate in their tissues. Further reports have found varying concentrations of fluoroacetate in samples of tea leaves from different origins[43] and it has been suggested that the ability to biosynthesise fluoroacetate by plants may be quite common.[43, 44] Although the biological conversion from fluoride to fluoroacetate in plants has yet to be studied in detail, its poisoning mechanism has attracted more attention and is better understood.

Once fluoroacetate is absorbed into a living cell, it is processed by the enzymes that utilise acetate. Its conversion to fluoroacetyl coenzyme A (CoA) **8** is readily observed[45, 46] and this derivatisation brings fluoroacetate into cellular metabolic pathways. Fluoroacetyl-CoA enters the Krebs cycle, the energy-generation cycle that oxidises acetate from the break-down of glucose. In the Krebs cycle fluoroacetyl-CoA undergoes a condensation reaction with oxaloacetate **9**, to generate fluorocitrate **5**. This reaction is catalysed by citrate synthase. Only the (2*R*, 3*R*) diastereoisomer of fluorocitrate **5** is produced, which is also the only toxic stereoisomer.[47] The other three possible stereoisomers of this molecule are non-toxic. (2*R*, 3*R*)-2-Fluorocitrate inhibits the next enzyme of the Krebs cycle, aconitase, which interconverts citrate and isocitrate (Scheme 1.2).[48, 49] Inhibition blocks this crucial metabolic pathway, hindering the complete oxidation of glucose and accounting for the very high toxicity of fluoroacetate in mammals and other organisms.
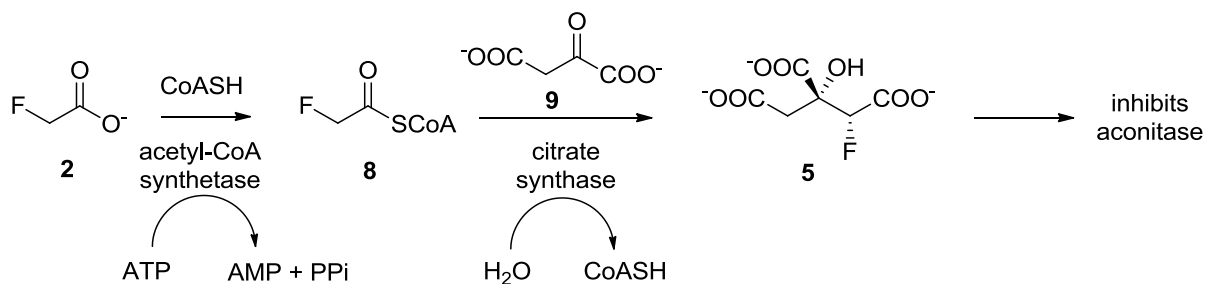
Scheme 1.2. The biosynthesis of 2-fluorocitrate.

Producers of fluoroacetate, such as the poisonous plants and the bacterium *Streptomyces cattleya* show high tolerance to the toxin, most likely due to the presence of a resistance mechanism that hydrolyses fluoroacetyl-CoA. One of the genes clustered with the fluorination enzyme of *S. cattleya*, *flK*, encodes for a fluoroacetyl-CoA thioesterase and the current view is that this gene expresses an enzyme which confers resistance by hydrolysing fluoroacetyl-CoA (Scheme 1.3).[50] It has also been observed that mitochondrial extracts of *D. cymosum* are also able to catalyse the hydrolysis of fluoroacetyl-CoA.[51] (2*R*, 3*R*)-2-Fluorocitrate **5** is not a true secondary metabolite as it is only formed as an unintended product when fluoroacetyl-CoA is substituted for acetyl-CoA in primary metabolic reactions. However it is a metabolite of fluoroacetate when it has been taken up by living cells and is also considered as one of the true organofluorine natural products.
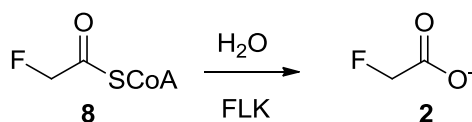
Scheme 1.3. The ester hydrolysis of fluoroacetyl-CoA.

The ω-fluorofatty acids also originate from fluoroacetyl-CoA. These fluorinated fatty acids are found in the seeds of a plant from Sierra Leone, *Dichapetalum toxicarium*.[33-35] The fatty acids are generated when fluoroacetyl-CoA is utilised as the starter unit in

fatty acid biosynthesis (Scheme 1.4), however presumably since fluoromalonyl-CoA is not formed, the fluorine is always located at the terminus of the fatty acid chain.[44] In this case, the fluorinated fatty acids are thought to be produced when fluoroacetyl-CoA is a surrogate substrate instead of acetyl-CoA for enzymes of the primary metabolic pathways.

Scheme 1.4. The biosynthesis of ω-fluorooleic acid in *D. toxicarium*. Other ω-fluorofatty acids are also formed similarly, differing by the number of chain extension cycles and the desaturation step.

From the five recognised organofluorine natural products, three are produced by *Streptomyces* bacteria.

Nucleocidin **7** is a nucleoside sulfonamide antibiotic isolated from *Streptomyces clavus*.[36-38] Shortly after its initial discovery in 1957, re-isolation of this fluorinated compound has proven very difficult. These cultures have been kept by several collections worldwide but despite many attempts by different research groups, there has been no successful reproduction of this metabolite. Therefore *S. cattleya* is the only strain of bacteria that is still able to perform fluorination *in vivo*.

4-Fluoro-L-threonine **3** is co-produced by *Streptomyces cattleya* with fluoroacetate **2**. They share a biosynthesis pathway, branching only at a late stage. The fluorination event and subsequent biosynthesis of **2** and **3** are now quite well understood and will be described in some detail below.

1.3.2 Fluorometabolite pathway of *Streptomyces cattleya*

Of the known species that produce fluorometabolites, *S. cattleya* was clearly the most practical organism to begin a biosynthesis study due to the relatively simple biochemistry of a bacterium compared to plants.

Early isotope labelling studies using deuterium labelled fluoroacetaldehyde identified fluoroacetaldehyde **12** as a common precursor of **2** and **3**.[52] This was followed by ^{19}F NMR experiments of cell free extracts (CFE), which allowed identification of all of the fluorinated intermediates and suggested the six enzyme activities involved in the conversion of fluoride ion to **2** and **3**. The pathway is illustrated in Scheme 1.5.
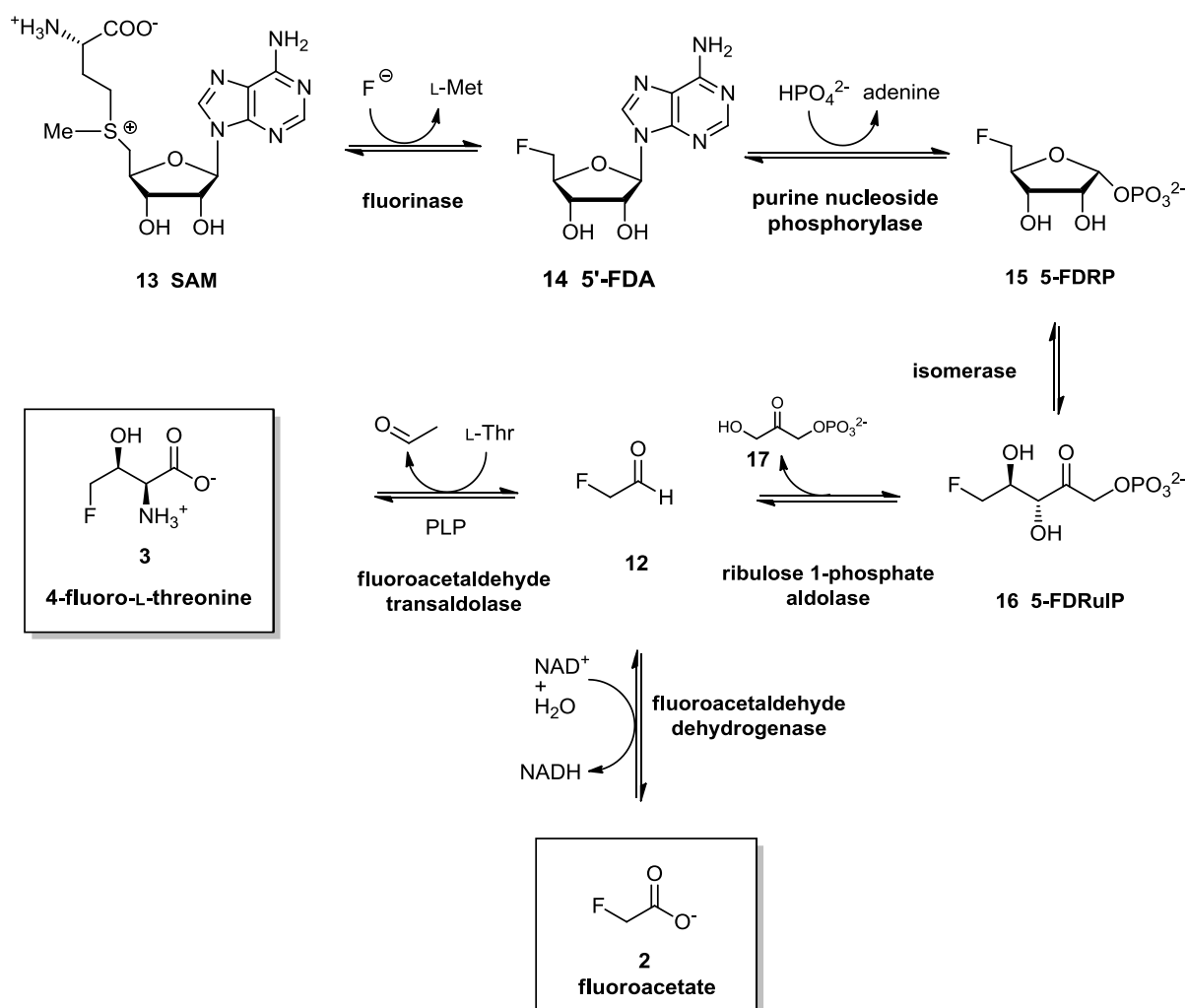
Scheme 1.5. The fluorometabolite pathway of *S. cattleya*.

Highlights of the studies carried out at the Universities of St Andrews and Durham investigating this pathway are described below.

1.3.2.1 Fluoroacetaldehyde as a common precursor

The first understanding of this pathway focused on fluoroacetaldehyde, the immediate precursor to fluoroacetate and 4-fluorothreonine, establishing the metabolic relationship between these two natural products. Isotope experiments concluded that fluoroacetate was not a precursor to 4-fluorothreonine and *vice versa*.[52]

Fluoroacetaldehyde **12** was then confirmed as a common intermediate by the efficient conversion of a synthetic sample of $[2,2\text{-}^2H_2]$fluoroacetaldehyde into deuterated **2** and **3** by resting cell cultures of *S. cattleya*.[52]

The oxidation of **12** into **2** by an NAD^+ dependent dehydrogenase was envisaged and a search was made to isolate such an enzyme from the cell-free extract (CFE). An active enzyme was purified that efficiently carried out the conversion. This enzyme had a native molecular weight of 200 kDa and is ten-fold more specific for fluoroacetaldehyde ($K_M = 80$ μM) than acetaldehyde ($K_M = 810$ μM).[53]
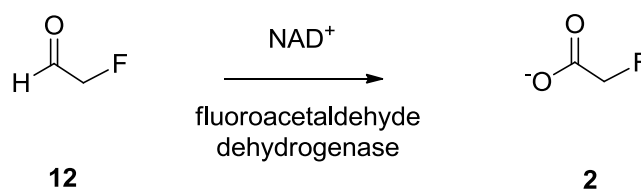
Scheme 1.6. The fluoroacetaldehyde dehydrogenase reaction.

The formation of 4-fluorothreonine **3** from fluoroacetaldehyde **12** was anticipated to be mediated by a pyridoxal 5'-phosphate (PLP) enzyme, similar to known bacterial threonine aldolases that couple glycine with acetaldehyde in the synthesis of the amino acid L-threonine.[54] Based on this, partially purified fractions of the cell-free extract were screened for activity using ^{19}F-NMR with various amino acids (glycine, threonine, alanine, serine, cysteine, aspartate) and fluoroacetaldehyde in the presence of the co-factor PLP. Only the amino acid L-threonine led to the production of 4-fluorothreonine. An active fraction of the CFE was located and purified to give a protein with a denatured mass of 60 kDa and which seemed to be a dimer in its native form.[55] This threonine transaldolase is distinct from other related enzymes in that it does not accept glycine as a substrate.
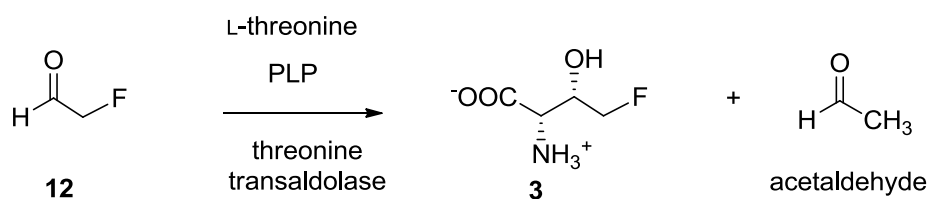
Scheme 1.7. The PLP-dependent threonine transaldolase reaction.

1.3.2.2 Identification of the fluorination enzyme

C-F bond formation is the key step of the biosynthesis of the two organofluorine natural products and thus research efforts were focused on studying how an organofluorine compound was produced from potassium fluoride in the CFE of *S. cattleya*. It was anticipated that, fluorophosphates may be candidates for an 'active fluoride' for the fluorination reaction. Accordingly, various phosphate-containing substrates, such as ATP, GTP and CTP were incubated with KF and the CFE. In this experiment, ATP incubations unexpectedly yielded three organic fluorine signals in the ^{19}F NMR, one of which was assigned as fluoroacetate.[56]

Based on the structure of ATP **18**, a product 5'-deoxy-5'-fluoroadenosine (5'-FDA) **14** was proposed where the triphosphate group was displaced by nucleophilic attack of fluoride. After further interrogation of the CFE, it was found that the addition of L-methionine with ATP increased the production of organic fluorine compounds. This suggested *S*-adenosyl-L-methionine (SAM) **13** as a more immediate substrate for fluorination, as ATP **18** is readily transformed to SAM by L-methionine adenosyltransferase,[57] an enzyme present in the CFE.
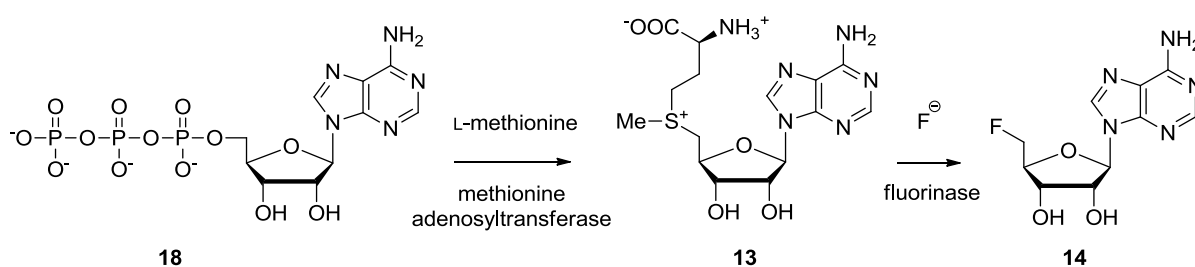
Scheme 1.8. The methionine adenosyltransferase and fluorinase reactions.

Incubation of SAM alone (without ATP) with KF in the CFE of *S. cattleya* allowed the generation of fluoroacetate. This confirmed SAM **13** as the substrate for fluorination, and 5'-FDA **14** was also confirmed as an intermediate on the pathway using a

synthetic reference sample. The reaction that converts SAM to 5'-FDA is the first step of secondary metabolism to fluorometabolites *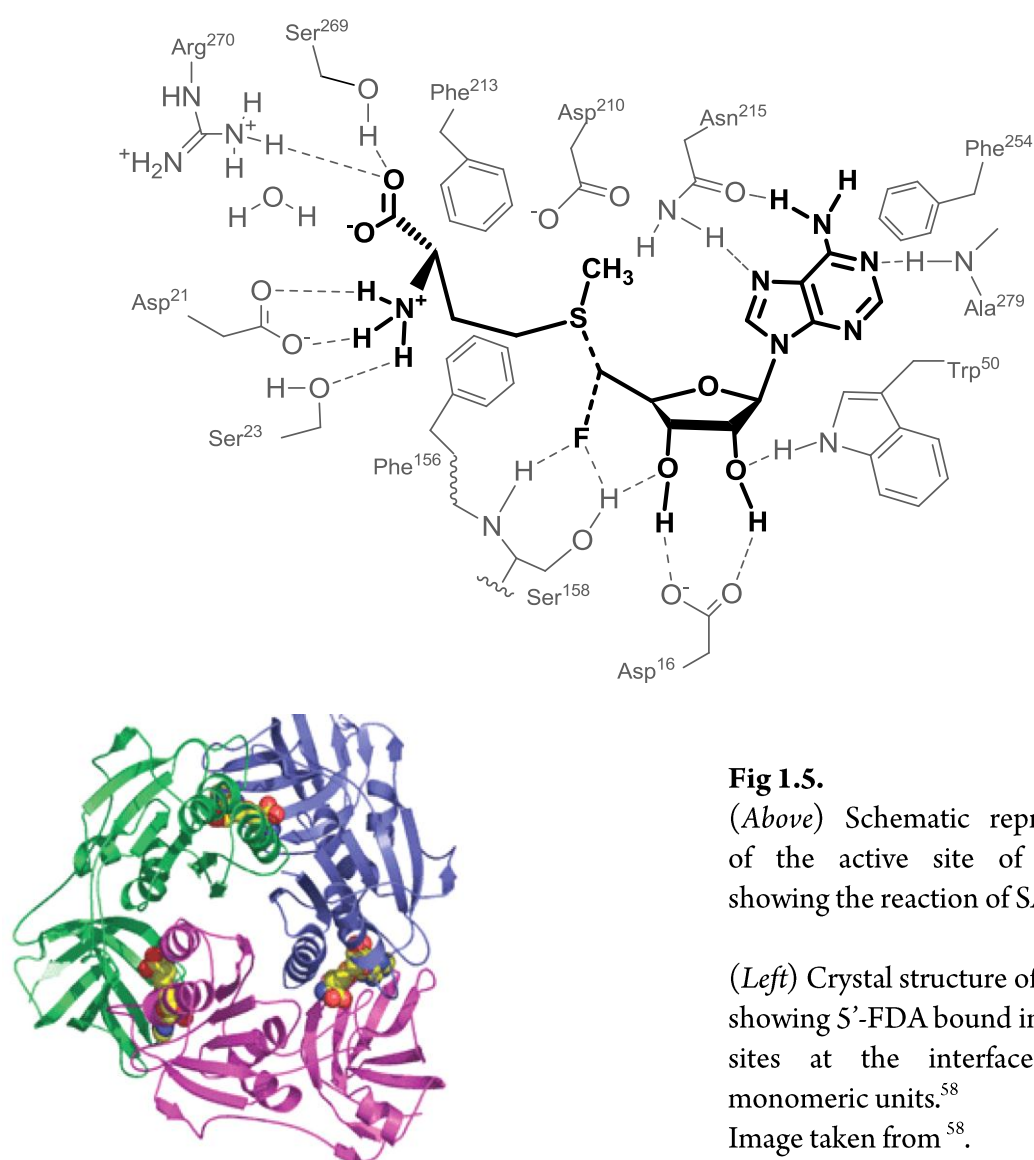*2** and **3**.[56] An enzyme was subsequently purified from the CFE that was capable of this fluorination reaction *in vitro*.[22] The enzyme was given the common name 'fluorinase' (5'-deoxy-5'-fluoroadenosine synthase) and is the first enzyme known to utilise inorganic fluoride to form a C-F bond in nature.

It became an objective at that time to over-express the genes encoding the fluorinase. A partial amino acid sequence was obtained by Edman degradation, from analysis of the wild-type protein, and this was used to design PCR primers for the fluorinase. After successful cloning of the fluorinase gene, it was found to over-express well in *E. coli* using a pET28A(+) plasmid.[58] Using this over-expression system, the fluorinase proved relatively easy to produce in mg quantities with a His_6 affinity tag. Thus it was possible to simplify the multi-step purification procedure to an elution by nickel affinity chromatography. The availability of the fluorinase enzyme assisted in solving the crystal structure of this protein.

The crystal structure revealed the protein to be a hexamer (dimer of trimers), and revealed three active sites, located at trimer interfaces. SAM was actually observed bound at the active site in the crystal structure through specific interactions of amino acid residues with the adenine ring, ribose ring and the L-methionine chain of SAM, indicating a very high specificity for SAM as a substrate. When fluoride was added to crystallisation trials, the enzyme turned over and a structure with the products 5'-FDA **14** and L-methionine bound at the active site was obtained. A comparison of the structures with the substrate and the products was consistent with an S_N2 reaction

course with the C-F bond of 5'-FDA and the C-S bond of SAM being arranged in an approximately planar fashion when the structures were overlapped (Fig. 1.5). [58]

Further studies using SAM, prepared with a stereogenic centre at the 5'-carbon, which was generated by substitution of one of the 5'-methylene hydrogens with deuterium, confirmed the fluorinase reaction goes through one inversion of configuration, consistent with an S_N2 mechanism.[59]

Fig 1.5.
(*Above*) Schematic representation of the active site of fluorinase showing the reaction of SAM and F⁻.

(*Left*) Crystal structure of the trimer showing 5'-FDA bound in the active sites at the interfaces of the monomeric units.[58]
Image taken from [58].

1.3.2.3 Purine nucleoside phosphorylase

With the knowledge of the first step of the fluorometabolite pathway, the fate of the product 5'-FDA **14** was investigated. 5'-FDA is a close analogue of adenosine and enzymes that metabolise adenosine were investigated, including that of a purine nucleoside phosphorylase (PNP).[60] The reaction outcome of 5'-FDA with a PNP enzyme results in the displacement of the purine by phosphate at the anomeric carbon. The validity of this intermediate was demonstrated when 5-deoxy-5-fluoro-α-D-ribose 1-phosphate (5-FDRP) **15** was converted into fluoroacetate after it was added to a CFE of S. *cattleya*.[61, 62]
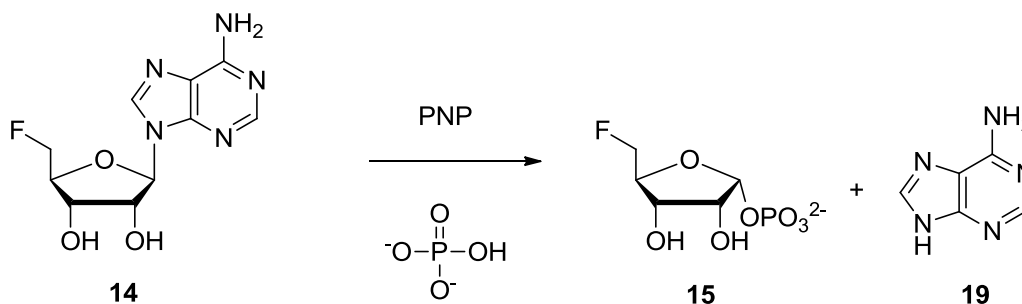
Scheme 1.9. The PNP reaction.

Subsequent studies led to the partial purification of a PNP from the CFE of S. *cattleya*, and showed that it converts 5'-FDA **14** into 5-FDRP **15**.[62] Later genetic studies (described in more detail in 1.3.2.7) would identify a gene (*flB*) encoding a PNP enzyme situated adjacent to the fluorinase gene (*flA*) in a gene cluster dedicated to fluorometabolite biosynthesis.[63]

1.3.2.4 Identification of a ribulose phosphate as intermediate

5-FDRP **15**, the product of the PNP reaction on 5'-FDA **14**, resembled 5-methylthio-α-D-ribose 1-phosphate (MTRP) **20** from the methionine salvage pathway (Fig. 1.6). [64]
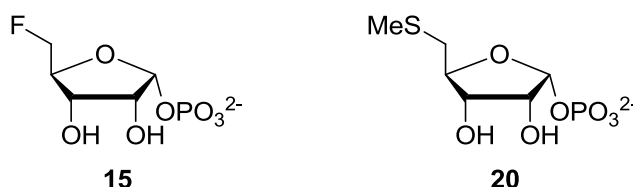
Fig 1.6. Intermediate ribose 1-phosphates from the fluorometabolite pathway and the methionine salvage pathway.

The methionine salvage pathway is a primary metabolic pathway, present in most organisms including bacteria[65, 66] and eukaryotes[67] with the purpose of regenerating L-methionine from waste by-products of SAM. In this way, valuable sulfur-containing resources are conserved within the cell.

In the methionine salvage pathway, **20** is a substrate for an isomerase which opens the hemiacetal group of the ribose ring into an acyclic keto-sugar (Scheme 1.10).[64]
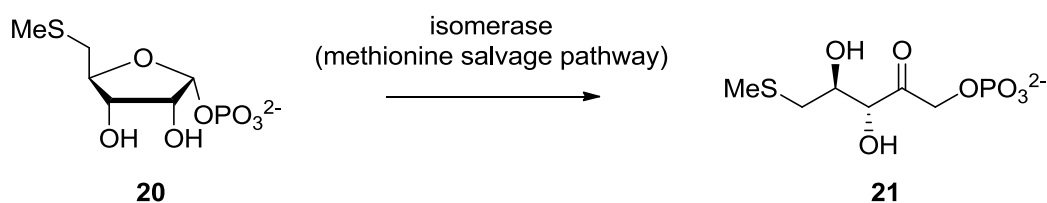
Scheme 1.10. The isomerase reaction from the methionine salvage pathway.

If 5-FDRP **15** was a substrate for this isomerisation reaction, the expected product would be 5-deoxy-5-fluoro-D-ribulose 1-phosphate (5-FDRulP) **16**.

This isomerase activity is an attractive prospect as a component of the fluorometabolite biosynthetic pathway, since such keto-sugars are common products of dihydroxyacetone phosphate (DHAP) dependent aldolases.[68] Such an aldolase could mediate a reverse reaction from the keto-sugar chain to give fluoroacetaldehyde. Using ^{19}F-NMR to monitor formation of new organofluorine species, a new compound was observed when 5-FDRP **15** was incubated with a CFE of *S. cattleya*. This compound was confirmed as 5-FDRulP **16** by GC-MS analysis after derivatisation of the reaction mixture with *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) **22**.[69] The persilylated 5-FDRP **23** had a different retention time to persilylated 5-FDRulP **24**, and their identities were assigned by their different ion fragmentation patterns (Scheme 1.11).
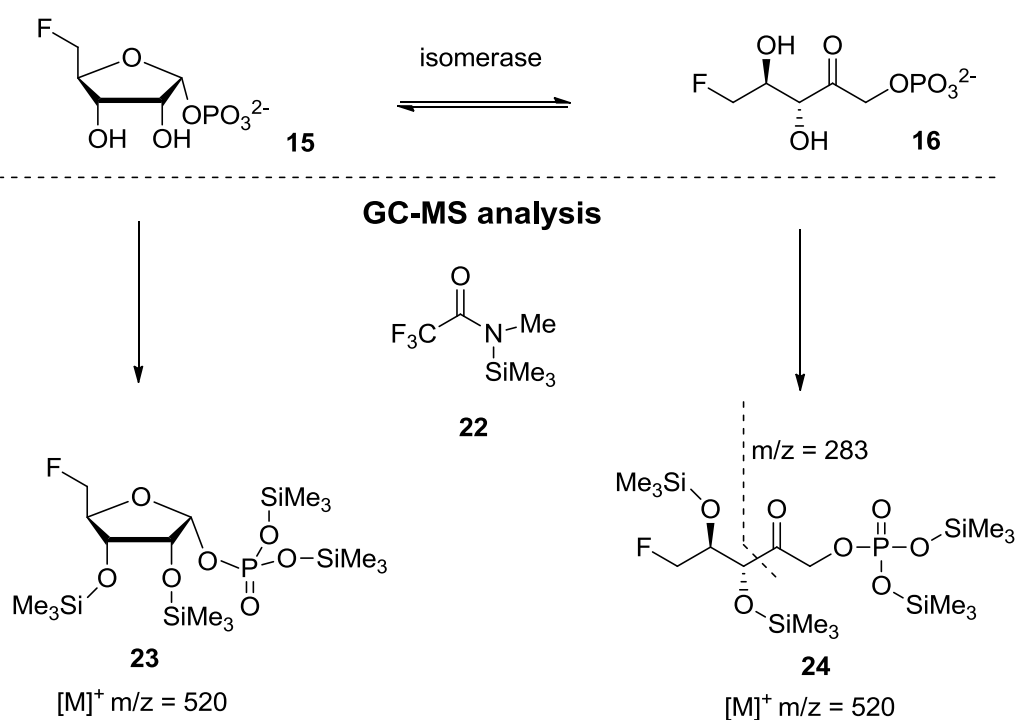
Scheme 1.11. The isomerase reaction from the fluorometabolite pathway and its GC-MS analysis.

An isomerase enzyme was purified from the CFE which had the capability to convert 5-FDRP to 5-FDRulP.[70] A mechanistic and crystallographic study of the isomerase is the focus of the thesis and will be discussed in greater detail in *Chapter 2*.

1.3.2.5 An aldolase completes the pathway

Fluoroacetaldehyde **12** has already been confirmed as an intermediate of the fluorometabolite pathway. A retro-aldol reaction on 5-FDRulP **16** offered a potential route to **12**, and this type of reactions is known to be catalysed by DHAP-dependent aldolases.[68] These aldolases are ubiquitous and catalyse coupling reactions between dihydroxyacetone phosphate **17** and (S)-lactaldehyde **25** or glyceraldehyde 3-phosphate (G3P) **26** to generate pentose sugars. The reaction generates two stereogenic centres. DHAP aldolases are classified into four types according to the stereochemistry of the product[71, 72] (Scheme 1.12). Based on the stereochemistry at C-3 and C-4 of D-ribulose, the aldolase of the fluorometabolite pathway belongs to the fuculose 1-phosphate aldolase class.
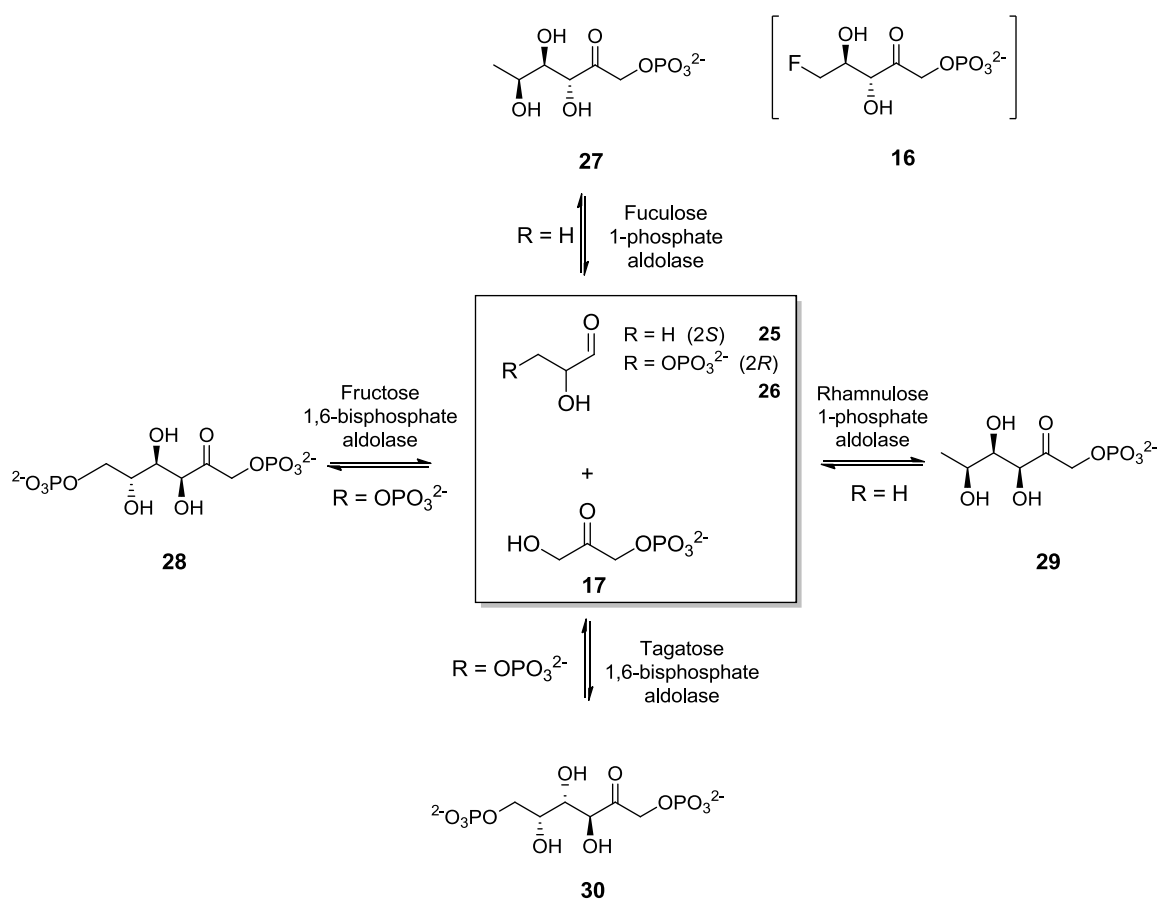
Scheme 1.12. The aldol condensation products from the four classes of aldolases.

There are several aldolases present in *S. cattleya*, and in early work identified two: one is a putative fuculose 1-phosphate aldolase and the other is a fructose 1,6-bisphosphate aldolase,[62] which was shown to produce an incorrect diastereomer, 5-deoxy-5-fluoro-D-xylulose 1-phosphate, in the reverse reaction between DHAP and fluoroacetaldehyde.[62]

Attempts to purify the putative fuculose 1-phosphate aldolase from the CFE of *S. cattleya* by chromatographic methods were not successful.[73] Cloning this gene from the genomic DNA of *S. cattleya* using PCR primers, designed from the known sequence of a related fuculose aldolase from *Streptomyces coelicolor*, were also frustrated.[73]
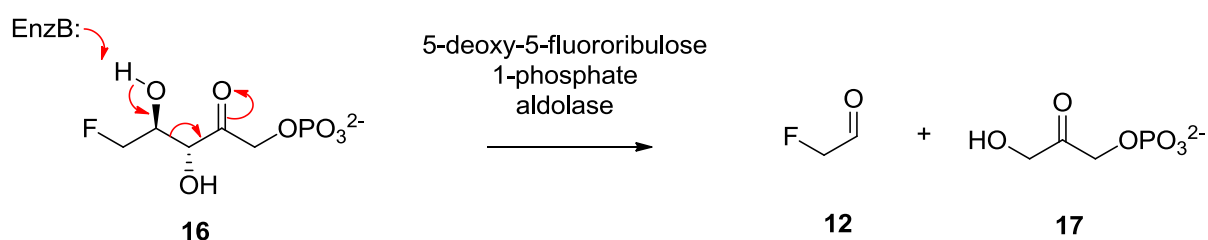
Scheme 1.13. The putative aldolase reaction for the generation of fluoroacetaldehyde from **16**.

To date, the aldolase involved in the fluorometabolite pathway in *S. cattleya* is uncharacterised at the protein and genetic level. However, the full genome of the organism has been recently sequenced and annotated and the putative fuculose aldolase gene sequence has become available. The over-expression of this and other candidate aldolases is underway to identify the last enzyme of the biosynthetic pathway.

1.3.2.6 The *in vitro* reconstitution of the fluorometabolite pathway

An experiment was designed to reconstruct the biosynthesis of 4-fluoro-L-threonine **3** *in vitro*.[74] This was carried out to demonstrate that the natural product **3** could be generated from KF and SAM **13**, through the reactions of five purified enzymes and co-factors alone.

The fluorinase, PNP, isomerase and fluoroacetaldehyde transaldolase from *S. cattleya* have an established over-expression system, however the aldolase gene had not been identified and a surrogate fuculose 1-phosphate aldolase from *S. coelicolor* was used in the experiment. Co-factors (Zn^{2+} ions and PLP) and reactants (KF, L-threonine) were supplied and the enzymatic reaction was carried out in phosphate buffer.[74]
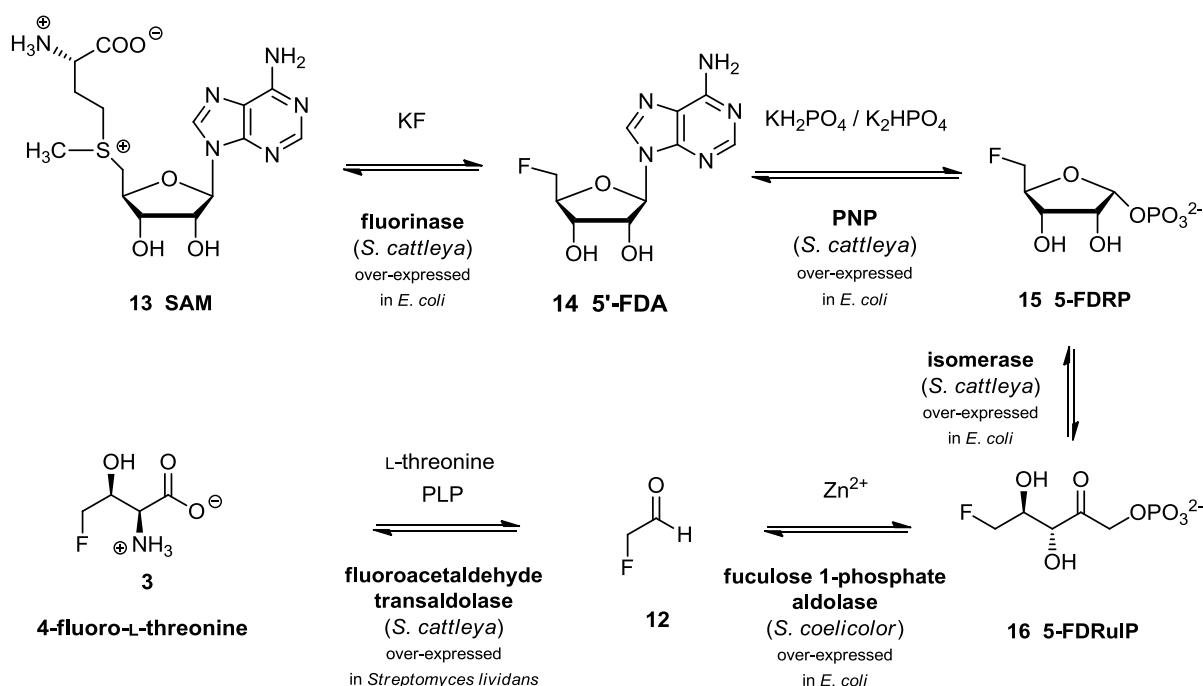
Scheme 1.14. The 4-fluorothreonine pathway *in vitro* reconstitution experiment.

The reaction was monitored by ^{19}F NMR and the production of each metabolite can be seen in the sequential enzyme omission experiments[73, 74] (Fig. 1.7). When all the enzymes were present the production of 4-fluorothreonine was observed *in vitro,* which substantially validated the *in vivo* pathway.

Fig. 1.7. The $^{19}F\{^{1}H\}$ NMRs of the 4-fluorothreonine pathway reconstruction experiment. The formation of the intermediates could be observed in the single-enz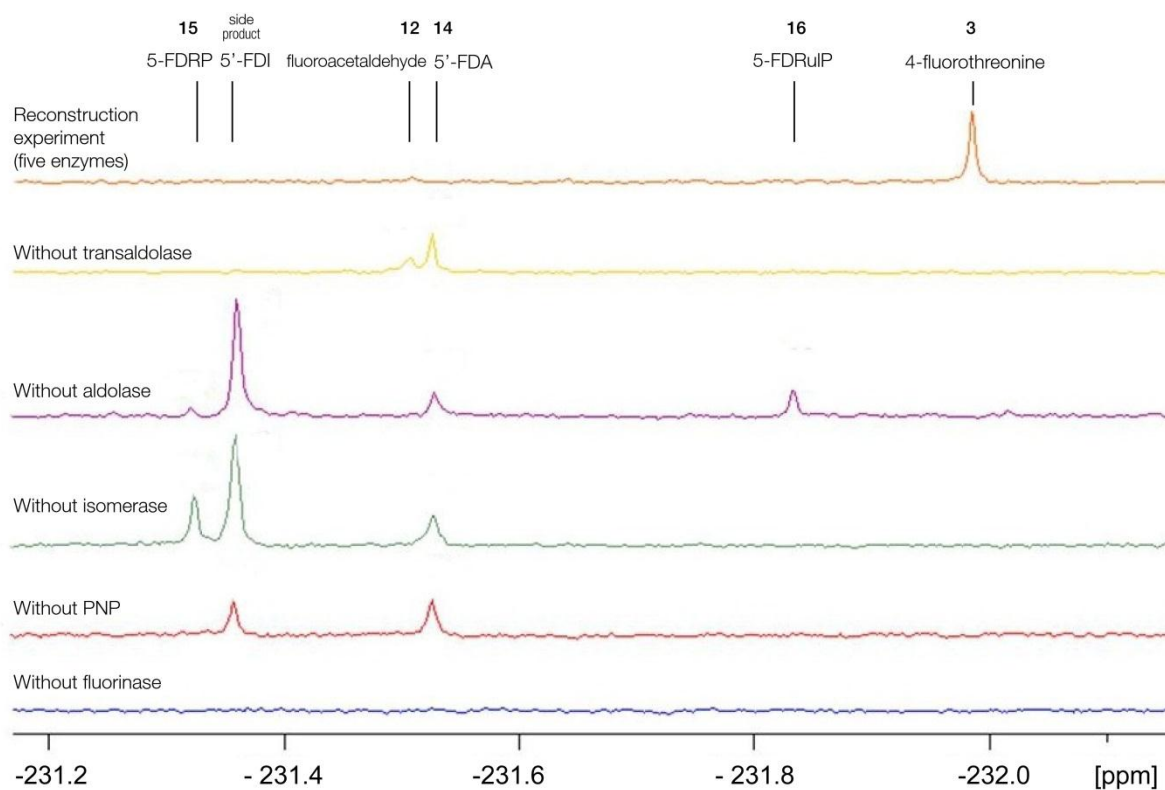yme omission experiments. The formation of 5'-FDI (5'-deoxy-5'-fluoroinosine) was due to a deaminase activity contamination in the protein preparation of the over-expressed PNP.[73]

1.3.2.7 The genetic studies of the fluorometabolite pathway

Successful cloning of the fluorinase gene in 2004[58] shortly after the identification of the enzyme[75] was a stepping stone for unravelling the details of fluorometabolite biosynthesis. The genetic information assisted in the biological studies of the genetic and molecular biology of fluorometabolite biosynthesis.
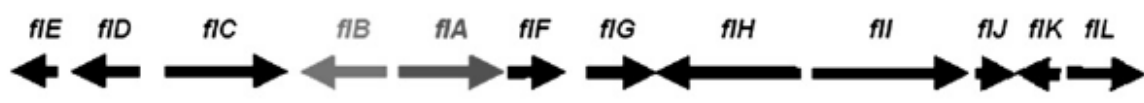
Secondary metabolism gene clusters are well known now in many strains stemming from the studies of the genetic arrangements in *Streptomyces coelicolor*.[76] All the genes involved in the biosynthesis, regulation and resistance of an antibiotic are usually located together on the chromosome, and this group of dedicated genes are known as a biosynthetic gene cluster.

It was thought that the genes for fluorometabolites might also be clustered, therefore Dr Jonathan B. Spencer and co-workers (University of Cambridge) sequenced the genes neighbouring the fluorinase by a gene-walking technique, and identified a 10 kb gene cluster[63] (known as the 'Spencer cluster') (Table 1.5). Within this cluster were twelve open reading frames (ORFs), including the fluorinase gene (*flA*) and a gene (*flB*) encoding for the purine nucleoside phosphorylase (PNP).

There are also other genes that may be involved in the regulation of the pathway, such as DNA binding regulatory proteins (*flE, flF, flG, flL*) and a Na^+/H^+ antiporter (*flH*). This antiporter may have a role involving fluoride uptake into the cell. However sequence homology does not confirm their functions. A fluoroacetyl-CoA thioesterase gene (*flK*) was also present, and may confer fluoroacetate resistance to the organism by breaking down fluoroacetyl-CoA **8** which leads to the toxic synthesis of 2-fluorocitrate **5** (Scheme 1.2). Interestingly, the complete biosynthetic pathway was

not featured within this gene cluster indicating that other genes must be located elsewhere on the chromosome remote from cluster.

Table 1.5. The fluorometabolite gene cluster (Spencer cluster).[63]

ORF	Start/Stop (bp)	Length (amino acids)	Annotated function
flE	130-795c	222	DNA binding regulatory protein
flD	857-1540c	216	Dehalogenase/Phosphatase
flC	1845-3036	397	MFS permease
flB	3057-3953c	299	**Purine nucleoside phosphorylase**
flA	4173-5069	299	**Fluorinase**
flF	5197-5751	185	DNA binding regulatory protein
flG	5951-6652	234	DNA binding regulatory protein
flH	6652-8052c	467	Na^+/H^+ antiporter
flI	8314-9780	489	Homocysteine lyase
flJ	9803-10195	131	DNA binding protein
flK	10592-10176c	139	Thioesterase/acyltransferase
flL	10700-11374	225	DNA binding regulatory protein

The full genome of *S. cattleya* was finally realised in 2011 with two groups completing a full genome sequence.[11, 12] The genome consists of a linear chromosome of 6.28 Mb (5822 genes) and a linear mega-plasmid of 1.81 Mb (1747 genes); the six biosynthetic genes involved in fluorometabolite biosynthesis are scattered on the genome without any recognisable pattern. A schematic illustration below shows the location of the six genes on the chromosome and the mega-plasmid. In particular, the transaldolase and

the putative aldolase are located on the megaplasmid, whereas the Spencer cluster is on the chromosome.
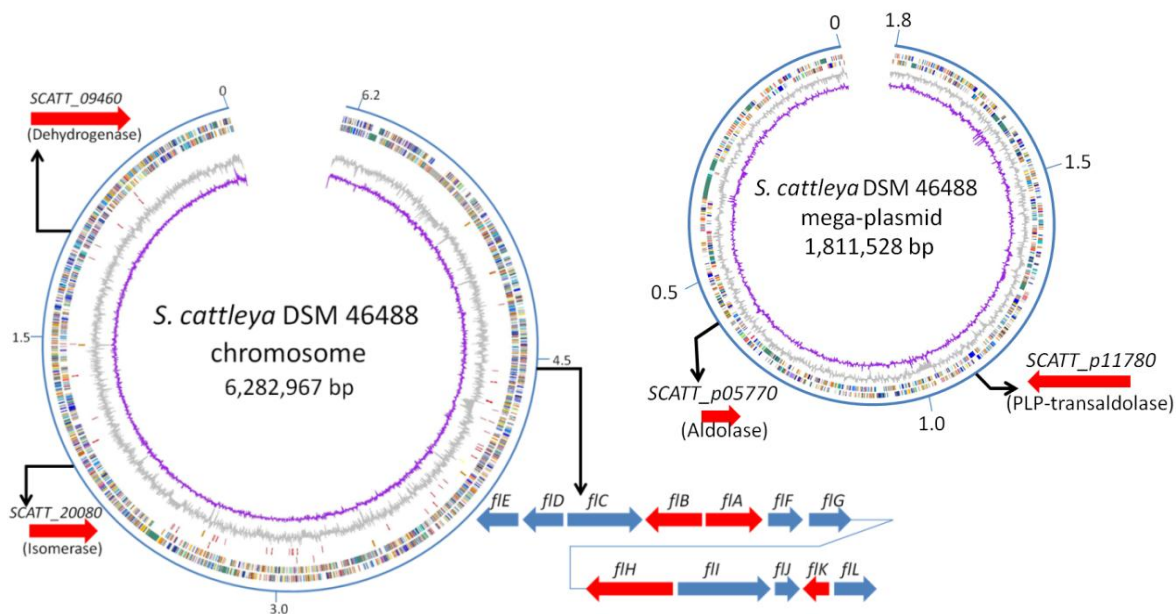
Fig 1.8. The full genome map of *S. cattleya*.

In order to validate the role of these six enzymes *in vivo* and other genes within the Spencer cluster, specific gene knock-out mutant strains of *S. cattleya* for all the relevant genes were prepared and the mutants were grown by fermentation and examined for fluorometabolite production.

The four mutants that had one of the first four enzymes of the pathway knocked out failed to produce any fluorometabolites, but the knock-out mutant of the threonine transaldolase was still able to produce fluoroacetate **2** consistent with the current understanding of branch point on the fluorometabolite biosynthesis pathway (Scheme 1.15).[77]
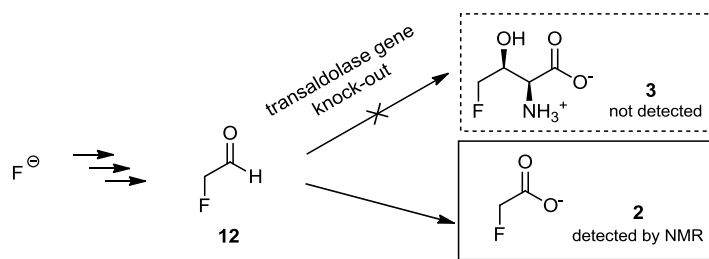
Scheme 1.15. The threonine transaldolase gene knock-out experiment.[77]

1.4 Acetylenic natural products

The alkyne functional group is found on a number of natural products isolated from bacteria, fungi, marine organisms, insects and some animals, and more than 2000 such structures have been isolated and characterised containing one or more triple bonds.[78] The group is essentially a rigid linear hydrophobic structure, which can impose a major influence on the shape and activity of the natural product.

The majority of acetylenic metabolites have been isolated from plant sources. The first such compound was (E)-dehydromatricaria ester **31** although its structure was unknown at the time.[79] The first structurally characterised acetylenic natural product was tariric acid **32** isolated in 1892.[80] As the isolation and analytical techniques for natural products advanced, an expansion occurred in the number of acetylenic compounds reported over the last 30 – 40 years. Most of these compounds are from plant sources especially from the family of Asteraceae and these compounds are largely related to fatty acid biosynthesis,[81, 82] many of which are polyacetylenic – containing multiple acetylene bonds in the structure.
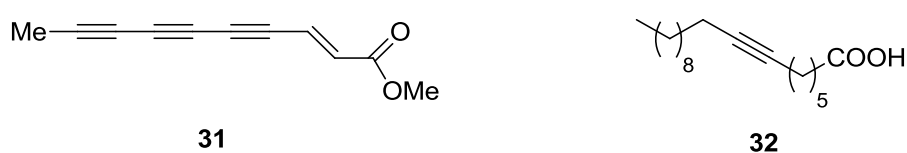
Fig 1.9. Representative acetylenic natural products.

The presence of the C≡C bond is not limited to long chain fatty acids, it is also found in a range of other natural products,[78] such as polyketides and non-ribosomal peptides (*incl.* enediynes, sporolides, jamaicamides), terpenoids, alkaloids, aromatic compounds derived from shikimate and small acetylenic amino acids.

1.4.1 Acetylenic fatty acids from plants and fungi

The majority of acetylenic natural products are elaborated from acetate-derived fatty acids through sequential desaturation of their alkyl chains. They comprise a very large number of compounds with varied chain-lengths and oxidative modifications. Many of these are polyacetylenic fatty acids, containing multiple alkynes in one chain. The acetylenes are also often stabilised by conjugation to adjacent alkene systems. Several examples are shown below (**33 - 36**).[83, 84, 85]

33

(*R*)-(-)-falcarinol
from *Apiaceae & Araliaceae* species

34

(*3R*, 8*S*)-falcarindiol
from *Apiaceae & Araliaceae* species

35

from *Chrysanthemum flosculosum* (plant)

36

(*R*)-(-)-marasin
from *Marasmius ramealis* (fungus)
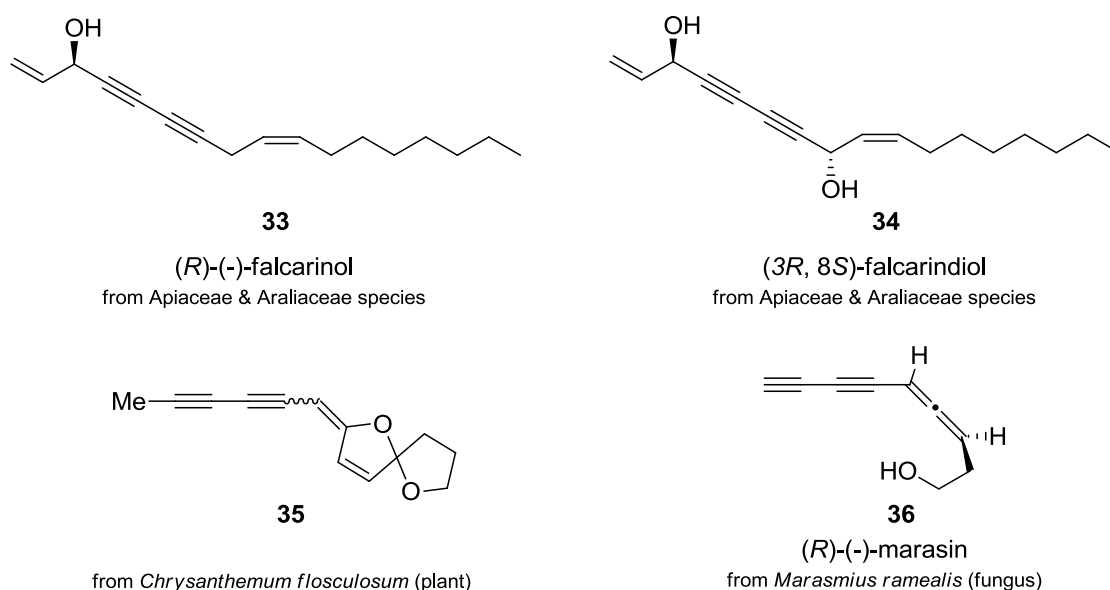
Fig. 1.10. Examples of polyacetylenes from plants and fungi.

In plants, these polyacetylenes are localised at specific locations thought to be related to their biological role. For instance, the polyacetylenes: falcarinol **33** and falcarindiol **34** were found to be concentrated in the exterior surface tissues of the root of the carrot plant (*Daucus carota*), where these compounds confer fungal resistance to the root.[86]

Falcarinol **33** has also been indicated as an active cancer-protective ingredient of the carrot and ginseng (*Panax quinquefolis*) plants and its use as an anti-cancer agent is under investigation as it was shown to induce apoptosis in large tumours.[87] However these beneficial properties are not expected to be general among the polyacetylenic lipids.

1.4.1.1 Biosynthetic pathways of acetylenic fatty acids

A large number of acetylenic compounds were isolated from plants and fungi by the pioneering work of Sir Ewart H. R. Jones (University of Oxford) and Professor Ferdinand Bohlmann (University of Berlin) in a span of around 30 years.[78] Collectively they created a library of over a thousand of these fatty acids derived acetylenic structures.

The biosynthesis of acetylenic compounds in plants and fungi were examined and all of them were found to have originated from three monoacetylenic fatty acid precursors: crepenynic acid **37**, stearolic acid **38** and tariric acid **32**. These precursors are the first entry points into secondary metabolism distinguished from fatty acid biosynthesis.

crepenynic acid
37

stearolic acid
38

tariric acid
32
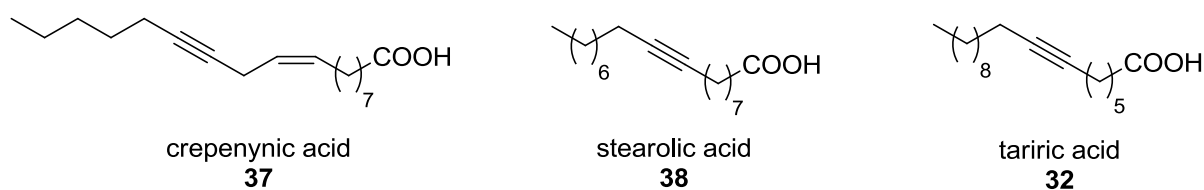
Fig 1.11. The monoacetylenic precursors to acetylenic natural products in plants and fungi.

The crepenynic acid pathway[88, 89] is the most common of the three. Crepenynic acid **37** is produced from C_{18} linoleic acid by an Δ^{12} acetylenase, which converts the C-12 alkene to an alkyne. Further desaturation of the lower alkyl chain into acetylenes can take place through two-step sequences of desaturase and acetylenase activities, generating both di- and tri- alkynes. The intermediates along this route are further elaborated into many acetylenic metabolites (Scheme 1.16).

The stearolic and tariric acid pathways are less widespread;[78] stearolic acid **38** is the product of an Δ^9 acetylenase on oleic acid (C_{18}) while tariric acid **32** arises from the reaction of petroselinic acid with an Δ^6 acetylenase. Both **38** and **32** are also further oxidised into polyacetylenes and these building blocks are then fashioned into the various metabolites.
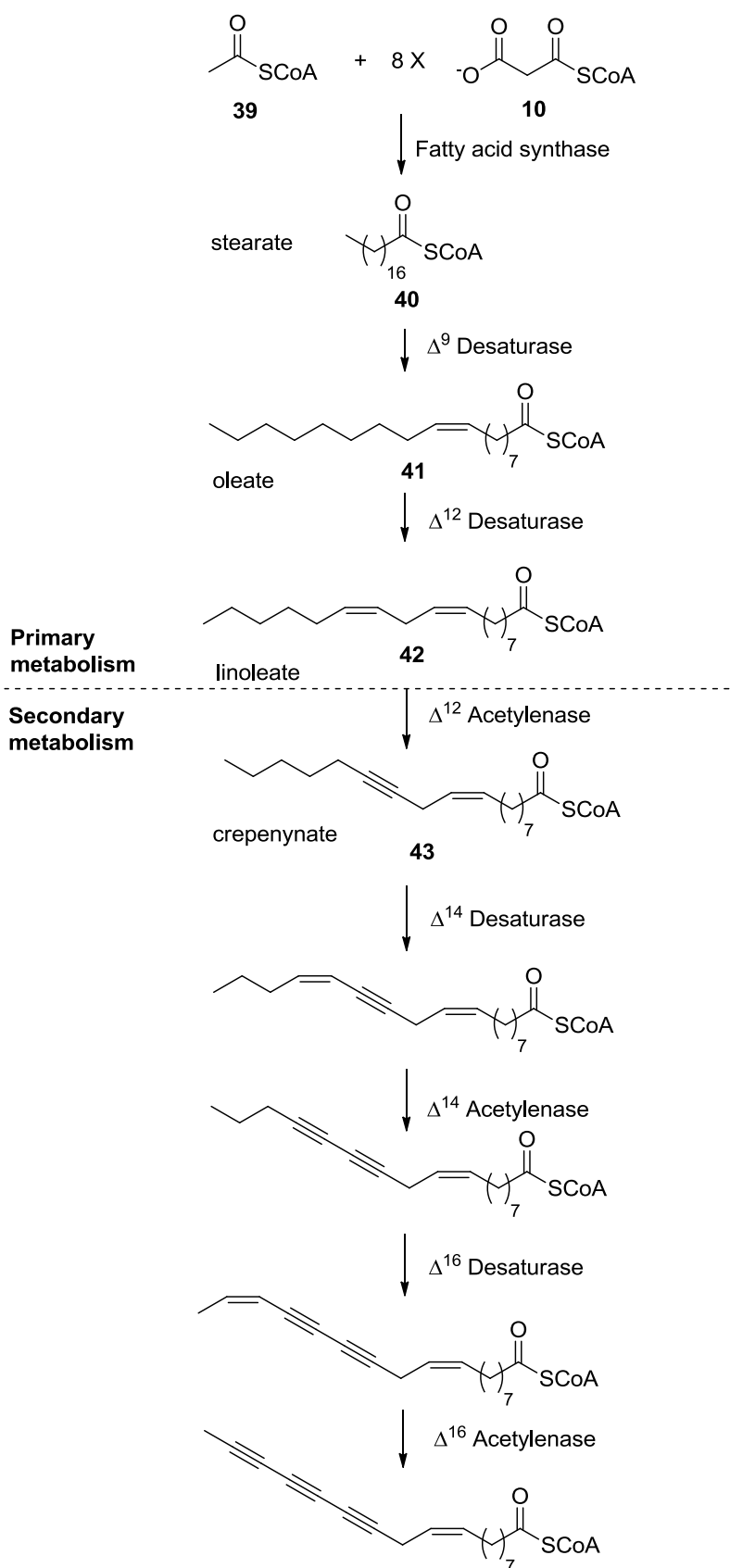
Scheme 1.16. The crepenynate pathway for acetylenic compounds in plants and fungi.[88, 89]

1.4.1.2 Acetylenase

Some acetylenases from plants and fungal sources have been cloned and expressed. The alkyne-forming reaction is a direct net dehydrogenation of a *cis*-alkene, which is coupled to NADPH oxidation and the reduction of oxygen to water.[90]
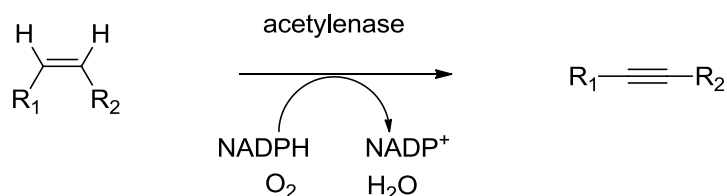
Scheme 1.17. The acetylenase reaction for the conversion of a (Z)-olefin to an acetylene.

Some mechanistic studies have been carried out with the Δ^{12} acetylenase from *Crepis alpina* which dehydrogenates the C-12 *cis*-alkene of linoleate **42**. A primary kinetic isotope effect (KIE) was measured for abstraction of H-12 ($k_H/k_D = 14.6$) and a secondary KIE for H-13 ($k_H/k_D = 1.25$), which suggests a stepwise hydrogen abstraction commencing with H-12, to generate either a cationic or radical intermediate species, followed by the loss of H-13 and formation of the new π bond (Scheme 1.18). [91] No further evidence has been published and the detail of this enzymatic mechanism is still unknown.
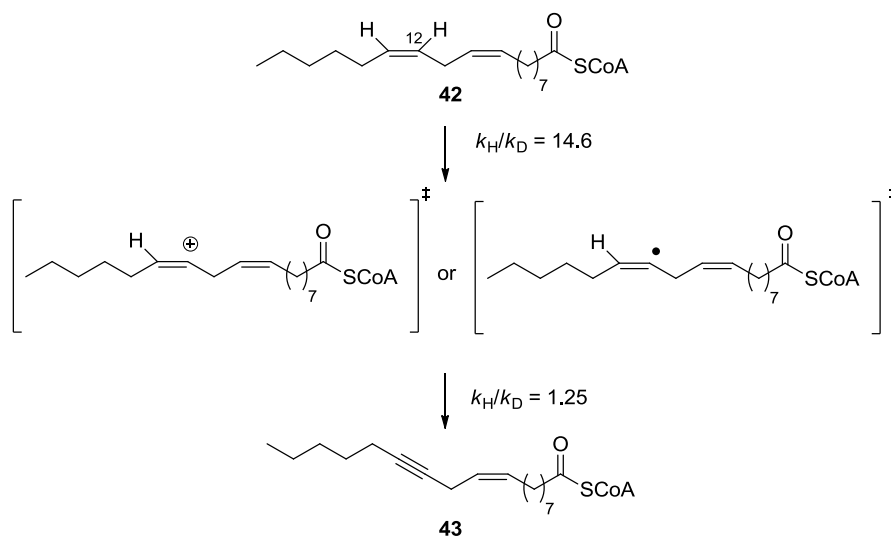
Scheme 1.18. The kinetic isotope studies on the mechanism of the acetylenase.[91]

1.4.2 Terminal acetylene metabolites from marine sources

When alkynes are found in natural products, they most frequently occur as internal acetylenes, consistent with derivation from the desaturation pathways described above. Terminal alkynes are comparatively rare and their origin may involve new enzyme chemistry.

Most of the natural products with terminal acetylenes have been isolated from marine organisms. The marine cyanobacteria *Lyngbya majuscula* was found to produce an unusual bromine terminated acetylene, jamaicamide A **44** and its debromo analogue **45**, for which a study of their biosynthesis has taken place.[92] An array of cyclodepsipeptides was also isolated from the same species.[93, 94] They contained a terminal acetylenic branch extending from the cyclic peptide core. Similar compounds were found in higher marine organisms that feed on cyanobacteria, such as from mollusc[95, 96] and sea hares.[97] In addition, some very long chain (C_{32} to C_{41}) terminal polyacetylenic alcohols and acids called triangulynes **46** were extracted from the marine sponge, *Pellina triangulata*[98] and a group of long chain polyacetylenic propargyl alcohols petroformynes **47** were isolated from the *Petrosia* sponges.[99, 100]

44 (X = Br) Jamaicamide A
45 (X = H) Jamaicamide B

47 Petroformyne A

46 Triangulynes D
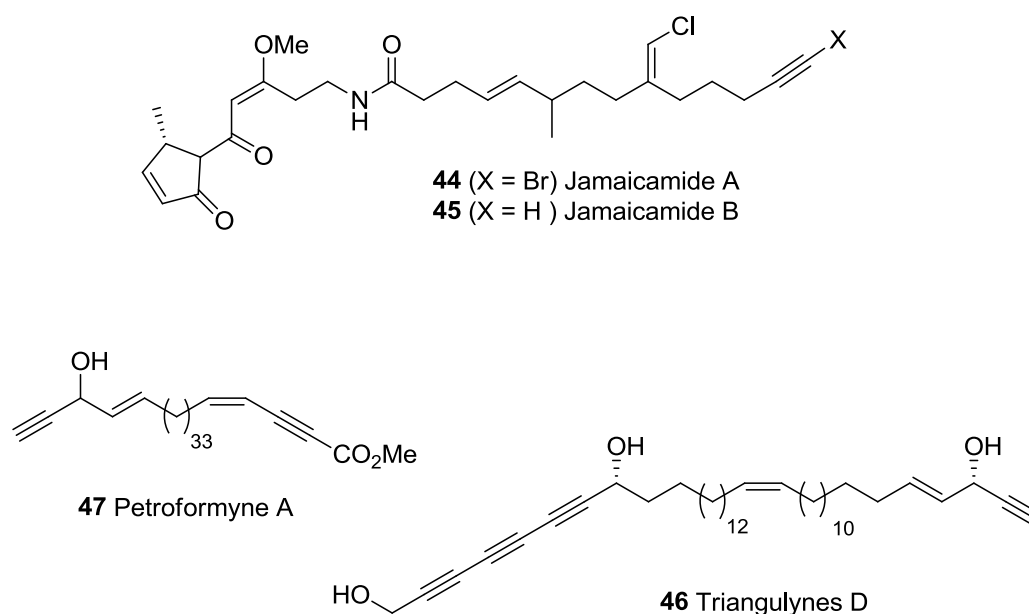
Fig. 1.12. Examples of terminal alkyne marine natural products.

1.4.2.1 Jamaicamides and a putative terminal acetylenase

L. majuscula produces several metabolites with terminal acetylenic groups, however the one that has attracted most attention is jamaicamide A **44** due to the unusual bromine cap on the ω-acetylene. The biosynthetic gene cluster with 17 ORFs was identified for this PKS/NRPS metabolite.[92]

The acetylene is installed prior to the PKS/NRPS reactions, and the starter unit utilised by the PKS appears to be hex-5-ynoic acid, which is efficiently converted into hex-5-ynyl-ACP by over-expressed JamA, an acyl-ACP synthetase which does not accept C_4, C_5, C_7 or C_8 substrates. One of the genes (*jamB*) encoded a desaturase that has a 48 % homology to the Δ^9-desaturase Ole1p from yeast.[92]

JamB is believed to be a terminal desaturase/acetylenase, the first example of such an enzyme, although no direct evidence has yet been published to support this.

The acetylenation of an unnatural substrate undec-10-enoate by *L. majuscula* had been observed in a feeding experiment and supports the hypothesis of terminal acetylenases at work.[92]

Terminal acetylenase activity is rare and a full study of JamB may also shed light on the biosynthesis of other terminal alkyne natural products, including β-ethynylserine **4** from *S. cattleya.*

1.4.2.2 Acetylenic cyclodepsipeptides and lipopeptides

Alongside the Jamaicamides, a wide array of cyclodepsipeptides and lipopeptides containing a terminal acetylene unit was isolated from *L. majuscula*.[93, 94] Other marine organisms, such as mollusc (*Onchidium* sp.[96], *Philinopsis speciosa*[95]) and sea hare (*Stylocheilus longicauda*[97]) that feed on blue-green algae such as *L. majuscula* were also found to contain these kinds of cyclic peptides but clearly the acetylenic bond may not be a *de novo* product but instead originated from their diet.

Three examples are shown below that were found in *Lyngbya* species (**48 - 50**).[101- 103]
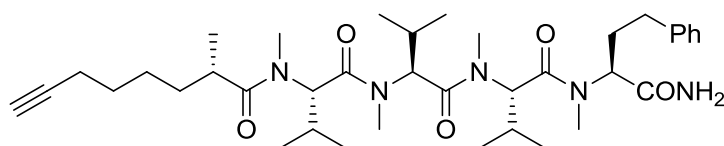
48 Dragonamide A
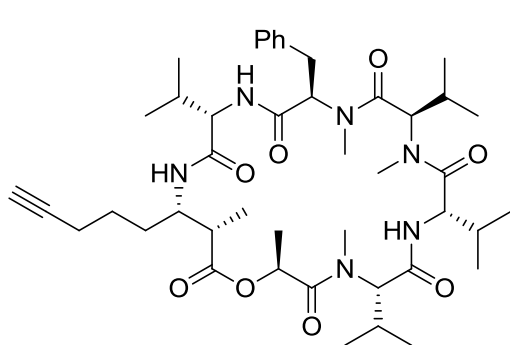
49 Ulongapeptin
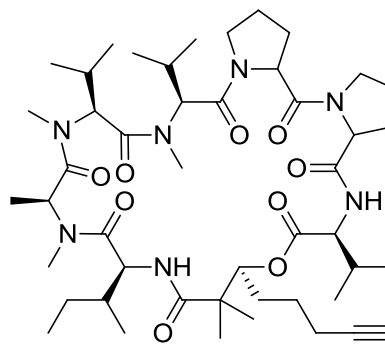
50 Wewakpeptin A

Fig. 1.13. Examples of terminal acetylenic cyclodepsipeptides and lipopeptides.

These peptides are products of NRPS or PKS/FAS and NRPS, which condense amino acids or acyl units together to build these large molecules. From the study of JamA and JamB, it seemed most likely that the ω-acetylene is already present on the amino acids or fatty acid before they are utilised in macrocycle assembly.[92] This suggests that many terminal acetylenic amino acids are biosynthesised by *L. majuscula* and other marine organisms. Some of the amino acid units found in these cyclic compounds are shown below in Figure 1.14 (**51 – 53**).[102, 97, 104]
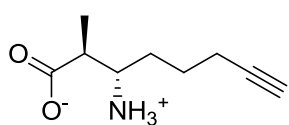
51 from Ulongapeptin
(marine cyanobacteria)
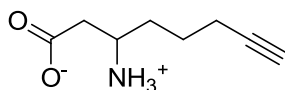
52 from Dolastatin-17
(sea hare)
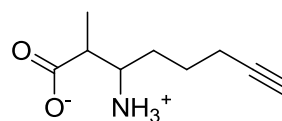
53 from Guineamide C
(Maoya)

Fig 1.14. Amino acid units found in some cyclodepsipeptides (not isolated as natural products).

1.4.3 Terminal acetylenic amino acids

Naturally occurring terminal acetylenic amino acids have been found in several organisms where they excrete or accumulate the amino acid as a toxin. The fungus *Sclerotium rolfsii* was found to contain high levels of β-ethynyl-L-serine **4**, which is responsible for the poisoning of poultry fed on maize contaminated with the fungus (LD$_{50}$ chicken, 150 mg/kg).[10] The same compound was also isolated from fermentation culture of *S. cattleya*.[9]

From the seeds of the Longan fruit (*Euphoria longan*), three C$_7$ terminal acetylenic amino acids **54** - **56** were isolated.[105] More recently, two acetylenic D-amino acids **57, 58** were found to be the lethal component in the Chinese mushrooms *Trogia venenata*.[106] Inadvertent consumption of this mushroom is estimated to have caused the death of 260 villagers over a 30-year period in the forests of Yunnan, China.
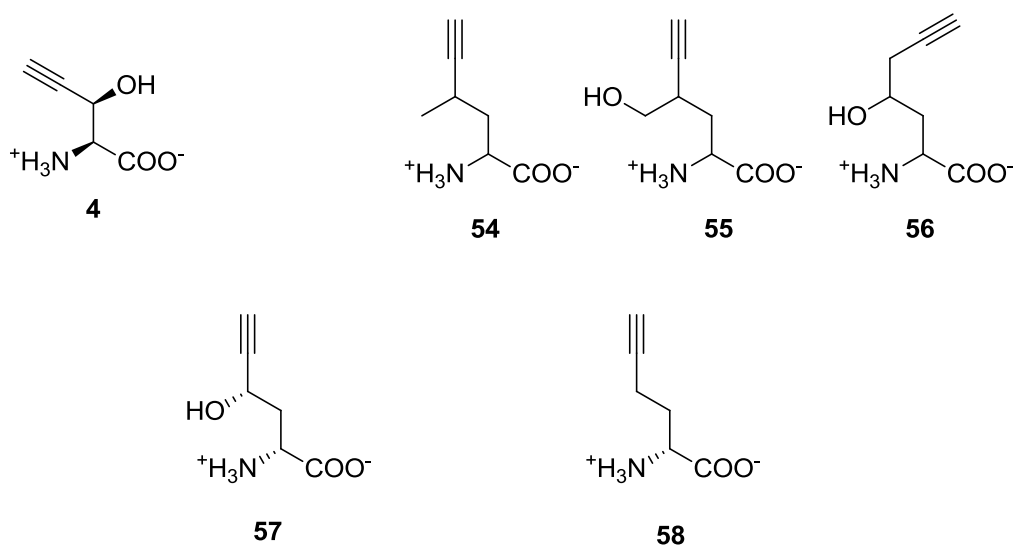
Fig 1.15. Some of the terminal acetylenic α-amino acids from bacteria, fungi and plant sources.

1.4.3.1 β-Ethynylserine and fluorometabolite biosynthesis

During the purification of 4-fluoro-L-threonine **3** from the cultures of *S. cattleya,*

β-ethynyl-L-serine **4** was identified as a co-product.[9] β-Ethynyl-L-serine was shown to

have good antibiotic activity against *Pseudomonas aeruginosa,* this bioactivity could be

reversed by supplementing the media with L-threonine.[9]
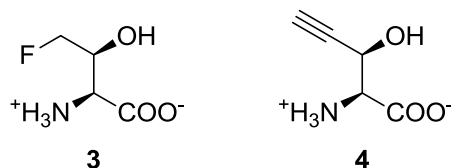
Fig 1.16. Two rare amino acids extracted from the fermentation broth of *S. cattleya.*

The structures of the two amino acids **3** and **4** are very similar and at the early stages of

the studies of 4-fluorothreonine biosynthesis, β-ethynylserine was investigated as a

probable precursor of **3**.[44] After some isotopic labelling experiments were carried out,

it became apparent that they did not arise from the same pathway.[107] Nevertheless an

interest in this acetylenic amino acid was established and from the labelling

experiments, it is known that both glycine and pyruvate are not precursor to **4** in *S.*

cattleya.[107]

A total synthesis and a method for detection of β-ethynyl-L-serine **4** from the

fermentation culture will be reported in this thesis in *Chapter 4.*

1.5 Chapter 1 References

1. J. B. McAlpine, *J. Nat. Prod.*, 2009, **72**, 566–572.

2. C. C. Thornburg, T. M. Zabriskie, K. L. McPhail, *J. Nat. Prod.*, 2010, **73**, 489–499.

3. R. A. Maplestone, M. J. Stone, D. H. Williams, *Gene*, 1992, **115**, 151-157.

4. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood, *Practical Streptomyces Genetics*, Norwich, The John Innes Foundation, 2000.

5. J. S. Kahan, F. M. Kahan, R. Goegelman, S. A. Currie, M. Jackson, E. O. Stapley, T. W. Miller, A. K. Miller, D. Hendlin, S. Mochales, S. Hernandez, H. B. Woodruff, J. Birnbaum, *J. Antibiotics*, 1979, **32**, 1-12.

6. M. Sanada, T. Miyano, S. Iwadare, J.M. Williamson, B.H.J. Arison, L. Smith, A.W. Douglas, J.M. Liesch, E. Inamine, *J. Antibiotics*, 1986, **39**, 259-265.

7. J. S. C. Marais, *Onderstepoort J. Vet. Sci. Anim. Ind.*, 1943, **18**, 203-206.

8. P. B. Oelrichs, T. McEwan, *Nature*, 1961, **190**, 808-824.

9. M. Sanada, T. Miyano, S. Iwadare, *J. Antibiotics* 1986, **39**, 304 – 305.

10. H. C. Potgieter, N. M. J. Vermeulen, D. J. J. Potgieter, H. F. Strauss, *Phytochemistry*, 1977, **16**, 1757 – 1759.

11. H. Y. Ou, P. Li, C. Zhao, D. O'Hagan, Z. Deng , *Nucleotide Sequence submitted to the EMBL/GenBank/DDBJ databases*, 2011.

12. V. Barbe, M. Bouon, S. Mangenot, B. Badet, J. Poulain, B. Segurens, D. Valleet, P. Marliere, J. Weissenbach, *J. Bacteriol.*, 2011, **193**, 5055-5056.

13. G. W. Gribble, *J. Chem. Edu.*, 2004, **81**, 1441-1449.

14. D. H. Williams, *Nat. Prod. Rep.*, 1996, **13**, 469-477.

15. D. H. Williams and B. Bardsley, *Angew. Chem. Int. Ed.*, 1999, **38**, 1173-1193.

16. C. S. Neumann, D. G. Fujimori, C. T. Walsh, *Chem. Biol.*, 2008, **15**, 99-109.

17. L. P. Hager, D. R. Morris, F. S. Brown, H. Eberwein, *J. Biol. Chem.*, 1966, **241**, 1769-1777.

18. A. Butler, J. N. Carter-Franklin, *Nat. Prod. Rep.*, 2004, **21**, 180-188.

19. M. N. Isupov, A. R. Dalby, A. A. Brindley, Y. Izumi, T. Tanabe, G. N. Murshudov, J. A. Littlechild, *J. Mol. Biol.*, 2000, **299**, 1035-1049.

20. S. Keller, T Wage, K. Hohaus, M. Holzer, E. Eichhorn, K. H. van Pee, *Angew. Chem. Int. Ed.*, 2000, **39**, 2300-2302.

21. A. S. Eustaquio, F. Pojer, J. P. Noel, B. S. Moore, *Nat. Chem. Biol.*, 2008, **4**, 69-74.

22. D. O'Hagan, C. Schaffrath, S. L. Cobb, J. T. Hamilton, C. D. Murphy, *Nature*, 2002, **416**, 279.

23. F. H. Vaillancourt, J. Yin, C. T. Walsh, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 10111-10116.

24. J. M. Bollinger Jr., J. C. Price, L. M. Hoffart, E. W. Barr, C. Krebs, *Eur. J. Inorg. Chem.*, 2005, 4245-4254.

25. L. M. Hoffart, E. W. Barr, R. B. Guyer, J. M. Bollinger Jr., C. Krebs, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 14738-14743.

26. D. P. Galonic, F. H. Vaillancourt, C. T. Walsh, *J. Am. Chem. Soc.*, 2006, **128**, 3900-3901.

27. J. G. Stark, H. G. Wallace, *Chemistry Data Book Second Edition*, Hodder Education, 1989.

28. R. A. Peters, M. Shorthouse, *Nature*, 1967, **216**, 80-81.

29. R. A. Peters, M. Shorthouse, *Nature*, 1971, **231**, 123-124.

30. X.-H. Xu, G.-M. Yao, Y.-M. Li, J.-H. Lu, C.-J. Lin, X. Wang, C-H Kong, *J. Nat. Prod.*, 2003, **66**, 285-288.

31. J. S. C. Marias, *Onderstepoort J. Vet. Res.*, 1944, **20**, 67-73.

32. R. A. Peters, R.W. Wakelin, P. Buffa, L.C. Thomas, *Proc. R. Soc. London, Ser. B.,* 1953, **140**, 497.

33. R. A. Peters, R.J. Hall, *Biochem. Pharmacol.,* 1959, **2**, 25.

34. J. T. G. Hamilton, D. B. Harper, *Phytochemistry,* 1997, **44**, 1179.

35. W. W. Christie, J. T. G. Hamilton, D. B. Harper, *Chem. Phys. Lipids,* 1998, **97**, 41.

36. S. O. Thomas, V. L. Singleton, J. A. Lowery, R. W. Sharpe, L. M. Pruess, J. N. Porter, J. H. Mowat, N. Bohonos, *Antibiotics Ann.,* 1957, **716**, 1956-7.

37. G. O. Morton, J. E. Lancaster, G. E. Van Lear, W. Fulmor, W. E. Meyer, *J. Am. Chem. Soc.,* 1969, **91**, 1535.

38. I. D. Jenkins, J. P. H. Verheyden, J. G. Moffatt, *J. Am. Chem. Soc.,* 1976, **98**, 3346.

39. R. J. Hall, R. B. Cain, *New Phytol.,* 1972, **71**, 839-853.

40. R. J. Hall, *New Phytol.,* 1972, **71**, 855-871.

41. L. E. Twigg, D. R. King, L. H Bowen, G. R. Wright, C. T. Eason, *Aus. J. Bot.,* 1996, **44**, 411-412.

42. M. M. De Oliveira, *Cell. Mol. Life Sci.,* 1963, **19**, 586-587.

43. T. Vartiaininen, P. Kauranen, *Analytica Chimica Acta.,* 1984, **157**, 91.

44. D. O'Hagan, D. B. Harper, *Nat. Prod. Rep.,* 1994, **11**, 123-133.

45. R. A. Peters, R. W. Wakelin, P. Buffa, *Proc. Roy. Soc. B,* 1953, **140**, 497-506.

46. J. F. Morrison, R. A. Peters, *Biochem. J.,* 1954, **58**, 473-479.

47. R. J. Dummel, E. Kun, *J. Biol. Chem.,* 1969, **11**, 2966-2969.

48. H. L. Carrell, J. P. Glusker, J. J. Villafranca, A. S. Mildvan, R. J. Dummel, E. Kun, *Science,* 1970, **170**, 1412-1414.

49. H. Lauble, M. C. Kennedy, M. H. Emptage, H. Beinert, C. D. Stout, *Proc. Natl. Acad. Sci. USA,* 1996, **93**, 13699-13703.

50. M. V. B. Dias, F. Huang, D. Y. Chirgadze, M. Tosin, D. Spiteller, E. F. V. Dry, P. F. Leadley, J. B. Spencer, T. L. Blundell, *J. Biol. Chem.*, 2010, **285**, 22495-22504.

51. J. J. M. Meyer, N. Grobbelaar, R. Vleggaar, A. I. Louw, *J. Plant. Physiol.*, 1992, **139**, 369.

52. S. J. Moss, C. D. Murphy, J. T. G. Hamilton, W. C. McRoberts, D. O'Hagan, C. Schaffrath, D. B. Harper, *Chem. Commun.*, 2000, 2281-2282.

53. C. D. Murphy, S. J. Moss and D. O'Hagan, *Appl. Environ. Microbiol.*, 2001, **67**, 4919-4921.

54. J. A. Robinson, D. Gani, *Nat. Prod. Rep.*, 1985, **2**, 293-319.

55. C. D. Murphy, D. O'Hagan, C. Schaffrath, *Angew. Chem. Int. Ed.*, 2001, **40**, 4479-4481.

56. C. Schaffrath, *Ph.D. Thesis*, University of St Andrews, 2002.

57. T-C Chou, P. Talalay, *Biochemistry*, 1972, **11**, 1065-1073.

58. C. Dong, F. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O´Hagan, J. H. Naismith, *Nature*, 2004, **427**, 561-465.

59. C. D. Cadicamo, J. Courtieu, H. Deng, A. Meddour and D. O'Hagan, *ChemBioChem*, 2004, **5**, 685-690.

60. S. L. Cobb, *Ph.D. Thesis*, University of St Andrews, 2005.

61. S. L. Cobb, H. Deng, J. T. G. Hamilton, R. McGlinchey, D. O´Hagan, *Chem. Commun.*, 2004, 592-593.

62. R. P. McGlinchley, *Ph.D. Thesis*, University of St Andrews, 2006.

63. F. Huang, S. F. Haydock, D. Spiteller, T. Mironenko, T-L Li, D. O'Hagan, P. F. Leadlay, J. B. Spencer, *Chem. Biol.*, 2006, **13**, 475-484.

64. F. R. Tabita, T. E. Hanson, H. Li, S. Satagopan, J. Singh, S. Chan, *Microbiol. Mol.Biol. Rev.*, 2007, **71**, 576-599.

65. A. Sekowska, A. Danchin, *BMC Microbiology*, 2002, **2**, 1-14.

66. Y. Dai, T. C. Pochapsky, R. H. Abeles, *Biochemistry*, 2001, **40**, 6379-6387.

67. I. Pirkov, J. Norbeck, L. Gustafsson, E. Albers, *FEBS J.*, 2008, **275**, 4111-4120.

68. C. Y. Lai, B. L. Horecker, *Essays Biochem.*, 1972, **8**, 149-178.

69. M. Onega, R.P. McGlinchey, H. Deng, J.T.G. Hamilton, D. O'Hagan, *Bioorg. Chem.*, 2007, **35**, 375-385.

70. M. Onega, *Ph.D. Thesis*, University of St Andrews, 2009.

71. R. Schoevaart, F. van Rantwijk, R. A. Seldon, *J. Org. Chem.*, 2000, **65**, 6940-6943.

72. R. Schoevaart, F. van Rantwijk, Roger A. Sheldon, *Chem. Commun.*, 1999, **24** 2465-2466.

73. S. Cross, *Ph. D. thesis*, University of St Andrews, 2008.

74. H. Deng, S. M. Cross, R. P. McGlinchey, J. T. G. Hamilton, D. O'Hagan, *Chem. Biol.*, 2008, **15**, 1268-1276.

75. C. Schaffrath, H. Deng, D. O´Hagan, *FEBS Lett.*, 2003, **547**, 111-114.

76. S. D. Bentley, K. F. Chater, A. M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, D. A. Hopwood., *Nature*, 2002, **417**, 141-147.

77. C. Zhao, P. Li, Z. Deng, H. Y. Ou, R. P. McGlinchey, D. O'Hagan, *Bioorg. Chem.*, 2012, **44**, 1-7.

78. R. E. Minto, B. J. Blacklock, *Prog. Lipid Res.*, 2008, **47**, 233-306.

79. M. A. Arnaud, *Bull. Soc. Chim. France,* 1902, 489-496.

80. M. A. Arnaud, *Bull. Soc. Chim. France,* 1892, 233-224.

81. F. Bohlmann, T. Burkhardt, C. Zdero, *Naturally occurring acetylenes,* London, Academic Press, 1973.

82. L. P. Christensen, J. Lam, *Phytochemistry,* 1991, **30**, 2453-2476.

83. P. M. Boll, L. Hansen, *Phytochemistry,* 1987, **26**, 2955-2956.

84. D. G. Crosby, N. Aharonson, *Tetrahedron,* 1967, **23**, 465-472.

85. Y. Zhang, Y. Wu, *Org. Biomol. Chem.,* 2010, **8**, 4744-4752.

86. B. Garrod, B. G. Lewis, *Physiol, Plant Pathol.,* 1978, **13**, 241-246.

87. M. Kobaek-Larsen, L. P. Christensen, W. Vach, J. R-Hoitinga, K. Brandt, *J. Agric. Food Chem.,* 2005, **53**, 1823-1827.

88. W. G. Haigh, A. T. James, *Biochim. Biophys. Acta.,* 1967, **137**, 391-392.

89. B. J. Blacklock, B. E. Scheffler, M. R. Shepard, N. Jayasuriya, R. E. Minto, *J. Biol. Chem.,* 2010, **285**, 28442-28449.

90. P. H. Buist, *Nat. Prod. Rep.,* 2004, **21**, 249-262.

91. D. W. Reed, D. R. Polichuk, P. H. Buist, S. J. Ambrose, R. J. Sasata, C. K. Savile, *J. Am. Chem. Soc.,* 2003, **125**, 10635-10640.

92. D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Gorger, M. A. Roberts, W. H. Gerwick, *Chem. Biol.,* 2004, **11**, 817-833.

93. S. Bunyajetpong, W. Y. Yoshida, N. Sitachitta, K. Kaya, *J. Nat. Prod.,* 2006, **69**, 1539-1542.

94. L. M. Nogle, W. H. Gerwick, *J. Nat. Prod.,* 2002, **65**, 1539-1542.

95. M. T. Reese, N. K. Gulavita, Y. Nakao, M. T. Hamann, W. Y. Yoshida, S. J. Coval, *J. Am. Chem. Soc.,* 1996, **118**, 11081-11084.

96. J. Rodriguez, R. Fernandez, E. Quiñoá, R. Riguera, C. Debitus, B. P. Onchidin, *Tetrahedron Lett.,* 1994, **35**, 9239-9242.

97. G. R. Petit, J-P Xu, F. Hogan, R. L. Cerny, *Heterocycles*, 1998, **47**, 491-497.

98. J. R. Dai, Y. F. Hallock, J. H. Cardellina, G. N. Gray, M. R. Boyd, *J. Nat. Prod.*, 1996, **59**, 860-865.

99. G. Cimino, A. De Giulio, S. de Rosa, S. de Stefano, G. Sodano, *J. Nat. Prod.*, 1985, **48**, 22-27.

100. Y. Guo, M. Gavagnin, E. Trivellone, G. Cimino, *J. Nat. Prod.*, 1995, **58**, 712-722.

101. K. L. McPhail, J. Correa, R. G. Linington, J. Gonzalez, E. Ortega-Barria, T. L. Capson, *J. Nat. Prod.*, 1998, **61**, 529-533.

102. P. G. Williams, W. Y. Yoshida, M. K. Quon, R. E. Moore, V. J. Paul, *J. Nat. Prod.*, 2003, **69**, 1539-1542.

103. B. Han, D. E. Goeger, C. S. Maier, W. H. Gerwick, *J. Nat. Prod.*, 2005, **70**, 3133-3139.

104. L. T. Tan, N. Sitachitta, W. H. Gerwick, *J. Nat. Prod.*, 2003, **66**, 764-771.

105. M-L Sung, L. Fowden, D. S. Millington, R. C. Sheppard, *Phytochemistry*, 1969, **8**, 1227-1233.

106. Z-Y Zhou, G-Q Shi, R. Fontaine, K. Wei, T. Feng, F. Wang, G-Q Wang, Y. Qu, Z-H Li, Z-J Dong, H-J Zhu, Z-L Yang, G Zeng, and J-K Liu, *Angew. Chem. Int. Ed.*, 2012, **51**, 2368-2370.

107. J. T. G. Hamilton, *Ph. D. Thesis*, Queen's University of Belfast, 1998.

2. Crystallisation of the 5-FDRPi isomerase

2.1 Aldose-ketose isomerases

Aldose-ketose isomerases mediate the reversible isomerisation between an α-hydroxyaldose and its corresponding α-hydroxyketose (Scheme 2.1). This is an important reaction found in primary metabolic pathways.
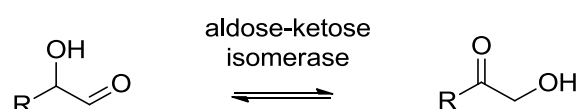
Scheme 2.1. Generalised reaction scheme of aldose-ketose isomerases.

There are two established mechanisms for this class of enzyme,[1] the most common involves a *cis*-enediol exemplified by triose phosphate isomerase. This enzyme interconverts dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (G3P) in the glycolysis pathway.[2]

There is evidence that the xylose isomerase from *Streptomyces olivochromogenes* that interconverts D-xylose and D-xylulose or D-glucose and D-fructose is mediated by a metal-catalysed 1,2-hydride shift mechanism.[3]

The isomerase on the fluorometabolite pathway of *S. cattleya* is also an aldose-ketose isomerase. At the genetic level it resembles 5-methylthioribose 1-phosphate isomerase (MTRPi) from the methionine salvage pathway in its chemical transformation. This Chapter describes the cloning of the isomerase gene and over-expression of soluble isomerase protein. Several mechanisms for this type of transformation are found in nature and these are reviewed in the next sections.

2.1.1 *cis*-Enediol mechanism in enzymatic isomerisation

A *cis*-enediol mechanism has been proposed for triose phosphate isomerase (Scheme 2.2).[2] It involves two tautomerisation reactions and proceeds through an intermediate *cis*-enediol motif. It was shown by isotope labelling studies that the pro-*R* proton on C-1 of DHAP **17** is abstracted by the enzyme utilising a single glutamate residue as general acid/general base.[4, 5] Since the Glu residue easily undergoes proton exchange with the solvent, when the reaction is carried out in D_2O, deuterium incorporation is observed at C-3 of DHAP **17** and C-2 of G3P **26**.[6, 7]

Scheme 2.2. The *cis*-enediol isomerase mechanism for the triose phosphate isomerase.[5]

2.1.2 1,2-Hydride shift mechanism in enzymatic isomerisation

A 1,2-hydride shift mechanism has been proposed for the xylose isomerase from *Streptomyces olivochromogenes*. This was deduced on the basis of the characterisation of xylose isomerase and its crystal structure to 0.95 Å resolution.[3] The enzyme active site contains two divalent metal ions which are predicted to be involved in the catalysis. They are secured by ligand coordination and act to stabilise the transition state of the reaction (Fig 2.1). The proposed mechanism is supported by the observation that deuterium exchange of the α-proton does not take place when the reaction is carried out in D_2O.[8] This is inconsistent with a *cis*-enediol mechanism and thus the working hypothesis favours a 1,2-hydride migration.
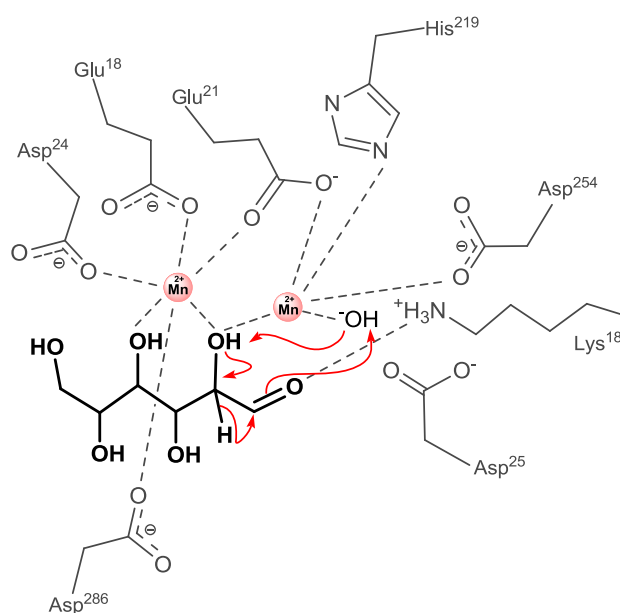

Fig 2.1. The active site arrangement of the xylose isomerase. A hydride shift is proposed as shown by red curly arrows.

2.2 Ribose 1-phosphate isomerases

2.2.1 The methionine salvage pathway

Organic sulfur is a precious commodity inside a cell, required to build essential molecules such as the amino acids L-methionine and L-cysteine, and co-factors such as SAM and coenzyme A. The assimilation of sulfur is energy-consuming and in order to conserve the thiomethyl group within the cell, most organisms including bacteria, plants and animals have a salvage pathway in place. This is known as the methionine salvage pathway.

The methionine moiety of SAM is metabolised in various ways. During the biosynthesis of spermidine **59**, SAM undergoes decarboxylation and then the aminopropyl group is subjected to nucleophilic attack from putrescine to generate spermidine (Scheme 2.3).[9] The side-product 5'-methylthioadenosine **60** is progressed around the salvage pathway which sees the thiomethyl group recirculated as L-methionine (Scheme 2.4).[10,11]
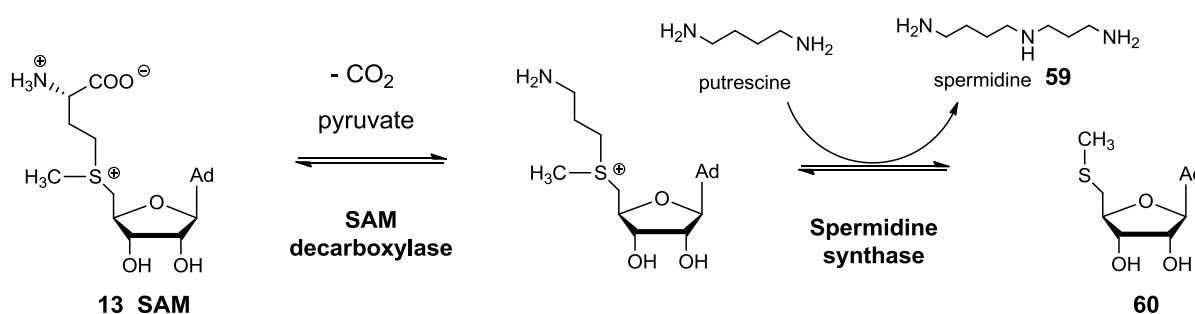
Scheme 2.3. Biosynthesis of spermidine from SAM.
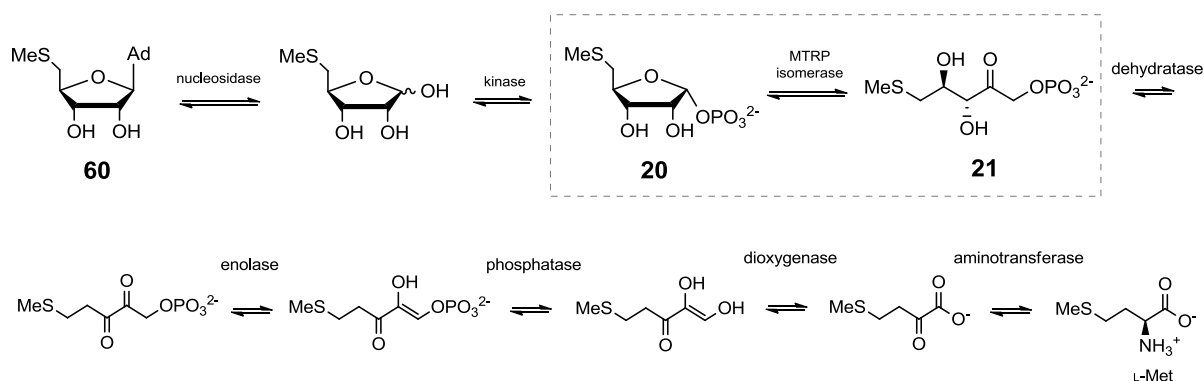
Scheme 2.4. The methionine salvage pathway found in bacteria.[10]

An isomerase (5-methylthio-α-D-ribose 1-phosphate isomerase, MTRPi) is involved as a key enzyme on this pathway. It catalyses the ring opening of **20** to **21**. This isomerase belongs to a class of proteins that has high homology to the 5-deoxy-5-fluoro-α-D-ribose 1-phosphate isomerase (5-FDRPi) of *S. cattleya*.

Two crystal structures of MTRPi's have been solved: the MTRPi of *Bacillus subtilis*[12] and YPR118W of *Saccharomyces cerevisiae* (yeast).[13] Due to their similarity with 5-FDRPi, the published data on the MTRPi's are valuable reference structures for studies on 5-FDRPi.

The protein alignment of MTRPi's from different sources illustrates that some residues of the active site are highly conserved, in particular Cys160 and Asp240 (numbered to the MTRPi of *B. subtilis*). These appear to be key catalytic residues. Both residues are also conserved in 5-FDRPi (Cys177 and Asp257).

2.2.2 The MTRPi from *Bacillus subtilis*

Several studies have been carried out on the MTRPi from the bacterium *Bacillus subtilis*. Matsumura and co-workers characterised the enzyme[14] and solved its crystal structure[12] in 2008. They observed that this enzyme did not require metals to exhibit activity.[14] Such an observation may support a *cis*-enediol mechanism. However, they also reported that deuterium was not incorporated into the ribulose product when the reaction was carried out in D_2O.[14] This was unexpected because a *cis*-enediol mechanism should promote the exchange of deuterium with the solvent. The absence of deuterium exchange raised the alternative possibility of a hydride-shift mechanism.
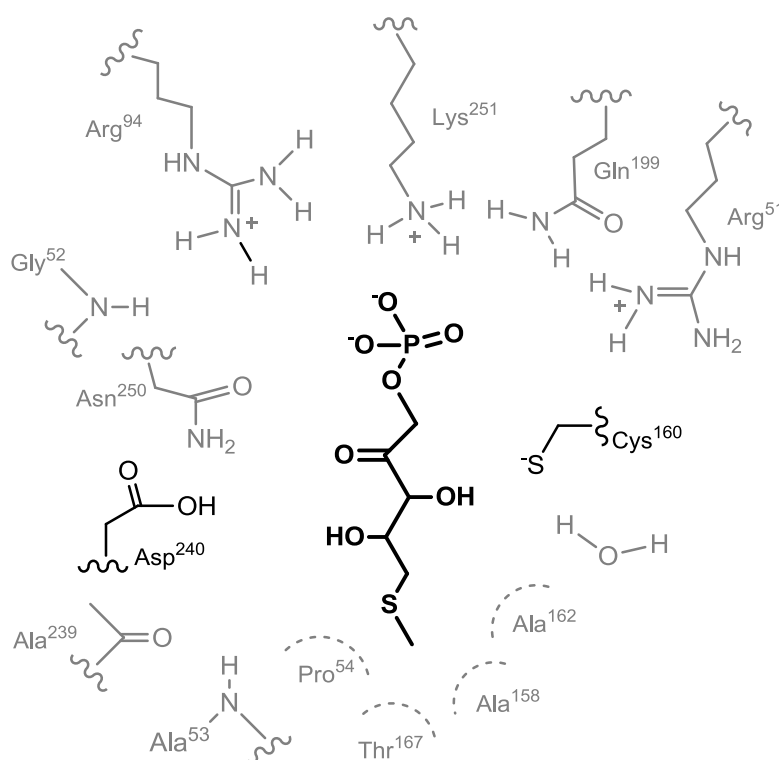

Fig 2.2. Active site residues in MTRPi from *B. subtilis*, bound to the product **13**.[12]

The 5-methylthio-D-ribulose 1-phosphate product **13** was co-crystallised with the enzyme in one of the structures (PDB 2YVK) and the product was found to be bound into the active site (Fig. 2.2). The active site residues were classified into three main categories: Those that form hydrogen-bonds or have an electrostatic interaction with the phosphate group of **13**, those which form hydrogen-bonds with the hydroxyl groups of **13** and those which interact with **13** *via* hydrophobic interactions.

However, the structure did not provide enough detail to distinguish between the two possible modes of catalysis. It did show that two highly conserved residues, Asp240 and Cys160 are situated in proximity to the ribulose carbon chain. The authors proposed that Asp240 may act as a proton donor/acceptor as it is situated within a hydrophobic pocket and this environment is likely to increase its pK_a from 4.4 to around 7. They also suggested that Cys160 might become deprotonated at the optimal pH of the enzyme (pH 8.1) and that sulfide anion may act as a base for catalysis in a *cis*-enediol mechanism or to stabilise a transient positive charge on the substrate during a hydride-shift (Scheme 2.5).

These structures have the linear open chain product bound. However a crystal structure of the enzyme with the substrate in the furanose ring-closed form in the active site is highly desirable. Such a structure would show the active site in a 'tight' configuration poised for reaction. The current product-enzyme co-complex is in a relaxed arrangement with the enzyme configured to release the product from the binding pocket.
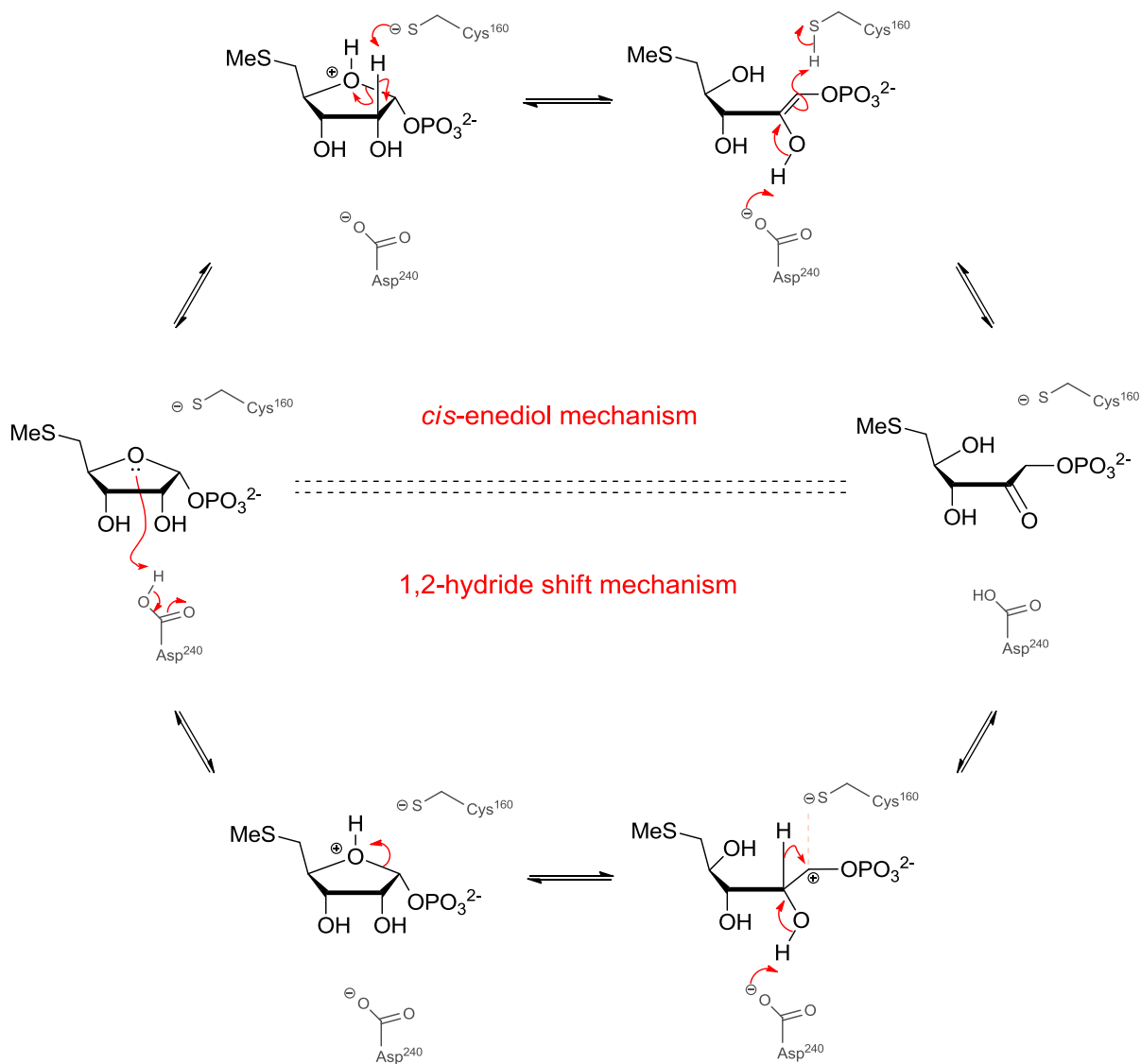
Scheme 2.5. The two postulated mechanisms of the MTRPi based on the crystal structure.

In a separate study on the methionine salvage pathway of *Geobacillus kaustophilus,* Imker and co-workers have shown that the MTRPi catalyses a regiospecific and stereospecific hydride transfer.[15] This was revealed using selective deuterium-labelled D-ribose 1-phosphate as a substrate. The reaction was followed by ^{1}H-NMR (Scheme 2.6). The study demonstrated that the C-1 and C-2 hydrogen atoms (H_A and H_B) of ribose 1-phosphate became the pro-*S* and pro-*R* hydrogens at C-1 of the resultant ribulose 1-phosphate after rearrangement. This indicated that the C-2 hydrogen had undergone a hydride-shift. Their findings are discussed in more detail in *Chapter 3* of this thesis.
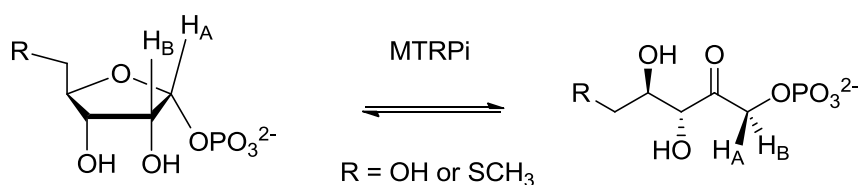
**Scheme 2.6**. Stereochemical course of hydrogen transfer by MTRPi.[15]

This experiment by Imker *et al.* provided an important insight supporting the hydride-shift mechanism for ribulose 1-phosphate isomerases. However there are aspects of this enzyme that merit continued investigation such as the role of the catalytic residues and to establish details regarding the mechanism of the process.

2.2.3 5-Deoxy-5-fluoro-D-ribose 1-phosphate isomerase
(5-FDRPi) from *S. cattleya* : Identification of the enzyme

As reported in *Chapter 1,* the activity of an isomerase similar to that of the 5-methylthioribose 1-phosphate isomerase (MTRPi) was previously explored in *S. cattleya*. Initially, the protein sequences of the two known MTRPi's from *B. subtilis* and *S. cerevisiae* respectively were used to probe the genome of *Streptomyces coelicolor,* for which a full genome sequence is available. From this search, gene-*SCO3014* of 1124 bp with high homology to the two MTRPi's was identified. It was mis-annotated as a translation initiation factor. This protein from *S. coelicolor* has 33 % and 38 % identity to the MTRPi's from *B. subtilis* and *S. cerevisiae* respectively. It also contains the putative catalytic residues: Cys160 and Asp240, which are conserved.

A set of PCR primers was then designed to clone the *SCO3014* gene and the protein was over-expressed in *E. coli*. The expressed protein was found to be catalytically active in the conversion of 5-FDRP to 5-FDRulP.[16] Thus it would appear that the methionine salvage pathway isomerase of *S. coelicolor* was able to act on the fluorinated substrate (Scheme 2.6).
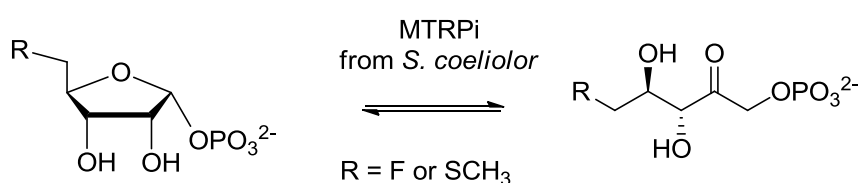
Scheme 2.6. MTRPi from *S. coelicolor* is able to catalyse the isomerisation of 5-FDRP.

With this success, the peptide sequence of *SCO3014* was aligned against the genome of *Streptomyces avermilitis,* a closely related strain to *S. cattleya*. This identified a homologous protein with 45 % identity. Alignment of the two sequences was carried out to locate highly conserved regions, from which degenerate DNA primers were

designed as a strategy to try to amplify a homologous gene from the genomic DNA of *S. cattleya* by PCR. This gave a 288 bp product with 90 % homology to *SCO3014*, and using a chromosome walking technique, the complete open-reading frame (ORF) of this fragment, 1161 bp in size was identified which had a 75 % identity to *SCO3014*.[16] This gene was cloned into an over-expression vector and the corresponding protein was then over-expressed in *E. coli*. This enzyme was also able to catalyse the isomerisation reactions on 5-deoxy-5-fluoro-D-ribose 1-phosphate.[16]

Subsequently, another homologous isomerase was identified from *S. cattleya* by alignment of both *SCO3014* and this new *S. cattleya* isomerase sequence, with a draft genome sequence of *S. cattleya*. This finding was confirmed when the full genome sequence was completed in 2011.[16, 17] There are only two isomerase genes in the full genome. The latter isomerase (*SCATT 32590*) was clustered with other genes of the methionine salvage pathway, which strongly suggests its role in primary metabolism (Table 2.1). By contrast isomerase *SCATT 20080*, the enzyme obtained from the cloning experiments described above, was not associated with any primary metabolic pathways (Table 2.2) and it is assumed to be involved in fluorometabolite biosynthesis.

A mutant of *S. cattleya* with a gene knock-out of the isomerase lost the ability to produce fluoroacetate and 4-fluorothreonine, but otherwise grew as a healthy organism.[16] Although a negative result, this experiment supported the product of *SCATT 20080* to be the 5-deoxy-5-fluoro-D-ribose 1-phosphate isomerase (5-FDRPi) dedicated to the secondary metabolic pathway of the fluorometabolites.

Table 2.1. Genes (± 5 ORFs) around the methionine salvage pathway isomerase of *S. cattleya* (*SCATT 32590*, located on the mega-plasmid)

Gene / direction	Location from	Location to	Annotation
SCATT 32540 →	3427869	3428353	hypothetical protein
SCATT 32550 ←	3428542	3428892	hypothetical protein
SCATT 32560 ←	3428912	3429742	methylthioadenosine phosphorylase
SCATT 32570 ←	3429739	3430479	2,3-diketo-5-methylthio-1-phosphopentane phosphatase
SCATT 32580 ←	3430476	3431054	putative oxidoreductase
SCATT 32590 ←	3431095	3432738	putative initiation factor eIF-2B alpha subunit-like protein (MTRP isomerase)
SCATT 32600 ←	3432799	3433515	hypothetical protein
SCATT 32610 ←	3433588	3435066	crotonyl-CoA reductase
SCATT 32620 ←	3435063	3436436	cytochrome P450 protein
SCATT 32630 →	3436761	3443915	non-ribosomal peptide synthetase
SCATT 32640 →	3443912	3445117	FAD-dependent monooxygenase

Table 2.2. Genes (± 5 ORFs) around the 5-FDRP isomerase of *S. cattleya* (*SCATT 20080*, located on the main chromosome)

Gene / direction	Location from	Location to	Annotation
SCATT 20030 ←	2130607	2131287	hypothetical protein
SCATT 20040 ←	2131894	2132616	hypothetical protein
SCATT 20050 ←	2132731	2134581	lipoprotein LpqB
SCATT 20060 ←	2134626	2136842	two-component system histidine kinase
SCATT 20070 ←	2136961	2317662	putative response regulator
SCATT 20080 ←	2137912	2138973	5-deoxy-5-fluororibose 1-phosphate isomerase
SCATT 20090 →	2139641	2140627	integral membrane protein
SCATT 20100 ←	2146570	2147184	hypothetical protein
SCATT 20110 ←	2147326	2148783	*S*-adenosy-L-homocysteine hydrolase
SCATT 20120 →	2149196	2151277	PTS system, fructose-specific IIABC components
SCATT 20130 ←	2151898	2152863	transport protein

2.2.4 A phosphonate analogue of the 5-FDRPi substrate

As part of the research probing the mechanism of the 5-FDRPi, the non-hydrolysable analogue **61** of 5-FDRP **15** was designed by replacing the bridging oxygen of the phosphate ester with a methylene group.[19] This modification changes the functionality at C-1 of the furanose ring from a hemiacetal phosphate to an ether phosphonate and **61** was designed to be an inert substrate analogue where the molecule should bind to the enzyme's active site, but not be turned over (Fig. 2.3).
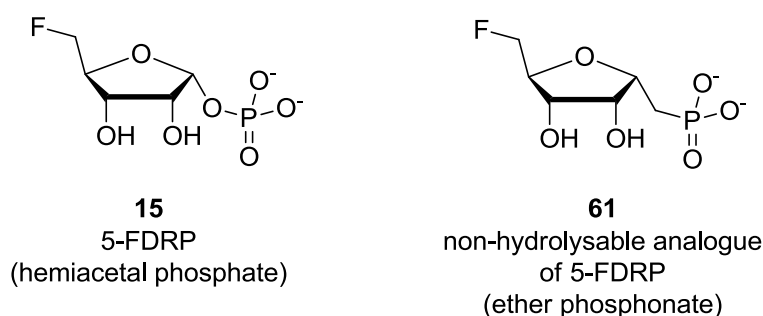

15
5-FDRP
(hemiacetal phosphate)

61
non-hydrolysable analogue
of 5-FDRP
(ether phosphonate)

Fig 2.3. The phosphate substrate and phosphonate analogue of 5-FDRPi.

It was an objective to carry out a crystallisation of the isomerase enzyme, co-crystallised with phosphonate analogue **61** in order to observe the substrate-bound structure of the enzyme, and to understand how the active site residues interact with the ribose ring. This unreactive analogue should be valuable in securing a co-crystal in the substrate-bound state. Such a structure would complement the crystal structure of the *B. subtilis* MTRPi which had a ring-opened product-bound structure.

The synthesis route to phosphonate **61**, carried out by Nasomjai[19, 20] began with D-ribose, which was suitably protected to leave a free OH group at the anomeric position. Then the phosphonate moiety was introduced through a Wadsworth-Emmons reaction. Removal of the acetal protecting group on the ribose and hydrolysis of the phosphonate esters gave **61** in the form of a white crystalline cyclohexylammonium salt (Scheme 2.7).
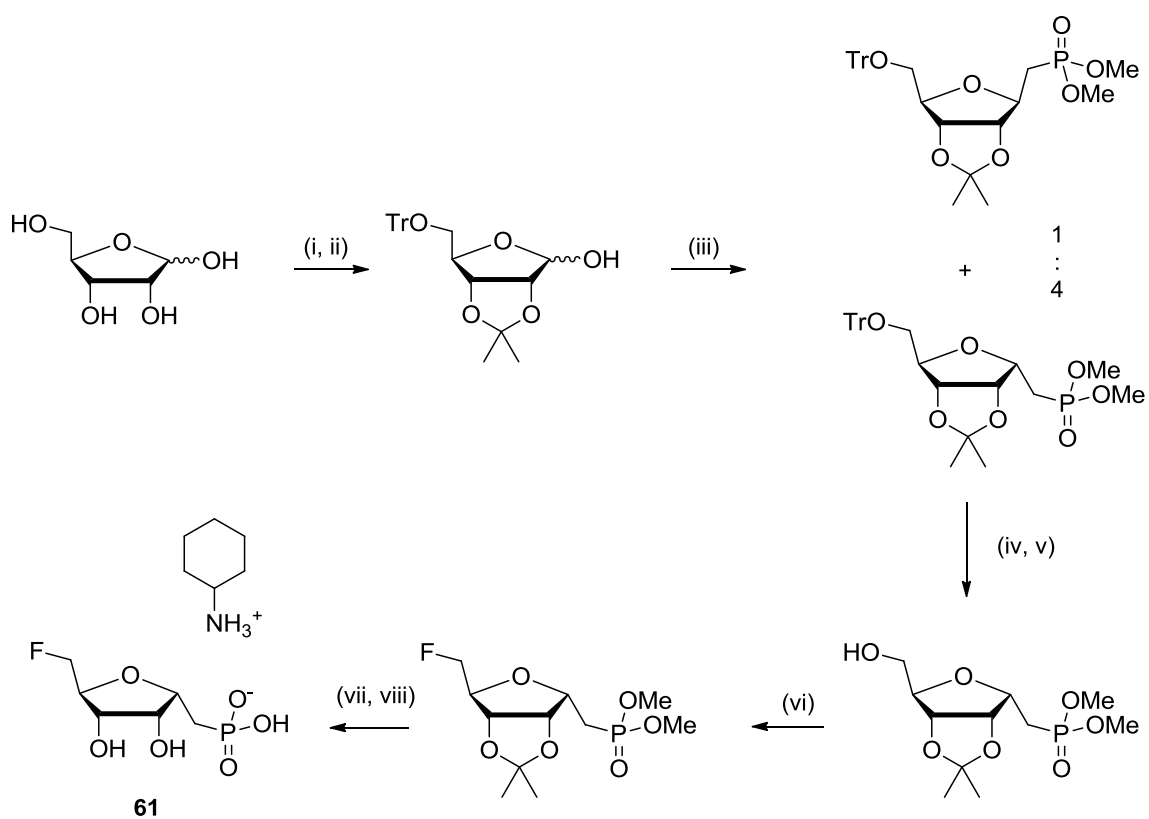
i. acetone, conc. H_2SO_4 (cat.)   ii. TrCl, Pyr, DMAP   iii. $(MeO)_2P\overset{O}{\parallel}\diagdown P(OMe)_2\overset{O}{\parallel}$, NaOH, H_2O/DCM
iv. chromatographic seperation   v. $ZnCl_2$, DCM   vi. TsF, TBAF, THF   vii. TMSBr, DCM
viii. TFA/H_2O (1:1), then $CyNH_2$

Scheme 2.7. The synthesis of 5-FDRP analogue **61** by Nasomjai. [20]

2.2.5 The aim of the project

This aspect of the project aimed to carry out crystallisation trials of the 5-FDRPi of *S. cattleya,* where the phosphonate analogue would be co-crystallised with the protein. This would involve designing a construct for the 5-FDRPi gene for its over-expression and to devise an efficient method for isolation and purification of the protein. Thereafter crystallisation trials could take place to produce protein crystals for X-ray diffraction.

In addition, the project also aimed to explore further the catalytic mechanism by mutation of active site residues, Cys177 and Asp257. Ring opening was studied using a deuterium labelled substrate, in order to probe the hydride-shift mechanism hypothesis.

2.3   Cloning and over-expression of the native 5-FDRPi

2.3.1 Cloning of the native 5-FDRPi using *Nco*I-*Xho*I cloning sites

Genomic DNA of *S. cattleya* was prepared from the mycelia of *S. cattleya* obtained from a 36 h culture grown in Tryptone Soy Broth-Yeast extract (TSBY) media. The mycelia were suspended in TE25S buffer (5 ml) and digested with lysozyme (2 mg/ml) at 37 °C for 1 h. Then proteinase K (50 μl) and 10 % sodium dodecylsulfate solution (300 μl) were added and the sample incubated at 55 °C for 1 h. After this incubation, 5 M NaCl (1 ml) was added, followed by phenol/chloroform (1:1 v/v, 5 ml) and the slurry was mixed by inversion for 30 mins and then centrifuged. The aqueous layer was transferred to a clean tube and the DNA was precipitated by addition of 3 M sodium acetate (0.6 ml) and isopropanol (6 ml). The DNA pellet was obtained by centrifugation, and was washed with 80 % ethanol and then redissolved in DNase-free water (1.5 ml).

The genomic DNA solution was used as the template for a PCR reaction. A set of PCR primers (Table 2.3) used previously in the lab were used to clone the gene using the restriction sites *Nco* I and *Xho* I.[21]

Table 2.3. PCR primers for isomerase cloning.

Primer	Restriction site	Primer sequence (5' → 3')
forward	*Nco* I 5'CCATGG3'	GCAGGAGGAATTCCATGGGTGATCAGTCCGTACAGCCTTTGGC NcoI 5-FDRPi
reverse	*Xho* I 5'CTCGAG3'	CGCCGCTCGAGCGGAAGGGACGGTCGTCATCGGTGAC XhoI downstream sequence of 5-FDRPi

The PCR reaction mixture contained KOD polymerase (1 μl), magnesium chloride (0.4 mM) and DMSO (5 % final concentration). The thermal cycles were carried out in a PCR machine (Techne) with 40 cycles for the sequence of 95 °C for 1 min, 56 °C for 1 min and 68 °C for 2 mins.

The 5-FDRPi is a protein with 386 amino acid residues and a gene of 1161 bp. As with other Streptomycetes DNA the sequence is GC-rich (77 %), and the melting point for the gene is thus slightly higher than average. In order to reduce the melting temperature for the PCR cycles and improve the yield, the PCR reactions were carried out with 5 % DMSO in the mixture. KOD polymerase was chosen for the PCR reactions for its superior fidelity and efficiency.

A PCR product of the expected size of *ca.* 1.2 kb was obtained and purified by running in an agarose gel electrophoresis and excising the band of correct mass (Fig 2.4).
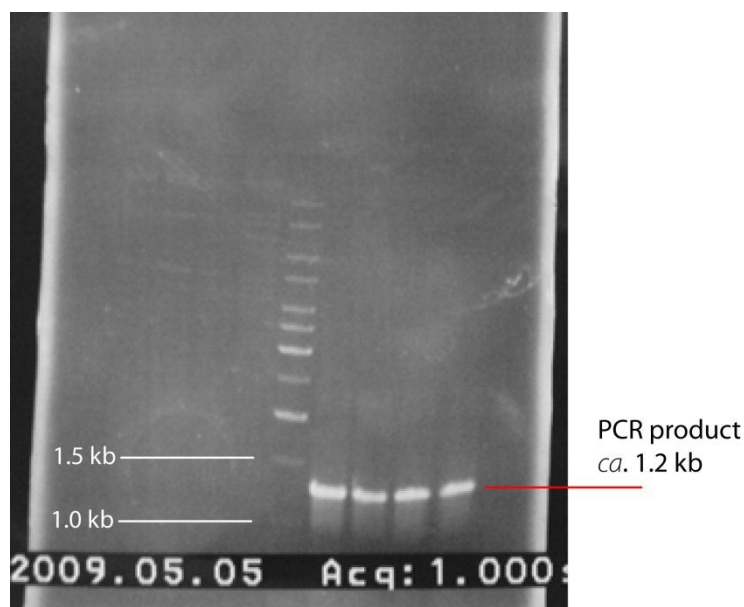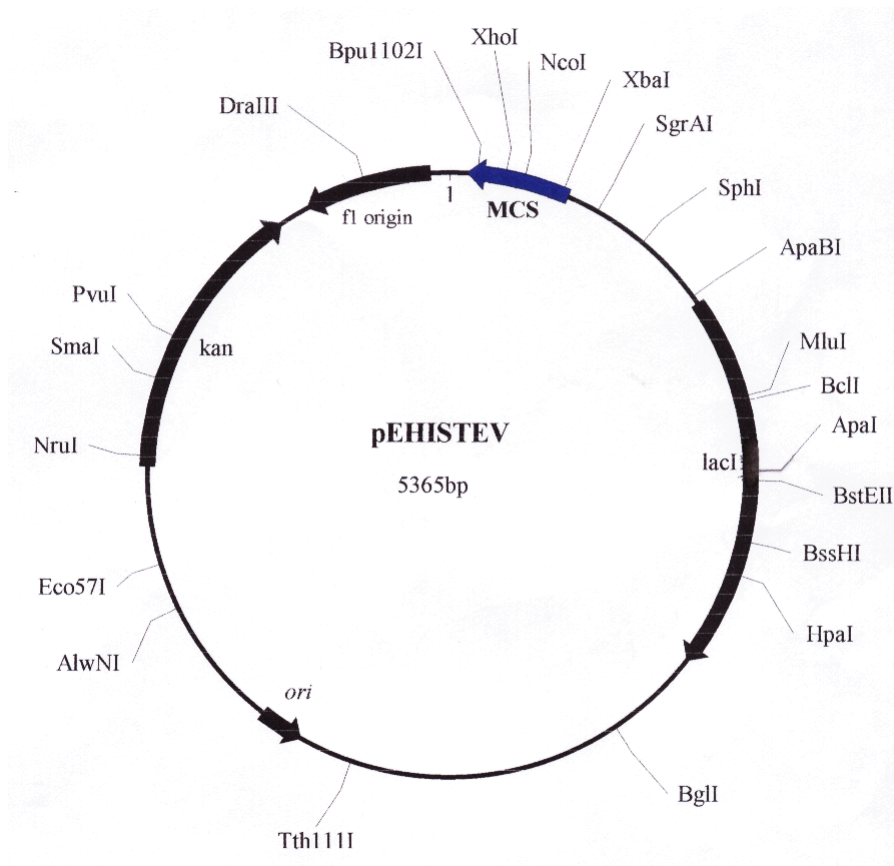

Fig 2.4. UV photograph of the 0.8 % agarose gel containing the 5-FDRPi PCR product from *S. cattleya* genomic DNA using the primers from Table 2.3.

The direct ligation of this fragment into the pEHISTEV vector has proven difficult and therefore the PCR products were cloned into a pGEM-T-vector (Promega). The PCR products were treated with *Taq* polymerase and dNTPs at 72 °C for 10 mins for the addition of a poly-A tail to the blunt end PCR products (as produced by KOD polymerase) prior to cloning into the pGEM-T-vector. The recombinant 5-FDRPi-T-vector plasmid was transformed into *E. coli* DH10B competent cells, and the transformed cells were allowed to develop on an agar plate containing ampicillin.

Colonies were picked and their plasmids were extracted and analysed by agarose gel electrophoresis. Those with the correct plasmid size were sequenced to confirm the accuracy of the inserted gene product. The correct T-vector-isomerase recombinant plasmid and the empty pEHISTEV vector were both subjected to a double digestion with the restriction endonucleases *Nco* I and *Xho* I, and the resultant linear DNA pieces were ligated by DNA T4 ligase to give the pEHISTEV-isomerase construct. This construct was transformed into *E. coli* BL21 (DE 3) Gold cells and plated onto an agar plate containing kanamycin.

The pEHISTEV vector for protein expression was chosen, as it has been used for many successful protein over-expressions (developed by Dr H. Lui, University of St Andrews).[22] This vector contains a multiple cloning site (MCS) with a preceding N-terminal His_6 affinity tag which is co-expressed with the protein using a start codon (ATG) upstream of this His_6 sequence. In between the affinity tag and the multiple cloning site is a sequence encoding a TEV protease cleavage site to allow removal of the His-tag from the product protein. Upstream of the start codon is a *lac*-operon to allow over-expression of the inserted gene. pEHISTEV also carries a kanamycin resistance gene and an origin of replication (ori) (Fig. 2.5).

Bpu1102I
XhoI
NcoI
XbaI
DraIII
SgrAI
SphI
1 MCS
f1 origin
ApaBI
PvuI
MluI
SmaI
kan
BclI
ApaI
NruI
lacI
BstEII
pEHISTEV
5365bp
BssHI
Eco57I
HpaI
AlwNI
ori
BglI
Tth111I

Multiple Cloning Sites of pEHISTEV

T7 Pro seq primer
*Xba*I
ATCGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCTCTAGAAATAAT
T7 promoter
lac operator

rbs   *Nde*I   6xHis
TTTGTTTAACTTTAAGAAGGAGATATACAT ATG TCG TAC TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT
                               M   S   Y   Y   H   H   H   H   H   H   D   Y   D

TEV protease site
GAT TAC GAC ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGC
D   Y   D   I   P   T   T   E   N   L   Y   F   Q   G

*Nco*I   *Eco*RV   *Bam*HI   *Eco*RI   *Sac*I   *Sal*I   *Hind*III   *Not*I   *Xho*I
GCC ATG GCT GAT ATC GGA TCC GAA TTC GAG CTC CGT CGA CAA GCT TGC GGC CGC ACT CGA GCA
A   M   A   D   I   G   S   E   F   E   L   R   R   Q   A   C   G   R   T   R   A

CCA CCA CCA CCA CCA CTG AGA TCC GGC TGC TAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCT
P   P   P   P   P   L   R   S   G   C   end
                                                                              T7 Ter

*Bpu*1102I
GAGCAATAACTAGCATAACCCCTTGGGGCC TCTAAACGGGTCTTGAGGGGTTTTTTGCTGAA
           T7 terminator
seq primer

Fig 2.5. The vector map for pEHISTEV. Image taken from [22].

2.3.2 The over-expression and purification of the native 5-FDRPi (*Nco*I – *Xho*I)

E. coli BL21 Gold (DE3) cells containing the 5-FDRPi-pEHISTEV recombinant plasmid were grown in LB media at 37 °C. Cells were grown to an OD_{600} of 0.5-0.6 and induced with isopropyl-β-D-1-thiogalactoside (IPTG) (0.5 mM final concentration). IPTG is a synthetic non-hydrolysable inducer of the *lac*-operon.[23] After induction with IPTG, cells were grown for a further 18 h at 37 °C to achieve maximum protein production. The cells were harvested by centrifugation (5000 rpm, 5180 *g*) and the cell pellet was resuspended in lysis buffer (50 ml) and subjected to sonication (Vibra Cell apparatus) at 0 °C for 10 mins at 1 min intervals with 1 min pause in between. The cell lysate was centrifuged (20,000 rpm, 48,384 *g*) to remove cell debris, and the soluble fraction was retained.

The soluble fraction was then passed through a column of Ni-NTA beads (3 ml) and the flow through was discarded. The loaded column was washed with wash buffer (100 ml), and then the protein was eluted by passing an elution buffer (10 ml) through the column. The protein solution was dialysed against 20 mM phosphate buffer (pH 7.8, 7.5 L) overnight at 4 °C and then concentrated by ultrafiltration to a volume of 2 ml (Amicon concentrators, 10 kDa MWCO), and the sample was applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated with gel filtration buffer. The isomerase containing fractions were pooled and concentrated to 15 mg/ml for crystallisation trials. The yield of the protein was 10-20 mg/L. The identity of the protein was confirmed by MS on the trypsin digest of the protein gel band (Fig. 2.6-2.8).
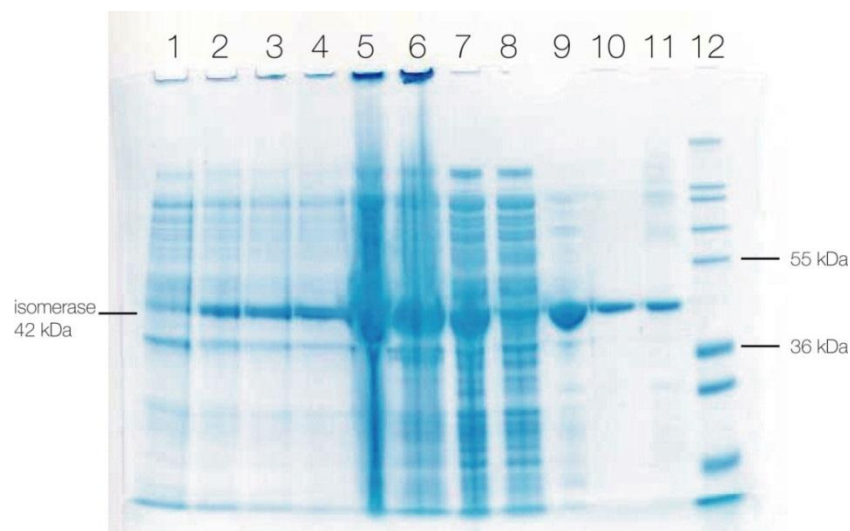
Fig 2.6. SDS-PAGE gel of samples taken from an isomerase over-expression and purification.
1. pre-induction **2.** induction after 2 h **3.** induction after 5 h **4.** induction after 24 h
5. cell lysate **6.** insoluble fraction **7.** soluble fraction **8.** flow through **9.** wash
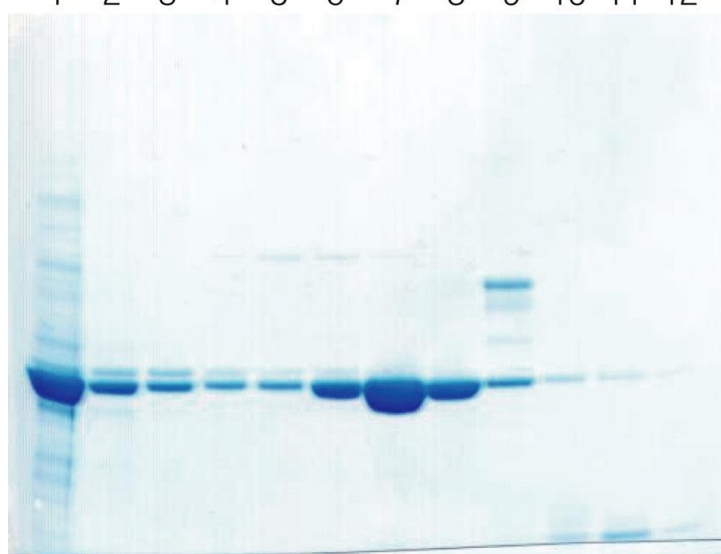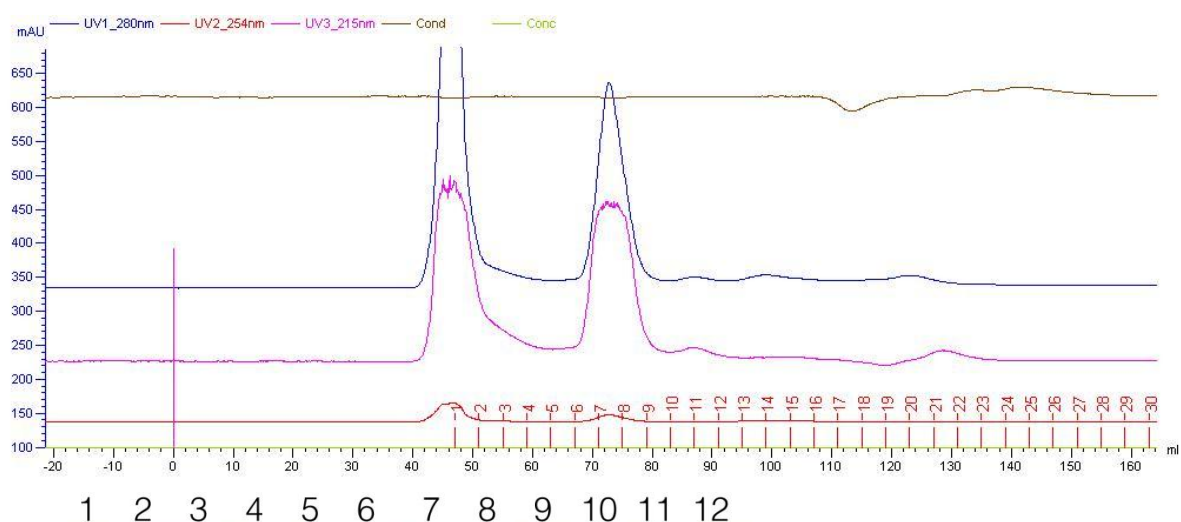**10, 11.** elutions **12.** protein unstained marker (Fermentas)

Fig 2.7. (*Above*) UV trace of the isomerase gel filtration purification.

Fig 2.8. (*Left*) SDS-PAGE gel image of fractions obtained in the gel filtration purification (Superdex 200).
1 to **12** corresponds to fractions 1 to 12 from the UV trace above indicated by red graduations.

2.3.3 Assay of 5-FDRPi

The activity of 5-FDRPi was assayed by ^{19}F-NMR as the chemical shifts of the fluorinated intermediates have been established.[16,24] The substrate of the isomerase reaction, 5-FDRP **15** was not readily available, therefore the assay was coupled with a purine nucleoside phosphorylase (PNP) reaction and synthetic 5'-FDA **14** was used as the substrate in a two-enzyme reaction (Scheme 2.8).
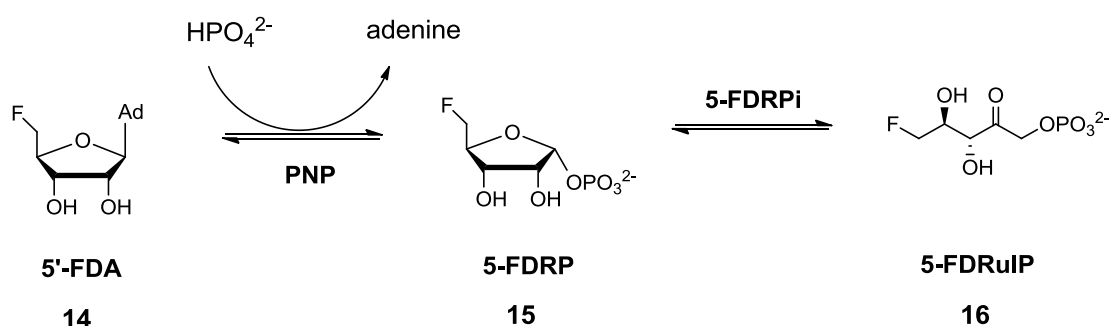
Scheme 2.8. The isomerase assay in a two-enzyme reaction by coupling to a PNP reaction.

An efficient chemical synthesis of 5'-FDA was developed by Ashton *et al.*[25] and their reported procedures were followed to prepare this compound. The PNP enzyme (FlB) from *S. cattleya* was over-expressed as a partially soluble fusion protein with a maltose-binding protein unit (FlB-MBP). This over-expression system was developed previously in St Andrews.[26]

When 5'-FDA was incubated with PNP-MBP in a phosphate buffer (pH 7.8, 20 mM) at 37 °C for 5 h, 5-FDRP was observed by ^{19}F NMR in an equilibrium mixture with 5'-FDA. The reaction stops at around 40 % conversion. When 5'-FDA was incubated with both PNP-MBP and over-expressed 5-FDRPi, three signals are observed in the ^{19}F\{^{1}H\} NMR as shown in Fig. 2.9: δ_F -231.35 ppm (5-FDRP), δ_F -231.54 ppm (5'-

FDA) and δ_F -231.82 ppm (5-FDRulP). The isomerase is a more efficient enzyme than the PNP in the assay and the conversion to 5-FDRulP from 5'-FDA is around 70 %.
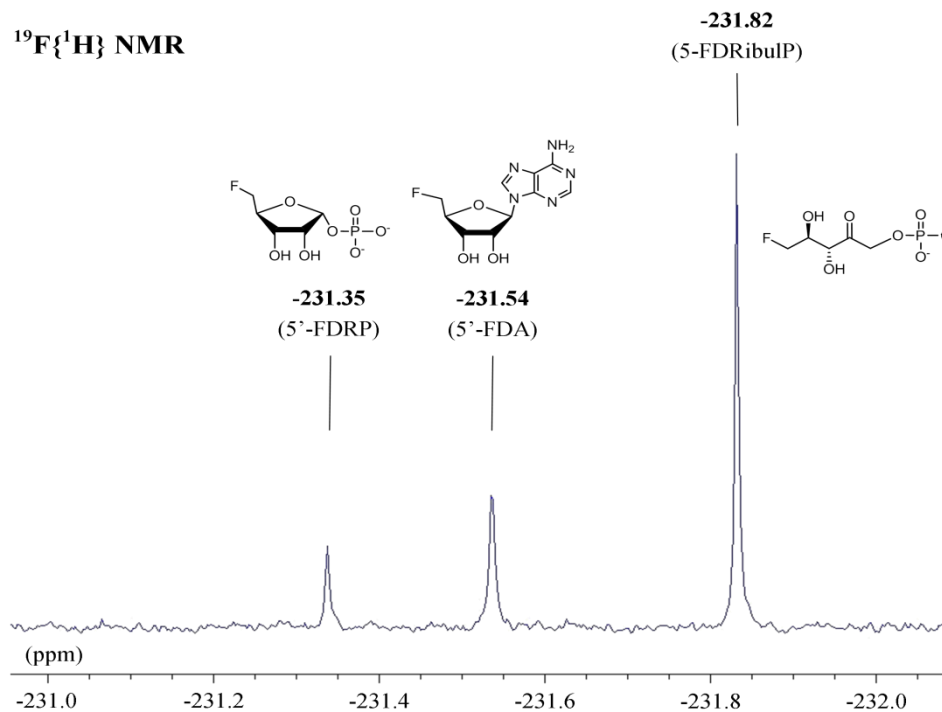

Fig 2.9. $^{19}F\{^1H\}$ NMR of an isomerase assay analysis carried out, with assignment of the fluorine shifts based on the reported chemical shifts in [24].

2.3.4 Removal of the His-tag on the native 5-FDRPi

The over-expressed 5-FDRPi has a His_6-tag extending from the N-terminus. This is removed to avoid a disordered loop which will hinder protein crystallisation.

As described earlier, a Tobacco Etch Virus (TEV) protease cleavage site[29] ENLYFQG is cloned onto the protein when using the pEHISTEV vector. The protein purification is also further streamlined by engineering the TEV protease to contain a His_6-tag (but lacking the TEV cleavage site). After proteolysis, the product is passed through a nickel affinity column. The liberated His_6-tag, any un-cleaved protein and the TEV protease enzyme are retained on the column. This allows the free isomerase to elute without these impurities.

When the reaction was carried out under the typical procedure with an 5-FDRPi to TEV ratio of 1 : 50 by mass for a 2 h digestion at 20 °C, there was almost no cleavage of the His_6 tag from the protein. Therefore the TEV reaction was optimised. It was found that a 5-FDRPi to TEV ratio of 1 : 5 and an incubation time of 4 days at 20 °C was required to achieve >95 % His-tag removal (Fig. 2.10). 5-FDRPi was stable under these conditions when analysed by SDS-PAGE. The digestion progress could be observed on the gel where the cleaved/uncleaved protein differ by 3.4 kDa in mass (26 residues).
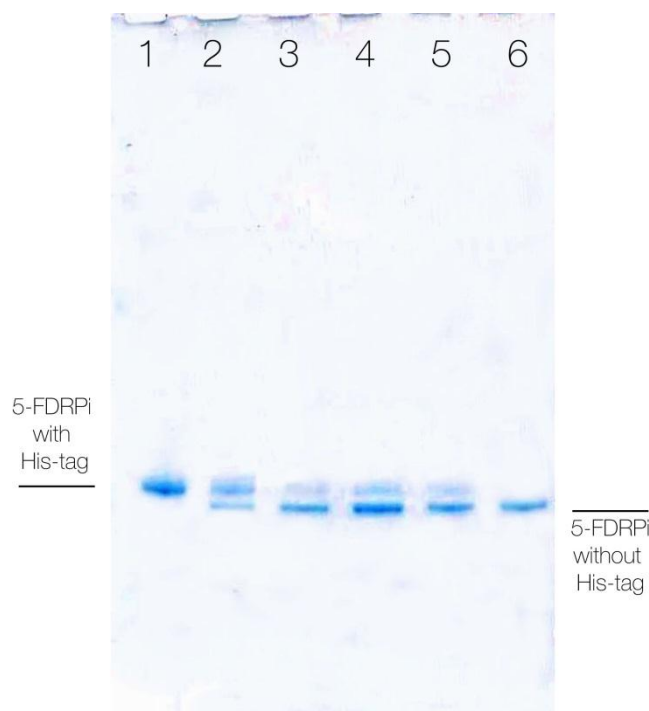
Fig 2.10. SDS-PAGE gel image of TEV protease digestion of His-tagged isomerase (*Nco*I – *Xho*I). Reaction time: **1.** 0 h   **2.** 2 h   **3.** 1 day   **4.** 2 days   **5.** 3 days   **6.** 4 days (a protein ladder was not loaded in this case)

The inefficient tag-removal was undesirable for a scale up preparation since large amounts of TEV protease enzyme were required. The present construct, employing the restriction sites *Nco* I at the N-terminus, gave a three-amino acid length spacer between the TEV cleavage site and the isomerase. It was anticipated that increasing the length of the spacer could improve the efficiency of tag-removal by making room for the TEV protease to access the hydrolysis site more easily. Therefore the restriction site was changed to *BamH* I which provided a linker of eight amino acids (chain length of 24 atoms) to put distance between the isomerase and the TEV site.

Nco I site :   MSYYHHHHHHDYDDYDIPTT**ENLYFQ**↓**GAM** – 5FDRPi

BamH I site : MSYYHHHHHHDYDDYDIPTT**ENLYFQ**↓**GAMADIGS** – 5FDRPi

Fig. 2.11. N-terminus of 5-FDRPi when expressed with the *Nco* I and *BamH* I restriction sites (pEHISTEV vector). Underline indicates TEV recognition sequence and ↓ denotes cleavage site.

2.3.5 The cloning and over-expression of the native 5-FDRPi (*BamH*I-*Xho*I)

A new set of PCR primers was designed to clone the new *BamH* I - *Xho* I restriction sites isomerase (Table 2.4).

Table 2.4. PCR primers for isomerase cloning.

Primer	Restriction site	Primer sequence (5' → 3')
forward	*BamH*I 5'GGATCC3'	GTCT<u>GGATCC</u>ATGGGTGATCAGTCCGTACAGCC BamHI · · · · 5-FDRPi
reverse	*Xho* I 5'CTCGAG3'	CTAGCTAA<u>CTCGAG</u>TTCACGGCTGGGCGCGGACGGG XhoI · · · · 5-FDRPi

The PCR primers were purchased (Eurogentec) and a PCR reaction was carried out using the 5-FDRPi-pEHISTEV (*Nco*I-*Xho*I) plasmid as the template. The PCR conditions were the same as those reported in Chapter 2.3.1.

The PCR product was ligated into the PCR blunt end vector (Roche) and then into pEHISTEV using the restriction sites *BamH*I and *Xho*I. The ligation product was transformed into *E. coli* BL21 Gold cells.

The cells containing the new plasmid were grown by the auto-induction method described by Studier at 20 °C for 48 h with vigorous shaking (250 rpm) for aeration. The cells were harvested by centrifugation (5000 rpm, 5180 g) and the pellet was resuspended in lysis buffer (200 ml) until homogeneous. The slurry was passed through a cell disruptor at 30 kPsi (Constant Systems) and the lysate was centrifuged (20,000 rpm, 48,384 g) to remove insoluble cell debris. The supernatant was filtered

(0.45 μm) and loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated with lysis buffer and connected to an Akta-Express Protein purification system. The column was then washed with wash buffer (approx. 150 ml) until impurities were no longer detected. The protein was eluted with elution buffer. Fractions containing 5-FDRPi were pooled and applied to a desalting column (Desalt 16/10, GE healthcare), the protein was washed with desalt buffer and the fractions containing the 5-FDRPi were again pooled.

A TEV digestion was carried out on the resultant 5-FDRPi. It was found that the new protein which carried an eight-amino acid spacer was efficiently cleaved (>95 %) from the His$_6$-tag using a protein to TEV ratio of 1 : 50 at room temperature for 2 h (Fig 2.12).
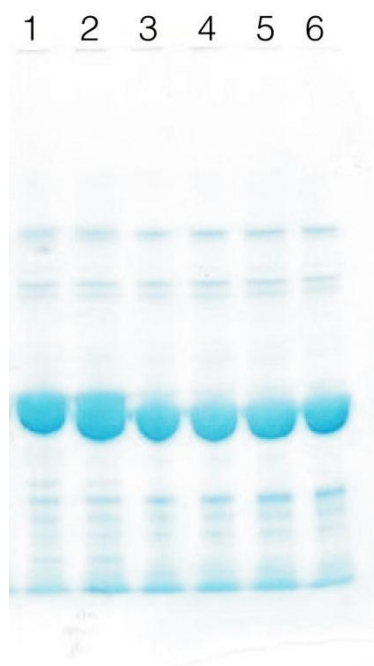

Fig 2.12. SDS-PAGE gel image of TEV protease digestion of His-tagged isomerase (*BamH*I – *Xho*I). Reaction time: **1.** 0 h **2.** 15 mins **3.** 30 mins **4.** 1 h **5.** 2 h **6.** 4 h (a protein ladder was not loaded in this case)

The free 5-FDRPi was then concentrated (Vivaspin concentrators, 30 kDa MWCO) and loaded onto a Ni-Sepharose 6 FF colum which had been equilibrated with desalt buffer. The flow through from the nickel column was collected. The 5-FDRPi fractions were pooled and applied to a gel filtration column (Superdex 200) and eluted with gel filtration buffer (Fig. 2.13). The fractions containing the 5-FDRPi were pooled and concentrated for protein crystallisation (Fig. 2.14).

This optimised procedure for the preparation of 5-FDRPi for protein crystallisation was then adopted for all crystallisation experiments.
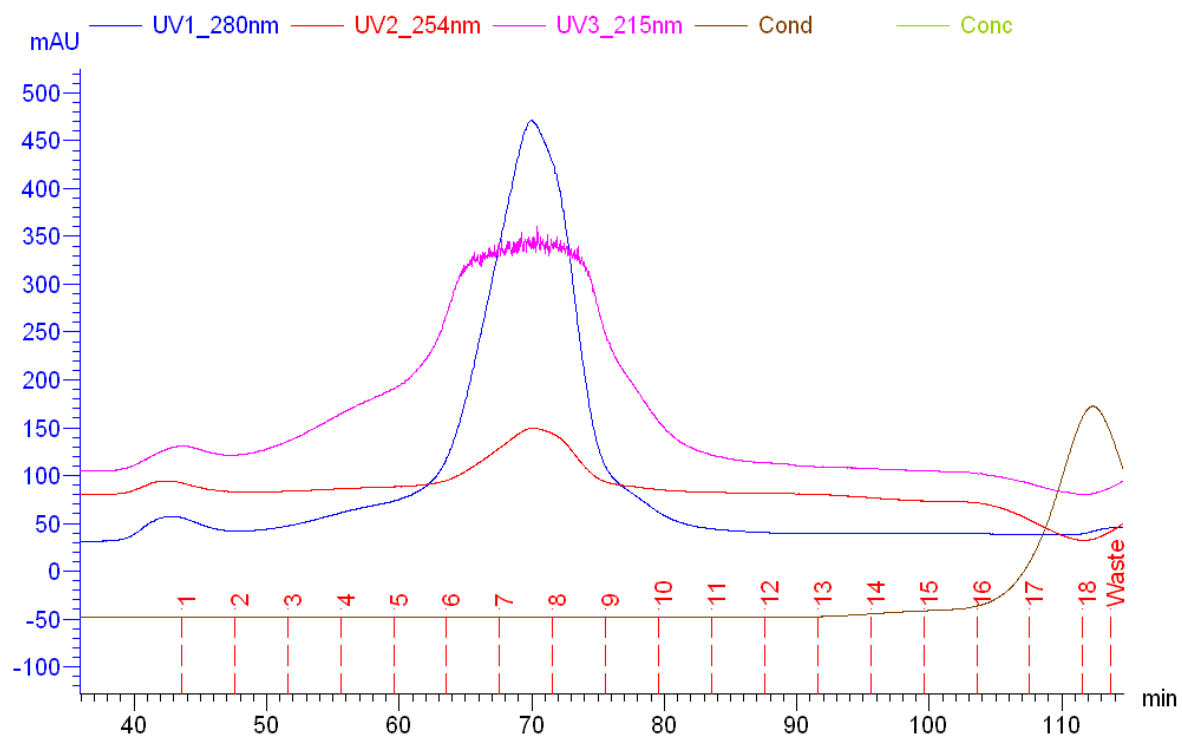
Fig 2.13. The UV-trace of a SEC purification of native isomerase. Blue= UV trace, Red = conductivity
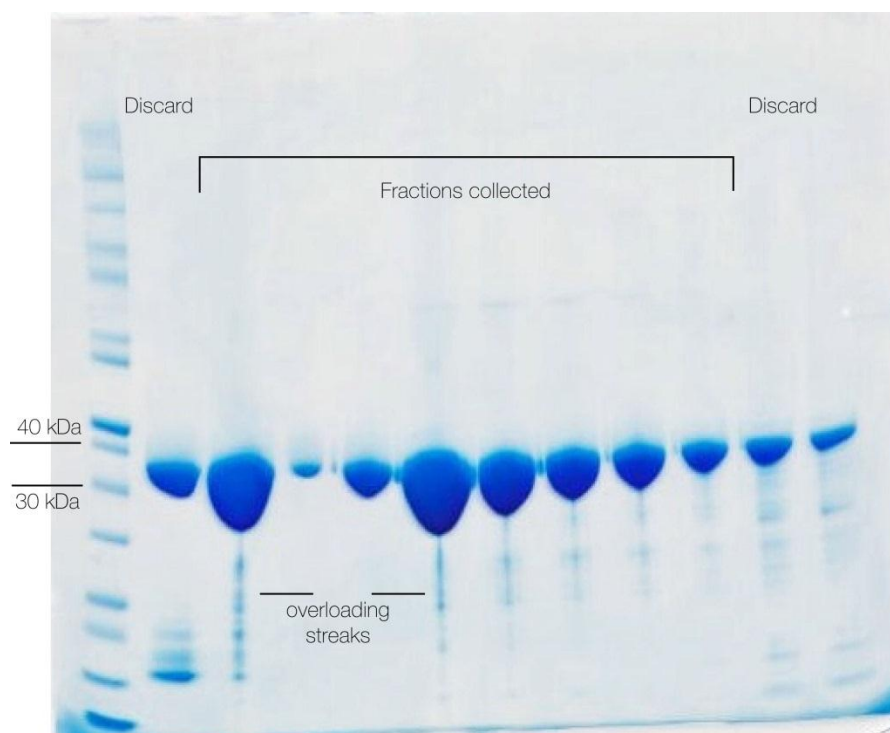

Fig 2.14. SDS-PAGE gel of the fractions from the SEC purification of the 5-FDRPi.

2.4   Preparation of non-native forms of 5-FDRPi for crystallisation

2.4.1 Lysine methylation of the native 5-FDRPi

The primary amino group of lysine is hydrophilic. A simple procedure has been developed to selectively modify the lysine residues on a protein into dimethylated lysines where the residues become hydrophobic. This reductive methylation technique has been shown in some cases to improve the crystallisation of proteins.[28, 29] The protein crystal structure of the methylated protein does not generally differ from the native protein.

The reductive methylation reaction is a one-pot procedure involving the addition of both formaldehyde and borane-dimethylamine complex to the protein solution.[30] The lysine amino groups couple to formaldehyde to form an imine, and then a borane reduction generates the methylamino group. This reacts again in a similar manner to give the dimethylated lysine (Scheme 2.8).
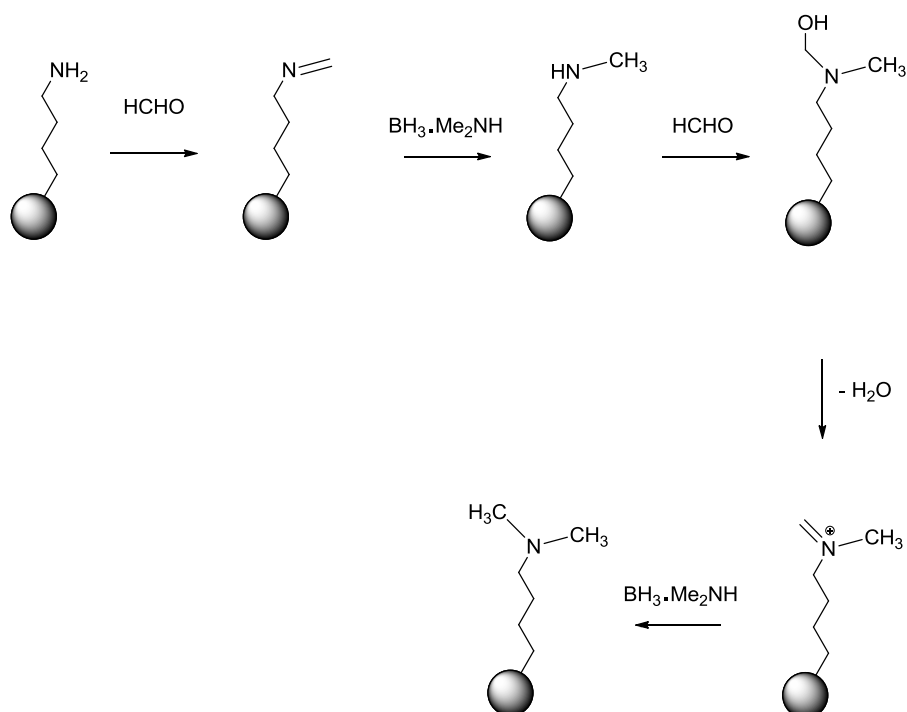
Scheme 2.8. The reaction sequence that generates dimethylated lysine residues.

5-FDRPi has three lysine residues (Lys11, Lys21 and Lys268) as well as its free N-terminus. The N-terminus is generally less reactive to this modification. Therefore exhaustive methylation would produce a protein with six or eight methyl groups. A sample of the purified native isomerase (10 mg in 1 ml phosphate buffer) was incubated with the borane-dimethylamine complex (6 mg) and formaldehyde solution (50 μl, 1M) for 24 h at 4 °C, and the reaction was quenched with the addition Tris buffer (125 μl, pH 7, 1M). The solution was dialysed against Tris buffer (5 L, pH 7.8, 20 mM) to remove the low MW reagents and the protein was then purified on a SEC column (Superdex 200). Fractions containing the eluted protein were pooled and concentrated for crystallisation (15 mg/ml). The dimethylations were confirmed by whole protein mass spectrometry, which showed an increase of 86 amu and 112 amu compared to the native protein (Fig 2.15).
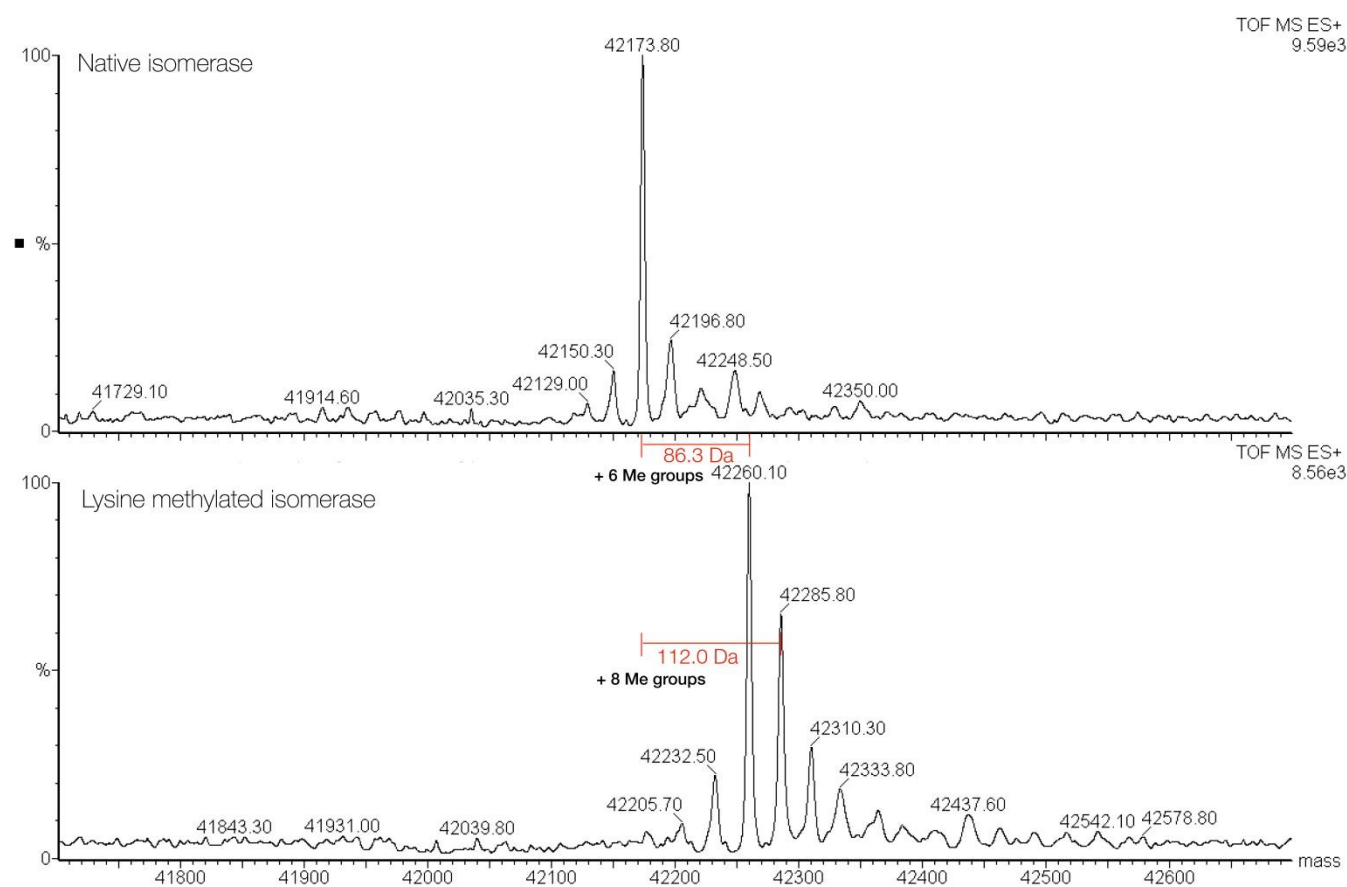

Fig 2.15. Whole protein mass spectra of native and 3 x and 4 x lysine-dimethylated 5-FDRPi (M+86, M+112).

2.4.2 Truncation mutants of the 5-FDRPi based on a secondary structure prediction

The isomerase sequence was analysed using a secondary-structure prediction tool, PROF,[31] provided by the University of Aberystwyth. This tool compares the linear sequence of amino acids to a database of secondary structures and predicts regions that are likely to fold into α-helices, β-sheets or loop structures (Fig 2.16).
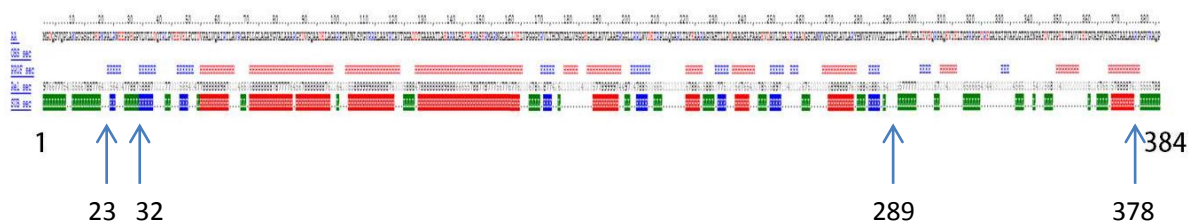

Fig 2.16. Secondary structure prediction of the 5-FDRPi with PROF.[31] Arrows indicate truncation loci chosen. **Red** = α-helix, **Blue** = β-sheet, **Green** = loop, **White** = no prediction available.

From this analysis, probable disordered loop regions were identified at both the N and C termini, whereas the middle region appeared to form defined α and β structures that are linked by short loops.

Truncation of the isomerase that removed the putative disordered regions at the ends were anticipated to have a better chance of crystallisation. The truncated versions were designed based on the structure prediction. At the N-terminus, two dissecting loci were chosen (Ala23 and Gly32). At the C-terminus, two loci were also chosen (Ala289 and Arg378). A total of six mutants were prepared (Table 2.5).

Table 2.5. Truncation mutants based on secondary structure prediction for 5-FDRPi.

Number	N-terminus	C-terminus	Length (aa)	Solubility
Native	Met1	Pro386	386	Yes
1	Ala23	Pro386	364	No
2	Ala23	Arg378	356	No
3	Ala23	Ala289	267	No
4	Gly32	Pro386	355	No
5	Gly32	Arg378	347	No
6	Gly32	Ala289	258	No

PCR primers were designed for each truncation and these mutants were all cloned into the pEHISTEV vector using the restriction sites *Nco* I and *Xho* I (Table 2.6). The recombinant genes were confirmed by DNA sequencing from both ends of the gene.

Table 2.6. PCR primers for cloning of truncation mutants of the 5-FDRPi.

Primer	Restriction site	Primer sequence (5' → 3')
forward **N -A23**	*Nco* I 5'CCATGG3'	ACGGTTTGGCCATGGCTCTCCGCTGGGAAGAGCC NcoI 5-FDRPi A23
forward **N –G32**	*Nco* I 5'CCATGG3'	ACGGTTTGGCCATGGGGCCCGTGCTGGTCCTCCT NcoI 5-FDRPi G32
reverse **C- R378**	*Xho* I 5'CTCGAG3'	CTAGCTAACTCGAGTTCAGCGGGCGGCCAGGGCGGCTA XhoI stop 5-FDRPi R378
reverse **C- A289**	*Xho* I 5'CTCGAG3'	CTAGCTAACTCGAGTTCAGGCGACCACGACGAAGGGGA XhoI stop 5-FDRPi A289

The proteins all over-expressed well in *E. coli* BL21 Gold cells and an over-expressed protein band of the expected size was found on SDS-PAGE of the cell lysates. The gel bands were also verified by mass spectrometry of the trypsin digest of the excised bands. However, all of the six mutants were found in the insoluble fractions. Further experiment in altering the over-expression temperature (16 °C, 25 °C, 37 °C) and IPTG concentration (0.1 mM, 0.5 mM, 1 mM) were attempted but without success in improving protein solubility.

2.4.3 The systematic truncation mutants of the 5-FDRPi

The truncations made based on structure prediction discussed above, did not produce soluble protein. Therefore a series of experiments was designed to explore in a systematic way whether this peptide can be shortened and also remain soluble. Accordingly, five residues at a time were sequentially shortened from each end of the isomerase peptide.

The mutants were cloned with the *BamH* I and *Xho* I restriction sites, and PCR primers were designed (Table 2.7). In each case, the sequence of the recombinant plasmid was confirmed. Each was then transformed into BL21 Gold *E. coli* and the over-expression was carried out in LB media at 25 °C for 24 h, induced by the addition of IPTG to a final concentration of 0.5 mM at OD_{600} of 0.5-0.6.

The *E. coli* cells were harvested by centrifugation and an aliquot was lysed by sonication. The cell lysates and the soluble and insoluble fractions were analysed by SDS-PAGE. This indicated over-expression of the individual proteins at the expected size for all of the mutants. It emerged that the N-terminal shortened mutants were all insoluble, while the C-terminal mutants were present in the soluble fraction (Table 2.8).

Table 2.7. PCR primers for cloning systematic truncation mutants of the 5-FDRPi.

Primer	Restriction site	Primer sequence (5' → 3')
forward native	*BamH* I 5'GGATCC3'	AGGATCCGGTGATCAGTCCGTACAGCC BamHI 5-FDRPi
forward N‑min5	*BamH* I 5'GGATCC3'	AGGATCCCAGCCTTTGGCCAAGGGCAC BamHI 5-FDRPi Gln7
forward N‑min10	*BamH* I 5'GGATCC3'	AGGATCCGGCACGGGGTCCGGGACCCC BamHI 5-FDRPi Gly12
reverse native	*Xho* I 5'CTCGAG3'	ACTCGAGTTCACGGCTGGGCGCGGACGGG XhoI stop 5-FDRPi
reverse C‑min5	*Xho* I 5'CTCGAG3'	ACTCGAGTTCAGGGGCCGGGGCGGGCGGCCA XhoI stop 5-FDRPi Pro381
reverse C‑min10	*Xho* I 5'CTCGAG3'	ACTCGAGTTCAGGCCAGGGCGGCTATGGAGG XhoI stop 5-FDRPi Ala376
reverse C‑min15	*Xho* I 5'CTCGAG3'	ACTCGAGTTCAGGAGGAGCCGGTGACCGGGG XhoI stop 5-FDRPi Ser371
reverse C‑min20	*Xho* I 5'CTCGAG3'	ACTCGAGTTCACGGGGAGGCCACGCCGGTCT XhoI stop 5-FDRPi Pro366
reverse C‑min25	*Xho* I 5'CTCGAG3'	ACTCGAGTTCAGGTCTCGGTGACCACGGCGG XhoI stop 5-FDRPi Thr361

Table 2.8. Systematic truncation mutants of the 5-FDRPi.

Name	N-terminus	C-terminus	Length (aa)	Solubility
N-*minus* 5	Gln7	Pro386	380	No
N-*minus* 10	Gly12	Pro386	375	No
C-*minus* 5	Gly2	Pro381	380	Yes
C-*minus* 10	Gly2	Ala376	375	Yes
C-*minus* 15	Gly2	Ser371	370	Yes
C-*minus* 20	Gly2	Pro366	365	Yes
C-*minus* 25	Gly2	Thr361	360	Yes

The systematically truncated mutants illustrate that the integrity of the N-terminus is crucial to solubility. Removal of five residues behind the start codon caused the protein to become insoluble. Clearly the N-terminus is essential for the proper folding of the protein *in vivo*. On the other hand, the lack of the C-terminus by up to 25 amino acids did not cause the protein to become completely insoluble.

Longer truncations are anticipated to reduce overall flexibility and possibly improve crystal formation. The C-*minus* 25 isomerase, which had the longest truncation among the soluble proteins, was therefore the first to be subject to crystallisation trials.

This mutant was over-expressed in *E. coli*, and purified as previously through a Ni-affinity column using an Akta protein purification system. The eluted protein was treated with TEV-protease to remove the His-tag. The resultant protein was loaded onto a SEC column (Superdex 75) for gel filtration chromatography. Unexpectedly, the protein eluted as three broad overlapping peaks, indicative of aggregate formation (Fig 2.17). SDS-PAGE analysis indicated multiple bands rather than a single band. This illustrated that the mutant displayed intrinsic instability and could not be purified

as a single homogenous protein. Even when 5mM DTT was added to the buffer, the stability of the protein did not improve. Since the protein was purified through Ni-affinity chromatography twice, it is only reasonable to conclude that the multiple bands were a result of self-degradation of this protein rather than impurities that originated from *E. coli.*
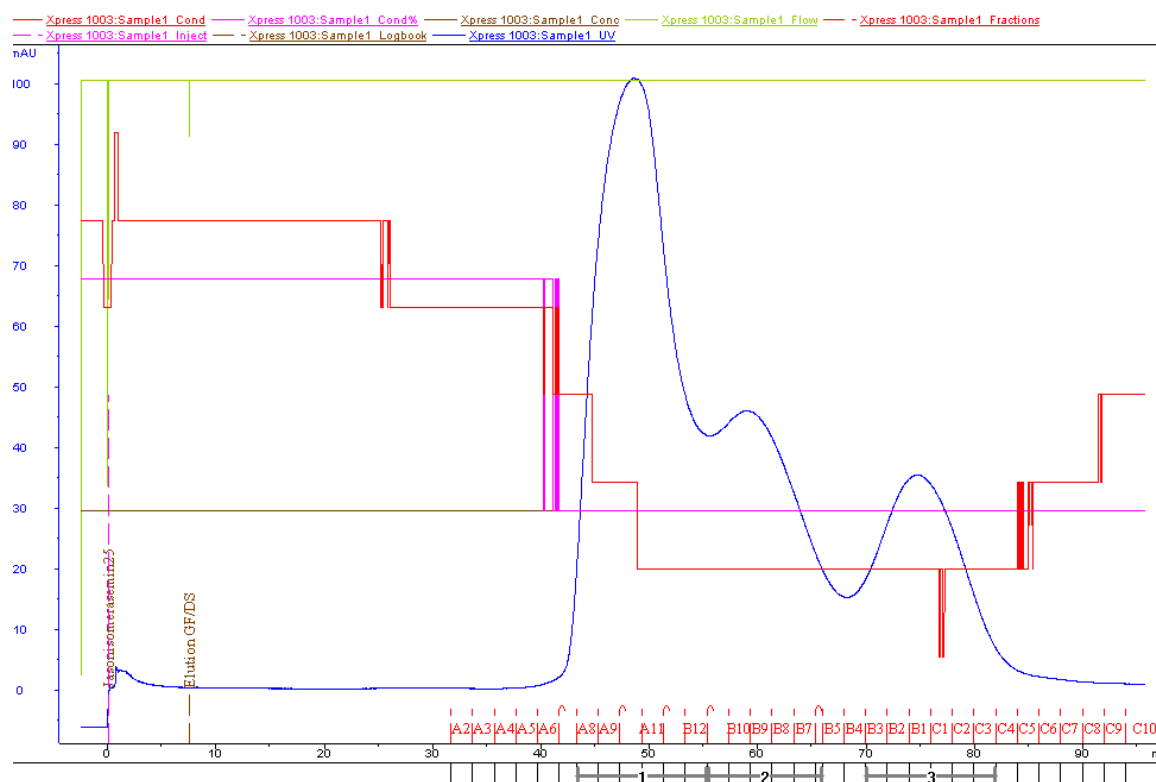

Fig 2.17. The UV-trace from the SEC (Superdex 75) of C-*minus* 25 mutant showing three broad overlapping peaks, indicative of the formation of aggregates.

Purification of the C-*minus* 15 mutant was then explored following the procedures described for the C-*minus* 25 mutant. In this case, the SEC purification (Superdex 75) resulted in two peaks. The major peak eluted first in the void volume, indicating that a protein of MW >75 kDa and suggesting a dimer or an aggregate, again even in the presence of 5 mM DTT.

When each peak was analysed by SDS-PAGE, the first appeared to consist of the C-*minus* 15 protein with some degradation products. The second contained mostly low MW peptides which may be impurities or may have arisen from self-degradation (Fig 2.18).
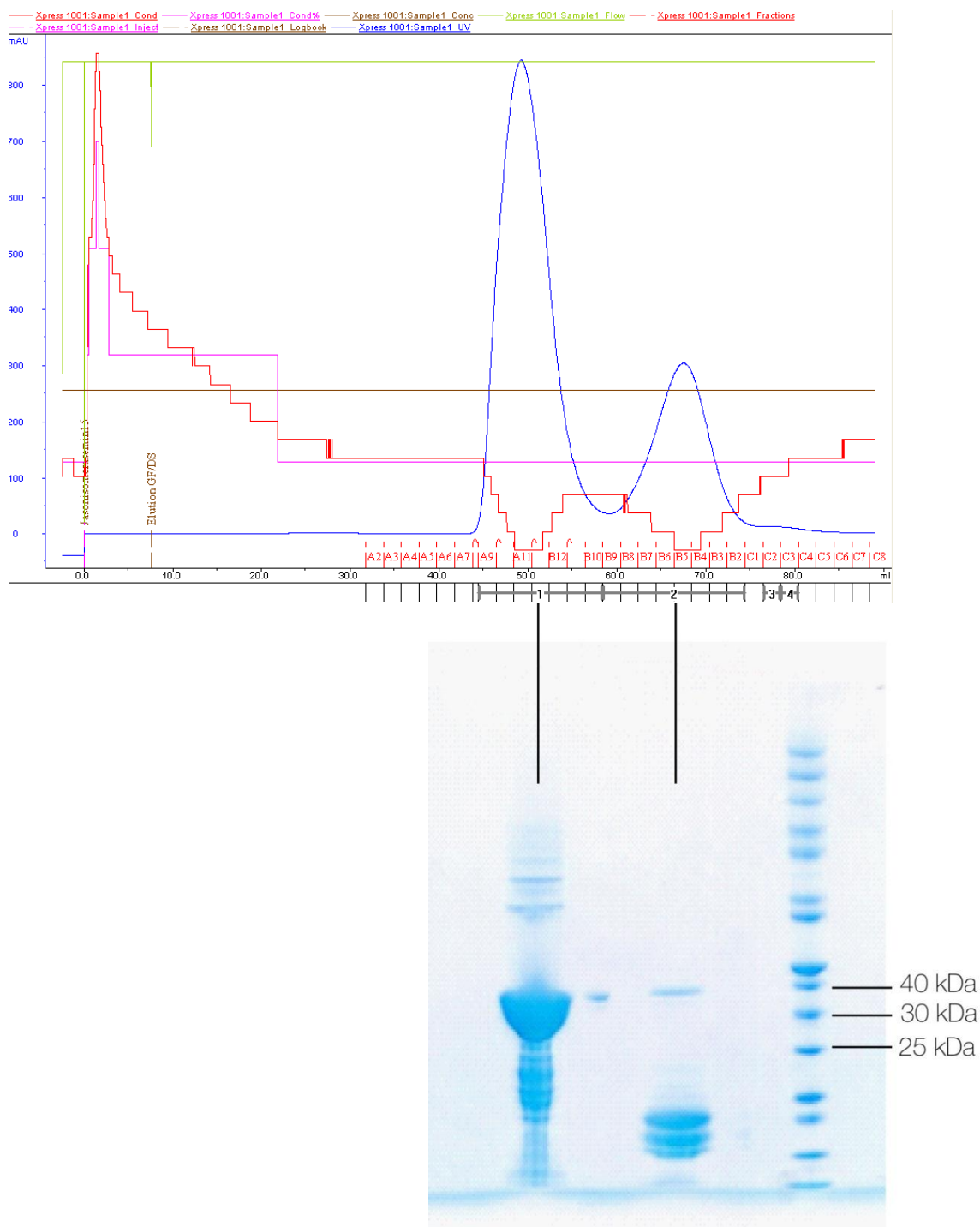

Fig 2.18.

(*Above*) The UV-trace from the SEC (Superdex 75) of C-*minus* 15 mutant of the 5-FDRPi.

(*Below*) SDS-PAGE analysis of the two peaks eluted from the SEC.

The major protein fraction eluting from the Superdex 75 gel filtration of the C-*minus* 15 isomerase was loaded onto a Superdex 200 column for gel filtration. Although the fraction purified as a single peak (Fig 2.19), the SDS-PAGE analysis of the fractions under this peak again showed a pattern consistent with protein degradation similar to Fig. 2.14. It was concluded that the C-*minus* 15 mutant is also unstable and prone to self-degradation.
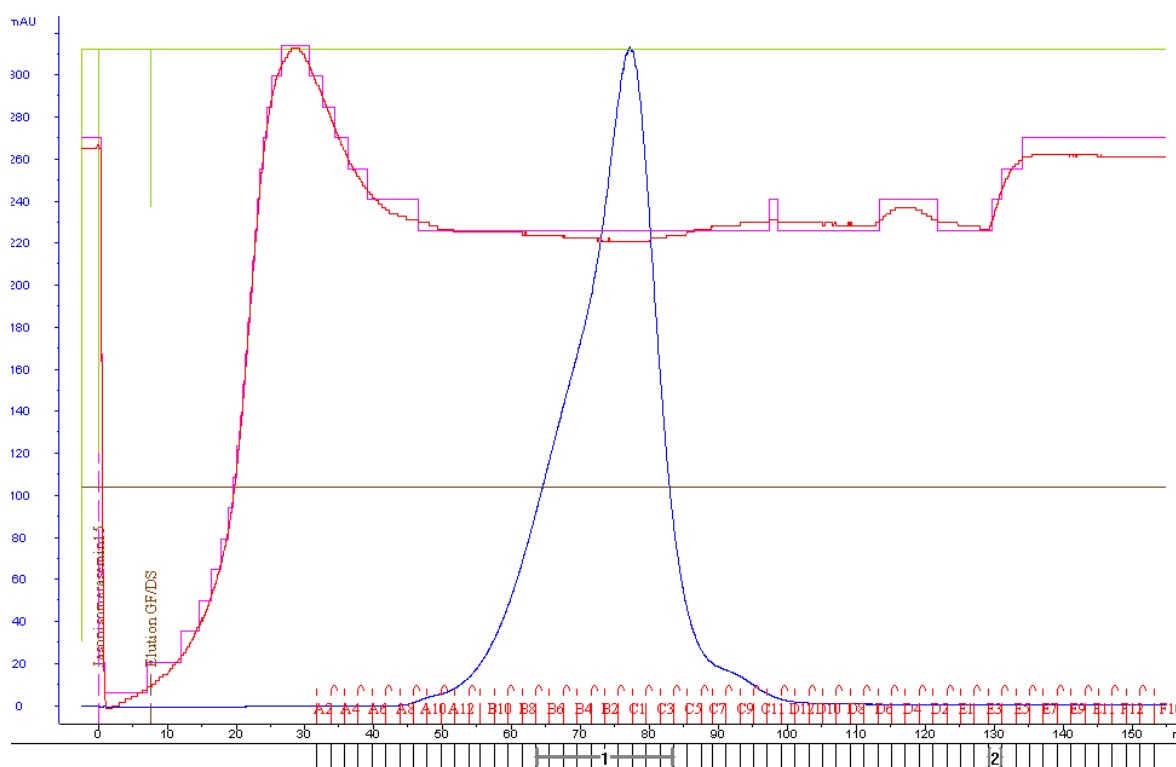

Fig 2.19. The UV-trace from a second SEC (Superdex 200) of the major fraction collected from the first SEC of the C-*minus* 15 mutant of the 5-FDRPi.

Purification of the C-*minus* 10 and C-*minus* 5 truncated 5-FDRPi's were also explored. Over-expression in *E. coli* was carried out and the cells were lysed by sonication. The cell lysates were cleared by centrifugation (20,000 rpm, 48,384 *g*) and the soluble and insoluble fractions were analysed by SDS-PAGE. The truncated peptide was present in both the soluble and insoluble fractions. The soluble fraction contained lower MW

bands which, from previous experience with other C-terminal truncations, indicated self-degradation products (Fig. 2.20). In view of the consistent instability of the truncation mutants, it was decided not to pursue this approach to crystallisation any further.
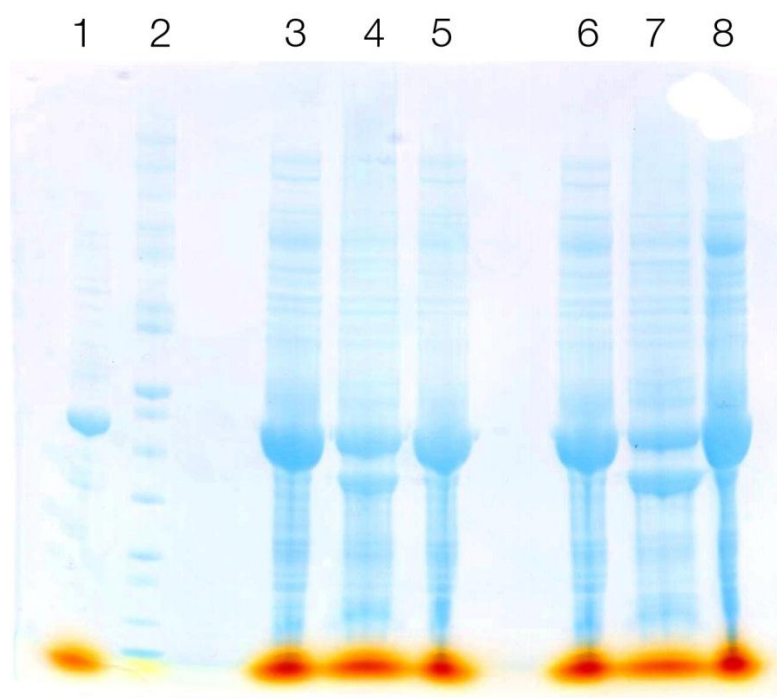

Fig 2.20. SDS-PAGE of the over-expression of C-minus 5 and C-minus 10 mutants.
1. native isomerase sample **2.** protein unstained ladder (PageRuler) **3.** C-*minus* 5 cell lysate **4.** C-*minus* 5 soluble fraction **5.** C-*minus* 5 insoluble fraction **6.** C-*minus* 10 cell lysate **7.** C-*minus* 10 soluble fraction **8.** C-*minus* 10 insoluble fraction.

In conclusion, the shortening of the 5-FDRPi appeared to cause the protein to become either insoluble or unstable. These truncated isomerase proteins were unsuitable for use in crystallisation trials.

2.4.4 Limited proteolysis of the 5-FDRPi

A truncated protein can sometimes be obtained by carefully controlled proteolysis using protease enzymes. The limited proteolysis procedure had been applied successfully in some cases where the proteolysed fragment of a full protein was crystallised and diffracted to solve a partial protein structure.[32]

A study of limited proteolysis on the 5-FDRPi was carried out with five proteases: tyrpsin, chymotrypsin, thermolysin, subtilisin and pepsin. The proteases were added to 5-FDRPi (40 µg in 20 ml PBS) in a ratio of 1: 100 and 1: 1000 by mass and the lysis reaction was carried out in each case for 1 h at 0 °C. The protein products were denatured and analysed on a SDS-PAGE gel.

The 1:1000 proteolysis reactions did not lead to observable degradation of the isomerase.

At 1:100 ratio, only thermolysin out of the five proteases produced a major protein band (25 kDa). Mass spectrometry analysis confirmed this fragment is part of the 5-FDRPi. None of the other lytic enzymes produced a coherent truncated peptide. Therefore a further study was carried out with thermolysin to optimise the ratio of protease to protein in order to prepare this peptide fragment (Fig 2.21).
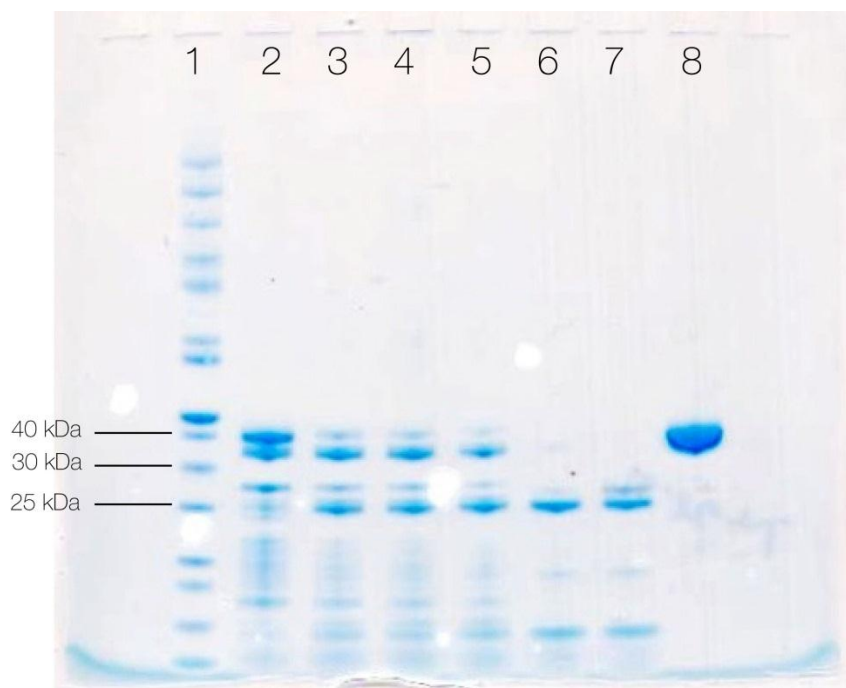
Fig 2.21. SDS-PAGE gel of thermolysin proteolysis of the 5-FDRPi. (Ratio = thermolysin : 5-FDRPi)

1. protein ladder **2.** 1:1000 **3.** 1:500 **4.** 1:300 **5.** 1:200 **6.** 1:100 **7.** 1:50 **8.** control

Accordingly, purified 5-FDRPi (20 mg in 2 ml) was treated with thermolysin in 1 : 100 ratio on ice for 1 hour, then EDTA-free protease inhibitor was added (Roche) to quench the reaction. The mixture was loaded to a SEC column (Superdex 200) and fractions of 2 ml each were collected and analysed by SDS-PAGE. The fractions containing the 25 kDa peptide were pooled. From the experiment starting with 20 mg of 5-FDPRi, the total yield of the 25 kDa fragment recovered was less than 1 mg (by Nanodrop measurement) and most of the protein mass had degraded into very small peptides. Due to this poor recovery, it was not feasible to obtain enough sample of the purified 25 kDa fragment for a protein crystallisation trial.

2.5   Crystallisation Trials

The vapour diffusion method is used widely to produce crystals of proteins.[33] For this technique drops of purified protein solution are usually mixed with a precipitant such as a salt or a large organic molecule. The drop is allowed to equilibrate with a reservoir of the precipitant mother liquor, and water is slowly removed from the drop into the mother liquor. This process increases the concentration of the protein and it becomes super-saturated. At this stage the protein can precipitate from solution into an amorphous structure, or it can remain in solution to form a meta-stable super-saturated solution. The meta-stable state is where crystallisation can occur if a nucleation event takes place. Upon nucleation the protein crystal grows, reducing the concentration of protein in solution. The protein crystal growth stops when the concentration has been reduced to below saturation (Fig. 2.22).
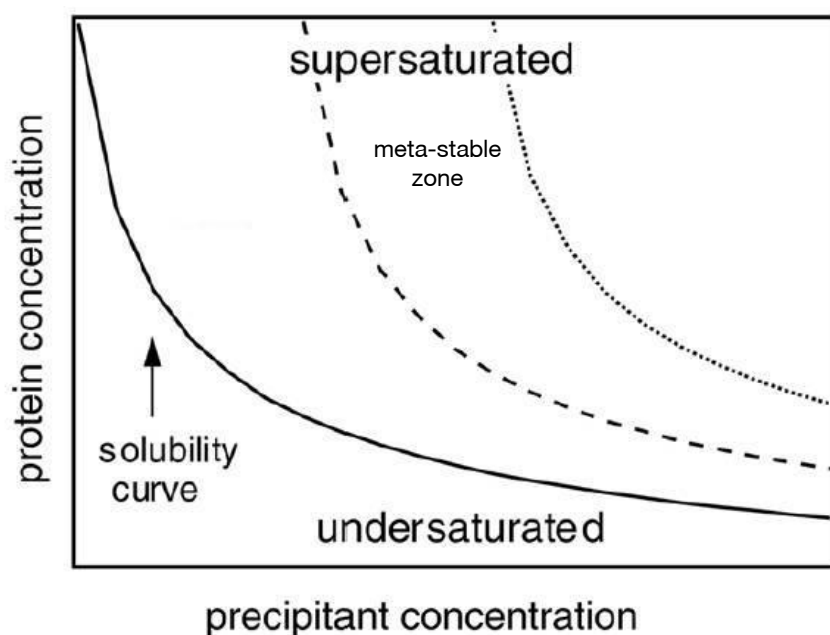

Fig 2.22. Protein concentration graph in a vapour diffusion droplet. Image adapted from [33].

2.5.1 Crystallisation of the native 5-FDRPi (*Nco*I-*Xho*I)

Protein crystallisation trials were conducted with the first batch of purified 5-FDRPi (10 or 15 mg/ml) from the first *Nco*I-*Xho*I construct. The sitting drop vapour diffusion technique was employed in 96-well crystallisation plates. Four different commercial screens: JCSG, PEGS, Wizards and Classics (Hampton Research) were used. The samples and mother liquors were pipetted by crystallisation robot (Honeybee 963, Genomic Solutions). Each condition was subjected to two drop sizes, which are 150 nl + 150 nl and 150 nl + 300 nl (protein + precipitant) and two sets of crystallisation plates were set up and kept separately in a cold room (4 °C) or at room temperature (20 °C).

Phosphonate analogue **61** was also used in some of the protein crystallisation screens. Phosphonate **61** (100 mM solution) was added to the 5-FDRPi solution at a final concentration of 1 mM and the mixture was incubated overnight at 4 °C and centrifuged (13,000 rpm, 16,060 *g*) to remove any aggregates prior to use in crystallisation experiments. The conditions are summarised in Table 2.9 below.

Table 2.9. The crystallisation conditions used for the first screening.

Screen	protein = 15 mg/ml without **61**	protein = 15 mg/ml with **61** (0.4 mM)	protein = 10 mg/ml without **61**	protein = 10 mg/ml with **61** (0.4 mM)
JSCG	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Basic PEGs	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Wizards	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Classics	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C

Three sets of conditions gave poorly formed needle-clusters. These hits were from the Wizards Screen (Emerald Biosystems) with phosphonate analogue **61** added to the drop (Table 2.10 & Fig. 2.23).

Table 2.10.   Crystal forming conditions

(protein concentration = 15 mg/ml with **61** at 0.4 mM, 20 °C)

Screen	Condition	Precipitant	Buffer	Salt
Wizards	F3	$(NH_4)_2SO_4$ 1.26 M	HEPES pH 7.5, 0.1 M	None
Wizards	G5	$(NH_4)_2SO_4$ 1.26 M	CHES pH 9.5, 0.1 M	NaCl 0.2 M
Wizards	H9	$(NH_4)_2SO_4$ 1.26 M	MES pH 6.0, 0.1 M	None

These crystallisation conditions highlighted ammonium sulfate as a common precipitant. They also showed that the protein tolerated a wide pH window(6 to 9.5) for crystal formation. These crystals formed at room temperature but not at 4 °C.
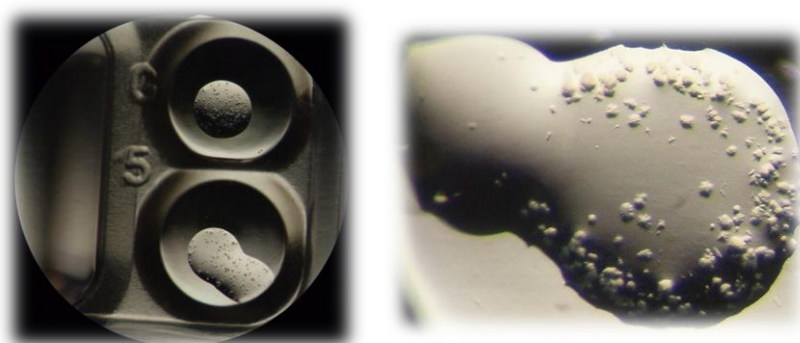

Fig 2.23. First crystals of the 5-FDRPi as needle clusters (from G5 Wizard screen).

In light of these initial results, optimisation screens were carried out aimed at improving the form of the crystals, since the small needle clusters were not suitable for X-ray diffraction.

Optimisation screens were set up using the hanging drop vapour diffusion method in 24 well crystallisation plates. Three protein + precipitant volume ratios were tested (0.6 µl + 0.6 µl, 0.6 µl + 1.2 µl and 1.2 µl + 0.6 µl) and applied by hand onto persilyated glass cover slips (Hampton Research). The drops were then suspended from an inverted cover slip of silylated glass over a reservoir of mother liquor (500 µl) and the chambers were sealed with grease. Both the pH (7.0, 7.5, 8.0, 8.5; HEPES 0.1 M) and the concentration of ammonium sulfate (0.7 M to 1.4 M, in increments of 0.1 M) were varied systematically in the screens.

Poorly formed spherolites/small cystals were obtained in these optimisation screens (Fig. 2.24) but these were not harvested for X-ray diffraction due to their low quality.
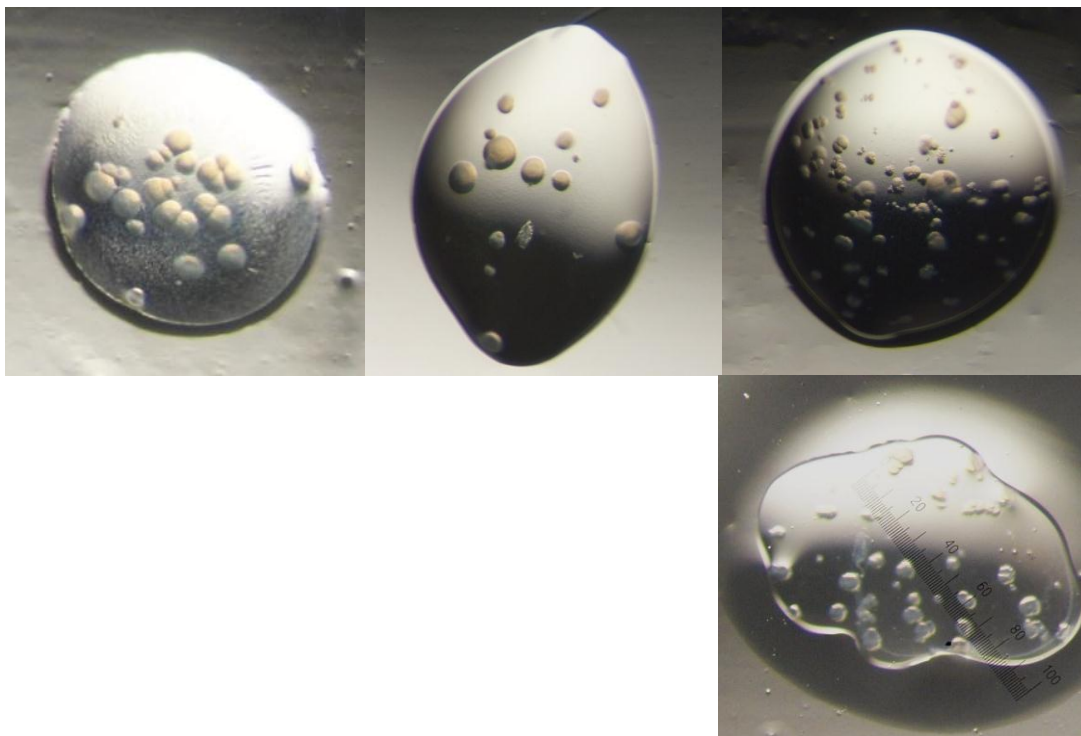

Fig 2.24. Selected images of the spherolite crystals of the isomerase from optimisation screens.

2.5.2 Crystallisation of the native 5-FDRPi (*BamH*I-*Xho*I)

As described above in 2.5.1, multiple crystallisation trials with the native form of 5-FDRPi that was prepared using the *Nco* I–*Xho* I cloning sites have failed to produce diffraction quality protein crystals. The associated protein purification involved a four-day TEV digestion at 20 °C and this prolonged standing at 20 °C could reduce the quality of the protein sample. For the new isomerase prepared using the *BamH* I–*Xho* I cloning sites, this incubation has been reduced to 2 h. In addition, more advanced techniques were introduced to improve the purification procedure. This included the use of a cell disruptor for cell lysis and the Akta system for protein chromatography. This resulted in a cleaner, higher quality protein sample and should improve crystallisation.

Stochastic screening was carried out using 16 different sparse matrix screens prepared from a range of known crystallisation conditions (Table 2.11). The samples and mother liquors were pipetted by crystallsation robot (Honeybee 963, Genomic Solutions). The drop sizes for the two wells under each set of conditions were 150 nl + 150 nl and 150 nl + 300 nl (protein + precipitant) and two sets of crystallisation plates were set up and kept separately in a cold room (4 °C) or at room temperature (20 °C). Two protein concentrations were used (13.7 mg/ml and 10 mg/ml).

The phosphonate analogue **61** was also used in some of these protein crystallisation screens. The phosphonate **61** (100 mM solution) was added to the 5-FDRPi solution at a final concentration of 1 mM and the mixture was incubated overnight at 0 °C and centrifuged (13,000 rpm, 16,060 *g*) to remove any aggregates prior to the crystallisation experiments.

**Table 2.11.** The crystallisation conditions used for the first screening.

Screen	protein = 13.7 mg/ml without inhibitor	protein = 13.7 mg/ml with inhibitor	protein = 10 mg/ml without inhibitor	protein = 10 mg/ml with inhibitor
JSCG	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Wizards	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Classics	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto PEGs 1	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto PEGs 2	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto PEGs 3	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto PEGs 4	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
$(NH_4)_2SO_4$ suite	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 14	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 16	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 17	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 18	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 19	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 20	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 21	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C
Sto 22	4 °C	4 °C	4 °C	4 °C
	20 °C	20 °C	20 °C	20 °C

Initially two sets of conditions gave small crystals after 4 days. These had a more defined shape and form than the irregular crystals obtained in the previous crystallisation attempts. Other conditions were also successful but the crystals took longer to grow. A list of the conditions that produced these improved crystals is given below (Table 2.12, Fig. 2.25).

Table 2.12.   Crystal forming conditions

(protein = 13.7 mg/ml with inhibitor at 1 mM where indicated, 20 °C)

Screen	Observed after	Precipitant	Buffer	Salt / additive
Sto21 G4 with inhibitor	4 days	PEG 8000 22.04 %	Tris HCl pH 7.5, 0.1 M	$MgSO_4$ 0.22 M
Sto PEGs 3 H1 with inhibitor	2.5 weeks	PEG 8000 15 %	none	$MgSO_4$ 0.22 M
Sto 20 H10 with inhibitor	4 weeks	PMME 550 36 %	Tris HCl pH 8, 0.1 M	$Mg(OAc)_2$ 0.18 M
Sto PEGs 4 B12 with inhibitor	4 weeks	PEG 3350 22.6 %	Sodium cacodylate pH 7.0, 0.1 M	$MgSO_4$ 0.27 M
Sto 17 B11 no inhibitor	8 weeks	PEG 400 38 %	Sodium cacodylate pH 7.0, 0.1 M	$(NH_4)_2HPO_4$ 0.15 M, β-mercaptoethanol 8.17 M
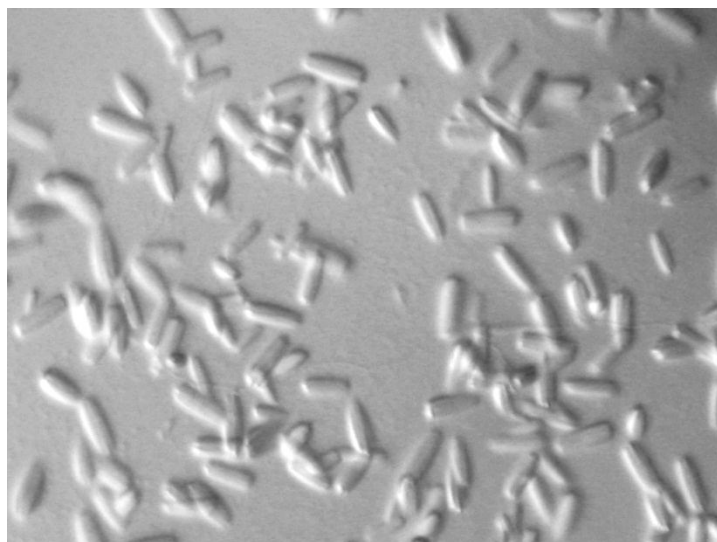
Fig 2.25. Small crystals of 5-FDRPi with phosphonate **61** obtained in Sto21 G4.

Poly(ethylene glycol) (PEG) was used as the precipitant in all the crystal forming conditions. Optimisation screening was carried out on the Sto21 G4 hit. The screens were set up by hand using the hanging drop vapour diffusion method, with droplet volumes of 2 μl + 1 μl, 1 μl + 1 μl and 1 μl + 2 μl (protein + precipitant) over a reservoir of the precipitant (0.5 ml) and then kept at room temperature (20 °C).

The refinements first focused on the PEG 8000 conditions. From the optimisation, many droplets produced crystals. The best crystals were formed at pH 7.7 to 7.8 and PEG 8k (15.5 – 16.0 %) and $MgSO_4$ (0.22 M). $MgSO_4$ was important for crystal formation, other sulfate salts (Li_2SO_4, Na_2SO_4, K_2SO_4, $(NH_4)_2SO_4$, $ZnSO_4$), group II chlorides ($CaCl_2$, $SrCl_2$, $BaCl_2$) and magnesium salts ($MgCl_2$, $Mg(OAc)_2$, $Mg(NO_3)_2$) were screened but did not encourage crystal formation.

Under these optimised conditions, larger crystals were observed after 1 week and continued to grow for about 3 to 6 weeks. The crystals appeared as oval rods (Fig. 2.26) with the facets displaying multiple layers, indicative of overlapping crystal lattices within each crystal. Most formed as single crystals, but twinned crystals were sometimes obtained. The best crystals were picked for diffraction at the in-house facility (described in 2.7), but they diffracted poorly.
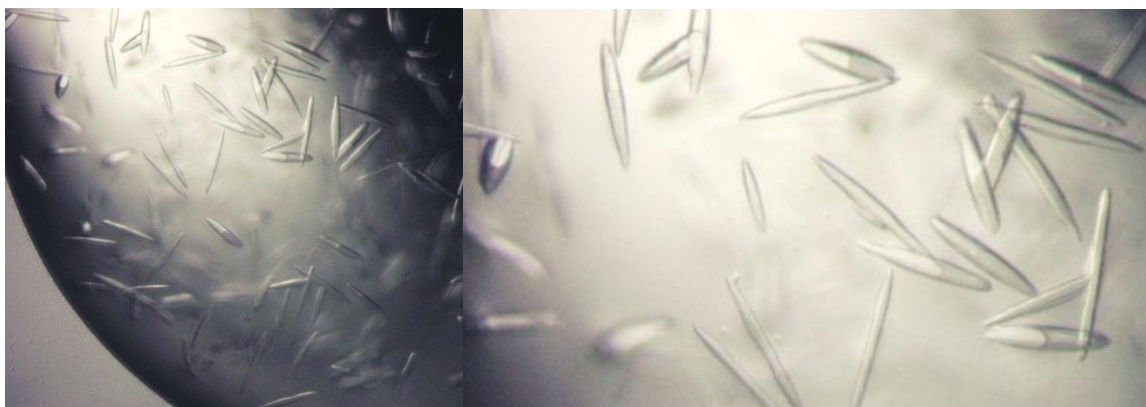

Fig 2.26.   Rod shaped crystals of the 5-FDRPi with phosphonate **61** obtained from optimisation screens based on the condition 16 % PEG 8k, Tris 0.1 M pH 7.7, $MgSO_4$ 0.22M, 20 °C .

Further iterations of the optimisation from the Sto 21 G4 conditions did not result in a significant improvement of the crystal form. All of these crystals displayed multiple facets and diffracted poorly.

The Sto17 B11 conditions (PEG 400 38 %, sodium cacodylate 0.1 M, ammonium phosphate 0.15 M and β-mercaptoethanol 8.17 mM) gave a single cubic crystal in the 96-well sitting drop plate after 2 months. The crystal was picked for diffraction but was damaged in the freezing procedure in liquid nitrogen. The resulting diffraction was poor.

Despite multiple repeats to optimise from Sto17 B11 conditions, in both 96-well sitting drop plates set up using the Honeybee robot and 24-well hanging drop plates set up by hand, in which all four precipitant ingredients and the protein concentration have been systematically varied, the crystallisation could not be reproduced.

2.5.3 Further optimisation with PEG conditions

Polyethylene glycols of different chain length have been effective as protein precipitants under different crystal-forming conditions. With this knowledge base, an experiment was designed to explore the effect of the many different PEGs available commercially.

As the chain length of the polyethylene glycols increase, they become stronger protein precipitants. In order to explore a range of PEGs with different molecular weights (MW), the known crystallisation conditions with PEG as the precipitant were plotted onto a graph of MW against the PEG concentration. The data points formed a pattern that could be fitted to a curve (Fig 2.27). Based on this curve, the concentration range of each PEG to test for 5-FDRPi crystallisation was obtained (Table 2.13) and solutions were made up in deep well blocks (96-well format) for screening.
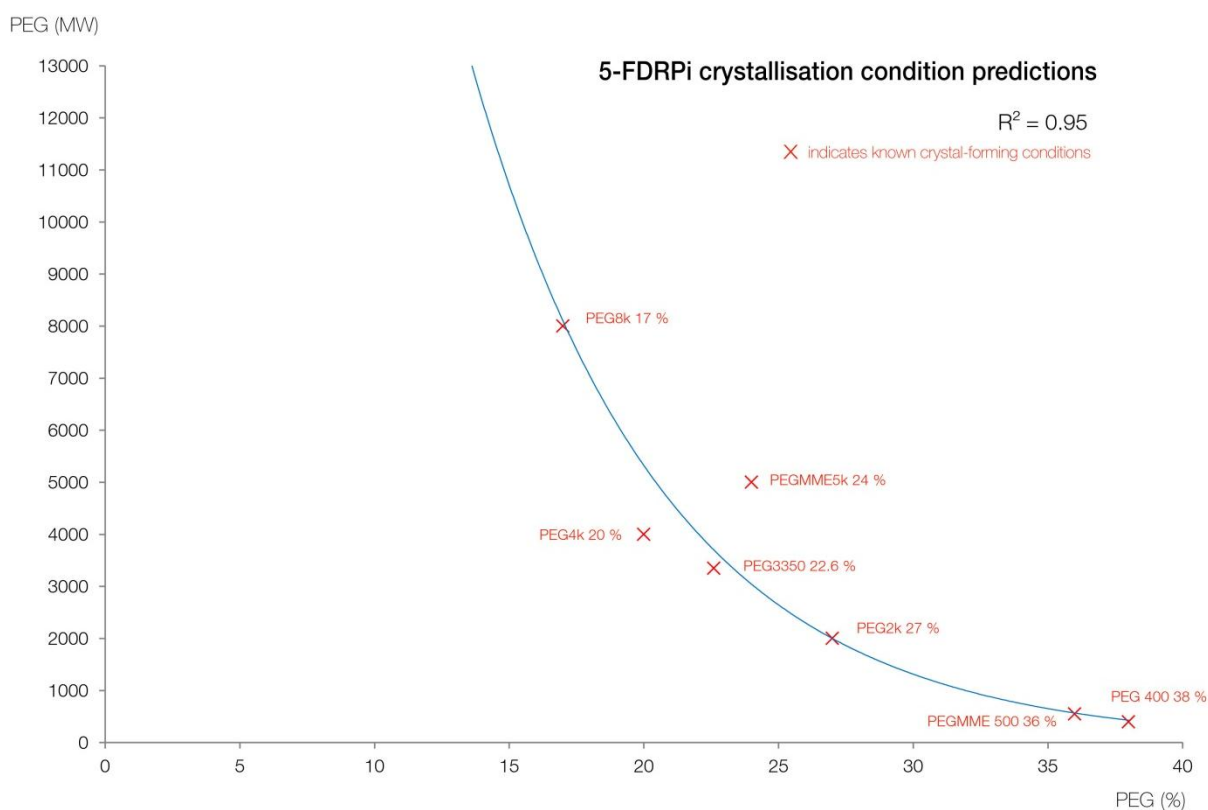
Fig 2.27. A graph plotting PEG molecular weights against PEG concentration for conditions that led to 5-FDRPi crystallisation.

Table 2.13. Different PEGs screened for isomerase crystallisation using concentration ranges derived from the graph of Fig. 2.23.

PEGs	Concentration range	Buffer	Salt
PEG 400	34 – 45 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG MME 500	31 – 42 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 1000	27 – 38 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 1450	24 – 33 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 2000	19 – 30 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 3350	17 – 28 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 4000	14 – 25 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG MME 5000	14 – 25 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 6000	12 – 23 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 8000	9 – 20 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 10000	6 – 17 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M
PEG 12000	4 – 15 %	Tris HCl pH 7.8, 0.1 M	$MgSO_4$ 0.22 M

Several sets of these conditions gave crystals. They grew as multiple rod-shaped crystals sticking together forming multi-layered crystals. However the multi-layered crystals did not diffract to a high resolution. The crystal forming conditions are listed in Table 2.14.

Table 2.14. Crystal forming conditions (protein = 13.7 mg/ml with inhibitor at 1 mM, 20 °C)

PEG	concentration	Observed after	Buffer	Salt / additive
PEG 2k	27-28 %	1 week	Tris HCl pH 7.8, 0.1 M	MgSO$_4$ 0.22 M
PEG 4k	19-20 %	2 weeks	Tris HCl pH 7.8, 0.1 M	MgSO$_4$ 0.22 M
PEG MME 5k	22-25 %	1 week	Tris HCl pH 7.8, 0.1 M	MgSO$_4$ 0.22 M
PGE 8k	15-18 %	1 week	Tris HCl pH 7.8, 0.1 M	MgSO$_4$ 0.22 M

Optimisation was carried out with all four conditions reported in Table 19 using 24-well hanging-drop crystallisation trays. The concentration of PEG, pH and protein to precipitant ratios were refined systematically. Although the crystallisation could be reproduced (Fig. 2.28), the crystals obtained did not improve in visual appearance.
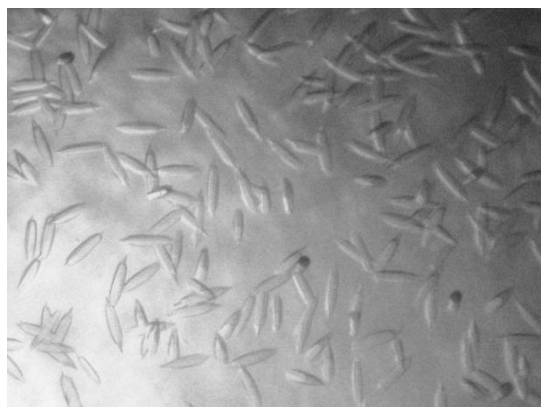
Fig 2.28. Crystals of 5-FDRPi with phosphonate **61** obtained from the optimisation screen with the condition of 22 % PEG MME 5k, pH 7.7, MgSO$_4$ 0.22 M, 20 °C.

2.5.4 Seeding Experiments

Seeding is a technique used to speed up the formation of crystals which by-passes the nucleation event before the crystal growth phase. The 'seeds' used for these experiments were fragments of a crushed isomerase protein crystal. Such fragments can act as a nucleation site where protein crystal growth may be induced. This technique had been applied successfully to produce large single crystals.[34, 35]

A single 5-FDRPi crystal from the optimisation experiments was extracted from the droplet using a micro-loop and viewing under a microscope. This was transferred into the precipitant solution (10 µl) and then crushed with a microtool (Hampton Research). The slurry was transferred to a tube containing precipitant solution (500 µl, PEG 8k 15.5 %, 0.1 M Tris-HCl, MgSO$_4$ 0.22 M). The mixture was vortexed vigorously to yield a stock suspension of crystal fragments.

Seeding experiments were set up in 24-well hanging drop plates by hand, with droplets containing 1 µl or 2 µl of 13.7 mg/ml 5-FDRPi mixed with 0.5 µl of crystal fragment stock and 1 µl or 2 µl of precipitant solution (PEG 8k 15.5 %, 0.1 M Tris-HCl, MgSO$_4$ 0.22 M).

Seeding did not improve the crystal form. The crystals which did form in these experiments were very small in size due to over-nucleation (Fig. 2.29).

Fig 2.29. A seeding experiment drop with very small crystals of the 5-FDRPi due to over-nucleation.

2.5.5 Additive Screen

Additives such as detergents, small organic molecules, multivalent metal salts and, co-factors are widely used to improve the success of obtaining diffraction quality protein crystals.[33, 36] There are cases where additives were used to alter the interactions between protein molecules and improved the quality of protein crystals.[37, 38] An additive screen in 96-well format is commercially available from Hampton Research. However since most of the additive reagents employed in their screen were already available locally, an additive screen was prepared in-house based closely on their formulation.[39]

The additive screens were set up using two reliable crystal forming conditions. (15.5 % PEG 8k/0.1 M Tris HCl pH 7.7/0.22 M $MgSO_4$ and 22.0 % PEG MME 5k/0.1 M Tris HCl pH 7.7/0.22 M $MgSO_4$). The additive solutions (50 nl) were initially added directly to the 5-FDRPi (13.7 mg/ml) and precipitant drops (150 nl + 150 nl and 300 nl + 150 nl) but this resulted in precipitation of the protein in most of the wells. Therefore the additive was first diluted into the precipitant (100 µl additive solution + 900 µl precipitant) and then the crystal trays were set up to assess crystallisation.

Most of the wells produced crystals after 4 days. The crystals obtained with additives in the droplets showed a variation in size and shape, but they did not improve in form. All crystals visually displayed multiple facets on the surfaces (Fig. 2.30).
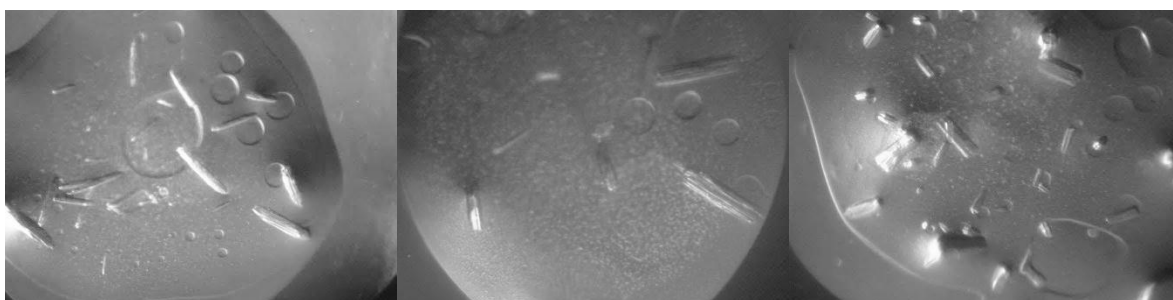
Fig 2.30. Crystals of 5-FDRPi obtained from the additive screens showing different sizes and shapes.

2.5.6 *In situ* limited proteolysis crystallisations

The limited proteolysis experiments (Section 2.4.4) on 5-FDRPi identified that thermolysin can produce a major protein fragment (*ca.* 25 kDa) under controlled conditions, but this fragment could not be easily prepared in sufficient quantity for crystallisation. Some success has been achieved by carrying out the limited proteolysis *in situ*.[40] In such experiments, a trace quantity of protease is added to each droplet and the proteolysis takes place during the crystallisation experiment.

Droplets containing a trace of thermolysin in a 1 : 1000 or 1 : 100 ratio to 5-FDRPi (13.7 mg/ml) were subjected to stochastic crystallisation screening using the same 16 in-house prepared screens and conditions (Table 2.11). In the event, protein crystals were not observed from these crystal screens.

2.5.7 Lysine methylated 5-FDRPi crystallisation

A lysine methylated 5-FDRPi solution (15 mg/ml) prepared as described in Section 2.4.1 was subjected to crystallisation screening using four commercial suites: JCSG, PEGs, Wizards and Classics (Hampton Research). The samples and mother liquors were pipetted by liquid handling robot (Honeybee 963, Genomic Solutions). The drop sizes for the two wells under each condition were 150 nl + 150 nl and 150 nl + 300 nl (protein + precipitant) and two sets of crystallisation plates were set up and kept separately in a cold room (4 °C) or at room temperature (20 °C).

Crystallisation was not observed in the any of eight screens prepared. Protein precipitation was observed in about 40 % of the wells.

2.6 Diffraction and data collection

The better formed crystals from the various crystallisation trials and optimisation plates were picked for X-ray diffraction (Table 2.15).

Table 2.15. X-ray diffraction experiments carried out with 5-FDRPi crystals.

Crystallisation conditions	N$^{\circ}$ of crystals diffracted	Cryoprotectant	crystal size (average)	X-ray source
PEG 8k 18.0 % Tris pH 7.8 0.1 M MgSO$_4$ 0.22 M	7	glycerol 30%	50 μm	in-house (St Andrews)
PEG 8k 16.0 % Tris pH 7.7 0.1 M MgSO$_4$ 0.11 M	9	glycerol 30%	50 μm	in-house (St Andrews)
PEG 8k 16.0 % Tris pH 7.8 0.1 M MgSO$_4$ 0.22 M	5	glycerol 30%	20 μm	Diamond beamline (Oxford)
PEG 8k 15.0 % MgSO$_4$ 0.22 M	4	sucrose 30%	50 μm	in-house (St Andrews)
PEG 8k 14.0 % Tris pH 7.7 0.1 M MgSO$_4$ 0.22 M	10	sucrose 30 %	20 μm	in-house (St Andrews)
PEG 400 38 % Na cacodylate pH 7 0.1 M BME 8.17 mM (NH$_4$)$_2$HPO$_4$ 0.15 M	1	glycerol 30 %	20 μm	in-house (St Andrews)
PEG MME 5k 20.0-22.0 % Tris pH 7.7 0.1 M MgSO$_4$ 0.22 M	3	sucrose 30 %	40 μm	in-house (St Andrews)

The crystals were picked with a micro-loop tool under the microscope. The looped crystals were cryo-protected with a mother-liquor supplemented with 30 % glycerol and 1 mM phosphonate analogue and then flash frozen in liquid nitrogen.

The data were collected in-house (University of St Andrews) at 100 K on a Rigaku 007HFM rotating anode X-ray generator with a Saturn 944 CCD detector.

A few crystals were sent to the UK National Synchrotron facility Diamond beamline, and data collection was carried out with a micro-focus beam with high intensity X-ray.

Some of the crystals diffracted up to 4 Å resolution (Fig. 2.31 and 2.32). However the reflections suffered very low intensity. There was a streaking pattern arising from the multiple crystal lattice overlap within one crystal.

Unit cell parameters were calculated on several of the diffraction patterns and the measurements from different diffractions matched. The measured crystal unit cell was orthorhombic. The parameters are shown below in Table 2.16.

Table 2.16.   Measure crystal unit cell parameter

Parameter	Measurement
space group	C221
a	158.023 Å
b	325.858 Å
c	88.709 Å
α	90.0°
β	90.0°
γ	90.0°
mosaicity	0.84°
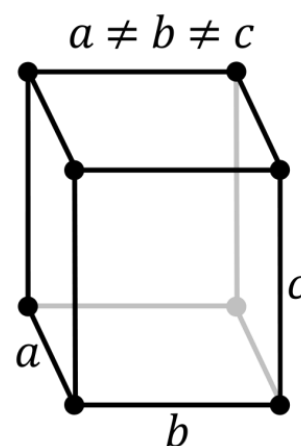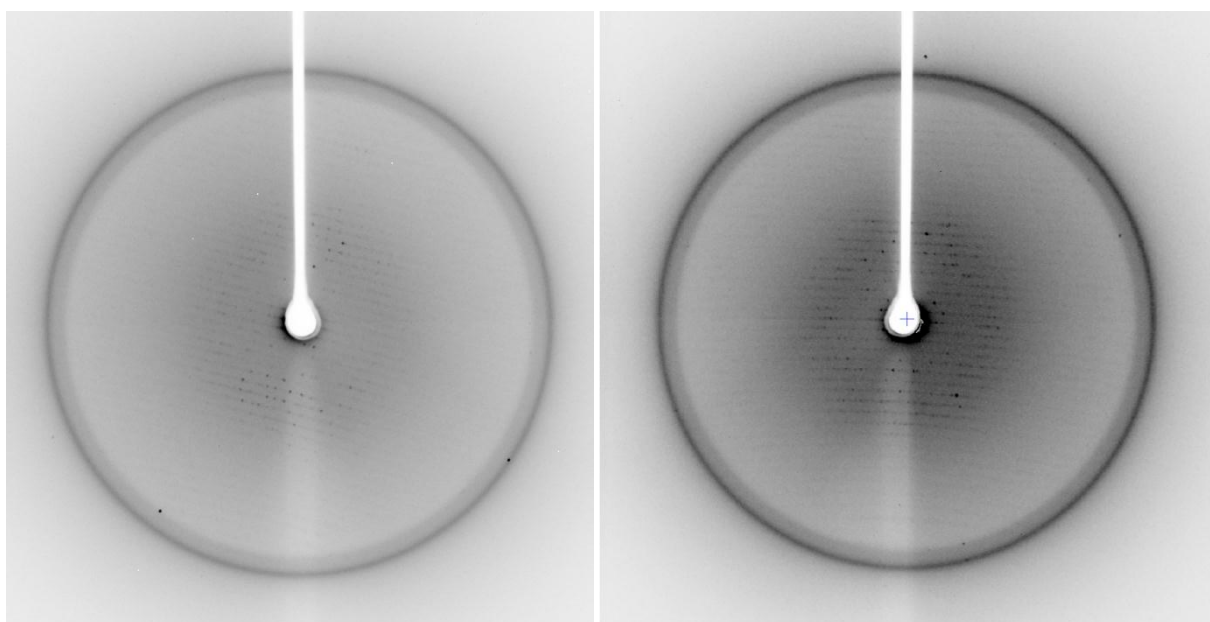
Fig 2.31. Two diffraction patterns of 5-FDRPi crystals obtained from X-ray facility in St Andrews.
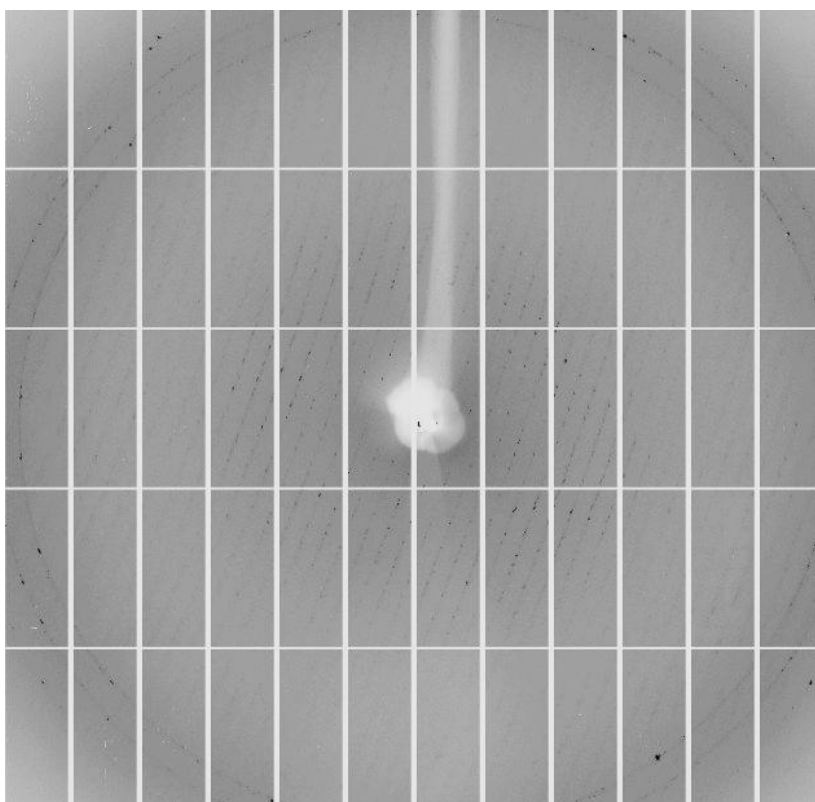
Fig 2.32. Diffraction pattern of 5-FDRPi crystals from a synchrotron micro-focus X-ray beam (Diamond beamline, Oxford).

2.7   Conclusions

The project set out to obtain an X-ray crystal structure of the isomerase (5-FDRPi) of the fluorometabolite pathway of *S. cattleya*, which would provide a valuable insight into exploring the mechanism of this enzyme.

The native enzyme was successfully cloned and over-expressed from an *E. coli* recombinant strain. The protein was successfully purified to homogeneity. Crystallisation of the native 5-FDRPi was carried out and crystals were obtained in many instances. These crystals appeared as rods with overlapping layers and diffracted poorly due to this defect. Optimisations did not result in an improvement in the diffraction quality of the crystals obtained.

In addition to the native enzyme, truncation mutants prepared by cloning experiments or proteolysis have been explored. Modification of the protein by lysine methylation was also carried out. However these experiments did not produce crystals for diffraction.

The production of high quality protein crystals remains a major bottleneck in protein structure determination,[41] and this is also the most unpredictable part of the process. In this case much effort was put into overcoming this challenge, but unfortunately the structure of 5-FDRPi was not obtained in the end.

2.8 Chapter 2 References

1. I. A. Rose, *Adv. Enzymol.*, 1975, **43**, 491-517.

2. S. V. Rieder, I. A. Rose, *J. Biol. Chem.*, 1958, **234**, 1007-1010.

3. T. D. Fenn, D. Ringe, G. A. Petsko, *Biochemistry*, **2004**, 43, 6464-6474.

4. S. V. Rieder, I. A. Rose, *J. Biol. Chem.*, 1959, **234**, 1007-1010.

5. T. Alber, D. W. Banner, A. C. Bloomer, G. A. Petsko, D. Phillips, P. S. Rivers, I. A. Wilson, *Phil. Trans. R. Soc. Lond. B*, 1981, **293**, 159-171.

6. A. C. O'Donoghue, T. L. Amyes, J. P. Richard, *Biochemistry*, 2005, **44**, 2610-2621.

7. A. C. O'Donoghue, T. L. Amyes, J. P. Richard, *Biochemistry*, 2005, **44**, 2622-2631.

8. K. N. Allen, A. Lavie, G. K. Farber, A. Glasfeld, G. A. Petsko, D. Ringe, *Biochemistry*, 1994, **33**, 1481-1487.

9. H. Tabor, S. M. Rosenthal, C. W. Tabor, *J. Biol. Chem.*, 1958, **233**, 907-914.

10. A. Sekowska, A. Danchin, *BMC Microbiol.*, 2002, **2**, 8.

11. A. Sekowska, V. Dénervaud, H. Ashida, K. Michoud, D. Haas, A. Yokota, A. Danchin, *BMC Microbiol.*, 2004, **4**, 9.

12. H. Tamura, Y. Saito, H. Ashida, T. Inoue, Y. Kai, A. Yokota, H. Matsumura, *Protein Sci.*, 2008, 17, 126-135.

13. M. Bumann, S. Djafarzadeh, A. E. Oberholzer, P. Bigler, M. Altmann, H. Trachsel, U. Baumann, *J. Biol. Chem.*, 2004, **279**, 37087-37094.

14. Y. Saito, H. Ashida, C. Kojima, H. Tamura, H. Matsumura, Y. Kai, A. Yokota, *Biosci. Biotechnol. Biochem.*, 2007, **71**, 2021-2028.

15. H. J. Imker, A. A. Fedorov, E. V. Fedorov, S. C. Almo, J. A. Gerlt, *Biochemistry*, 2007, **46**, 4077-4089.

16. H. Deng, S. M. Cross, R. P. McGlinchey, J. T. G. Hamilton, D. O'Hagan, *Chem. Biol.*, 2008, **15**, 1268-1276.

17. C. Zhao, P. Li, Z. Deng, H. Y. Ou, R. P. McGlinchey, D. O'Hagan, *Bioorg. Chem.*, 2012, **44**, 1-7.

18. V. Barbe, M. Bouon, S. Mangenot, B. Badet, J. Poulain, B. Segurens, D. Valleet, P. Marliere, J. Weissenbach, *J. Bacteriol.*, 2011, **193**, 5055-5056.

19. P. Nasomjai, *Ph. D. thesis*, University of St Andrews, 2010.

20. P. Nasomjai, D. O'Hagan, A. M. Z. Slawin, *Beilstein J. Org. Chem.*, 2009, 5, 37.

21. S. Cross, *Ph. D. thesis*, University of St Andrews, 2009.

22. H. Liu, J. H. Naismith, *Protein Expression Purif.*, 2009, **63**, 102-111.

23. L. H. Hansen, S. Knudsen, S. J. Sørensen, *Current Microbiol.*, 1998, **36**, 341-347.

24. M. Onega, R. P. McGlinchey, H. Deng, J. T. G. Hamilton, D. O'Hagan, *Bioorg. Chem.*, 2007, **35**, 375-385.

25. T. D. Ashton, P. J. Scammells, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3361-3363.

26. R. McGlinchey, *PhD thesis*, University of St Andrews, 2006.

27. R. B. Kapust, J. Tözsér, T. D. Copeland, D. S. Waugh, *Biochem. Biophys. Res. Commun.*, 2002, **294**, 949-955.

28. T. S. Walter, C. Meier, R. Assenberg, K. F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart, J. M. Grimes, *Structure*, 2006, 11, 1617-1622.

29. Y. Kim, P. Quartey, H. Li, L. Volkart, C. Hatzos, C. Chang, B. Nocek, M. Cuff, J. Osipiuk, K. Tan, Y. Fan, L. Bigelow, n. Maltseva, R. Wu, M. Borovilos, E. Duggan, M. Zhou, T. A. Binkowski, R-G Zhang, A. Joachimiak, *Nat. Methods*, 2008, **5**, 853-854.

30. *JBS Methylation Kit Manual*, Jean Bioscience GmbH.

31. B. Rost, C. Sander, *J. Mol. Biol.*, 1993, **232**, 584-599.

32. S. Quevillon-Cheruel, B. Collinet, L. Trésauques, P. Minard, G. Henckes, R. Aufrère, K. Blondeau, C. Z. Zhou, D. Liger, N. Bettache, A. Poupon, I. Aboulfath, N. Leulliot, J. Janin, H. van Tilbeugh , *Methods Mol. Biol.*, 2007, **363**, 21-37.

33. R. C. Stevens, *Curr. Opin. Struct. Biol.*, 2000, **10**, 558-563.

34. C. Thaller, L. H. Weaver, G. Eichele, E. Wilson, R. Karlsson, J. N. Jansonius, *J. Mol. Biol.*, 1981, **147**, 465-469.

35. T. Bergfors, *J. Struct. Biol*, 2003, **142**, 66-76.

36. B. D. Santarsiero, D. T. Yegian, C. C. Lee, G. Spraggon, J. Gu, D. Scheibe, D. C. Uber, E. W. Cornell, R. A. Nordmeyer, W. F. Kolbe, J. Jin, A. L. Jones, J. M. Jaklevic, P. G. Schultz, R. C. Stevens, *J. Appl. Cryst.*, 2002, **35**, 278-281.

37. R. Sousa, *Acta Cryst.*, 1995, **51**, 271-277.

38. S. Trakhanov, F. A. Quiocho, *Protein Sci.*, 1995, **4**, 1914-1919.

39. Hampton Research Additive Screen HR2-138 Reagent Formulation, http://hamptonresearch.com/documents/product/hr003789_binder1.pdf

40. A. Dong, X. Xu, A. M. Edwards, *Nat. Methods*, 2007, **12**, 1019-1021.

41. N. E. Chayen, E. Saridakis, *Nat. Methods*, 2008, **5**, 147-153.

3. Studies explaining the isomerase mechanism

3.1 Protein sequence of 5-FDRPi

The sequence of 5-FDRPi (*SCATT* 20080) from the fluorometabolite pathway of *S. cattleya* was compared with known protein sequences using the Basic Local Alignment Search Tool (BLAST).[1, 2] 5-Methylthioribose 1-phosphate isomerases (MTRPi) and the α-subunit of the eukaryotic translation initiation factor eIF-2B[2] emerged as the most homologous genes in the database. The closest homologues (70 – 80% homology) were from various *Streptomyces* species. These proteins belonged to the COG-0182 superfamily.[3]

From the protein crystal structure of the MTRPi from *Bacillus subtilis,* the enzyme active site was identified. The active site contained several key catalytic residues which were highly conserved.[4] The linear sequence of 5-FDRPi was aligned with the sequences of MTRPi from *B. subtilis* (35 % homology),[5] *Streptomyces lividans* (76 % homology)[6] and *Saccharomyces cerevisiae* (31 % homology)[7] using the alignment tool CLUSTAL omega (Fig 3.1).[8] X-ray crystal structures of the *B. subtilis* and *S. cerevisiae* MTRPi have been solved.[4, 9]
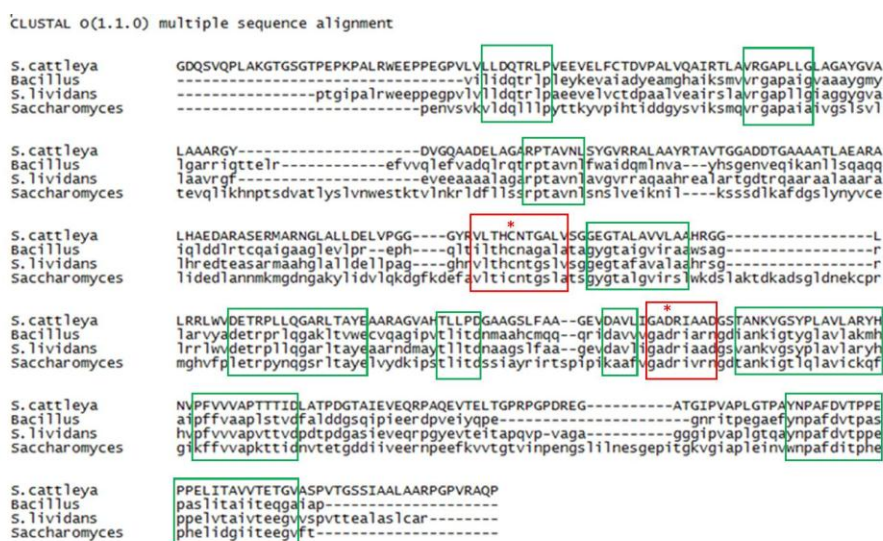
Fig 3.1. Clustal-omega sequence alignment of 5-FDRPi with other MTRPi.[10] **Red boxes** indicate the two key residues C177 and D257. **Green boxes** indicate other conserved regions.

The catalytic residues Cys177 and Asp257 were conserved through all four isomerases, and the five residues adjacent to these two sites were mostly conserved. This suggests that the active site arrangement in 5-FDRPi will closely resemble that of the other MTRPi's. The other regions of the protein have some other conserved contigs by comparison. Therefore the arrangement of the secondary structures of 5-FDRPi is expected to loosely resemble the structure models of Bacillus and yeast MTRPi's.
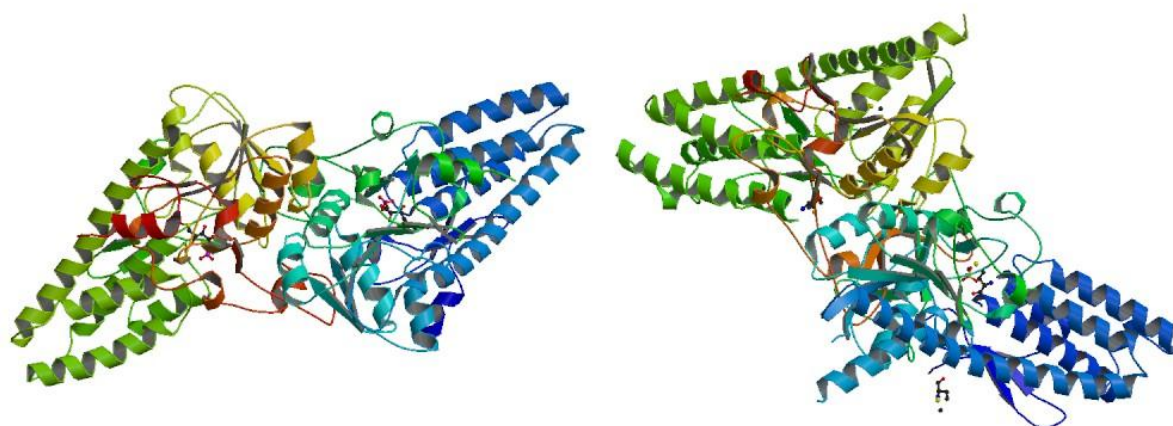
Fig 3.2. X-ray crystal structures of two MTRPi's. **Left:** 2YVK from *B. subtilis*;[4] **Right:** 1W2W from yeast.[9]

The crystal structures of Bacillus MTRPi (2YVK) and yeast MTRPi (1W2W) are dimers with two active sites (Fig 3.2). The two structures resemble each other closely. Their active sites are located in a pocket formed by the α-helices held together by loops. These structures are assumed to show the relaxed state of the enzyme. The Bacillus MTRPi was co-crystallised with the product (5-methylthioribulose 1-phosphate) and the yeast MTRPi with sulfate ions. In the relaxed state, the active site pocket is opened to the solvent to allow the substrate/product to enter/depart.

The secondary structure from the X-ray crystal structures of these two enzymes showed a similar arrangement. In each case their monomeric units consist of segments of α-helices and β-strands, linked by short β-turns and loops (Fig. 3.3). The secondary structure prediction of 5-FDRPi based on its linear sequence displayed a similar pattern (Fig. 3.4).[11]
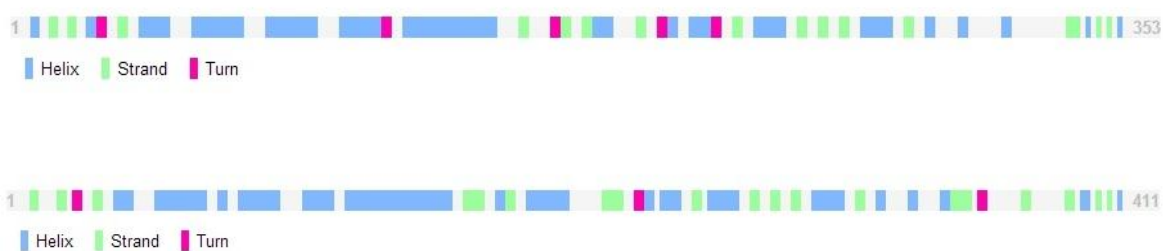
Fig. 3.3 Secondary structures of the MTRPi of *B. subtilis* (**top**) and yeast (**bottom**) based on their X-ray crystal structures.
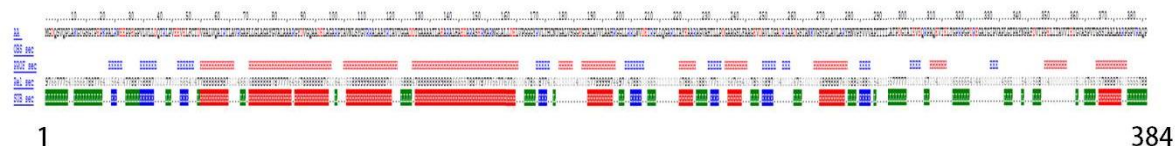
Fig. 3.4 The secondary structure prediction of the 5-FDRPi with PROF.[11]

      Red = α-helix, **Blue** = β-sheet, **Green** = loop, **White** = no prediction available.

It is envisaged that 5-FDRPi will have a similar structure to MTRPi's from *B. subtilis* and yeast, based on their similar secondary structure arrangements.

3.2 Active site mutagenesis

The highly conserved active site residues Cys177 and Asp257 contains the catalytic functional groups for the isomerase reaction. The reaction could take place through a hydride shift or a *cis*-enediol intermediate[4] as previously described. In both scenario Asp257 is predicted to be a general acid/base, first as an acid to protonate the furanose ring oxygen to assist ring-opening, and then as a base to remove a proton from the OH group at C-2 to generate the carbonyl centre. Cys177 is predicted to be deprotonated, and would provide stabilisation to a cationic centre on C-1 during a hydride shift or as a base to assist the formation of the enediol intermediate.
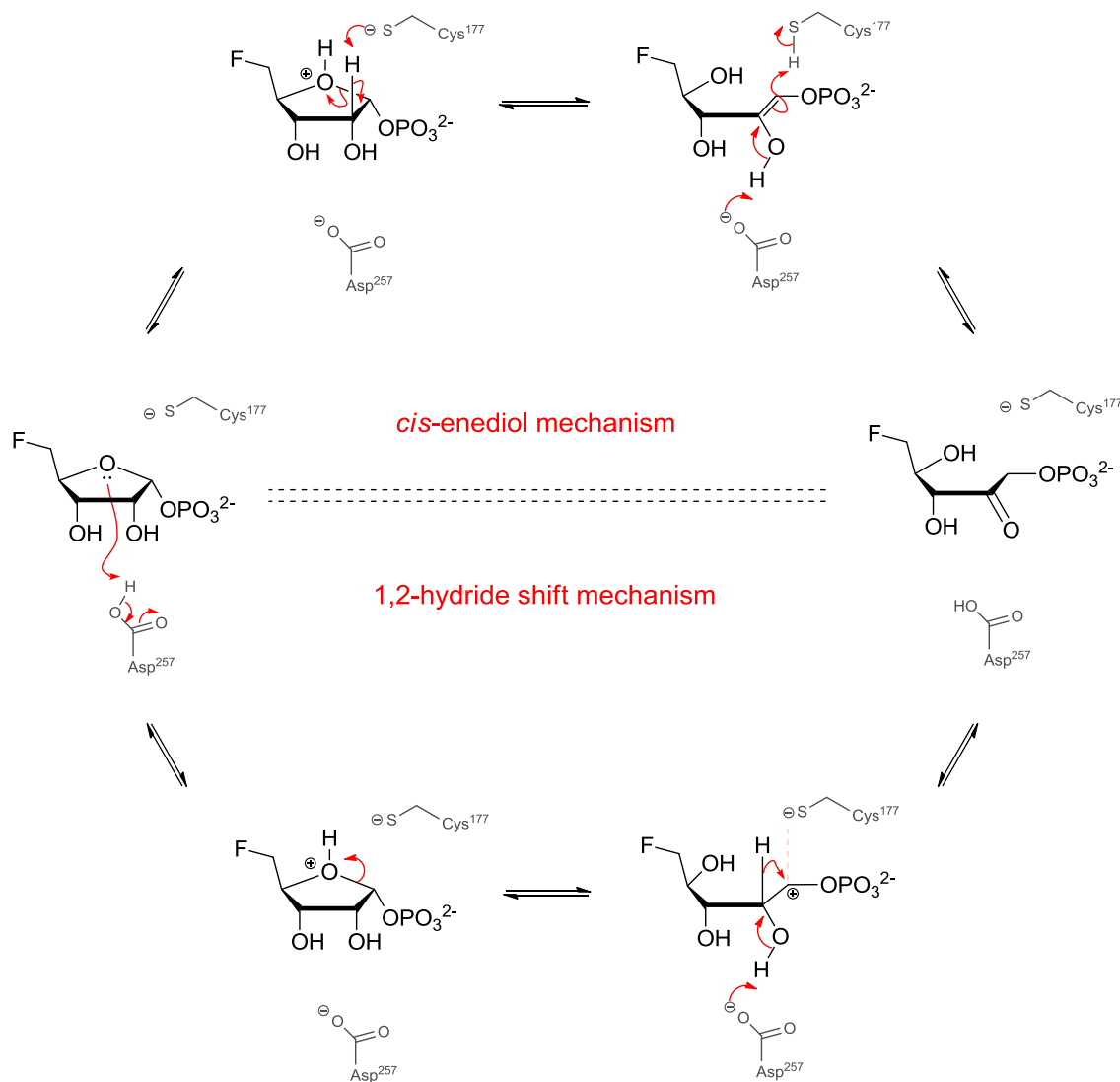
Scheme 3.1. Two putative mechanisms of the isomerase reaction.

In order to confirm the requirement of these residues for the isomerase reactivity, site-directed mutagenesis[12] was carried out to produce 5-FDRPi with Cys177 or Asp257 replaced by another amino acid. These mutants were then assayed for activity.

Cys177 was mutated to alanine as an unreactive functional group replacement. In this case the methylthiol group is replaced by a methyl group. Cys177 was also mutated to a serine residue, where the SH became OH, as a similar functional group replacement. Asp257 was mutated to glycine as an unreactive residue replacement and also to asparagine, where the carboxylate $(-CO_2H)$ is replaced as a primary amide$(-CONH_2)$, while keeping the atom distances similar.

Mutagenic primers were designed with 11-13 matching bases flanking each side of the mismatched mutation codon.[13] A reverse complimentary sequence was used as the reverse primer for PCR (Table 3.1). The 5-FDRPi-T-Vector recombinant plasmid (*ca.* 4 kb) was used as the template.

Table 3.1. PCR primers for mutagenesis cloning.

Primer	Mutation	Primer sequence (5' → 3')
forward	C177A	GGTGCTGACCCAC AGC AACACCGGCGCCC
reverse	C177A	GGGCGCCGGTGTT GCT GTGGGTCAGCACC
forward	C177S	GGTGCTGACCCAC GCC AACACCGGCGCCC
reverse	C177S	GGGCGCCGGTGTT GGC GTGGGTCAGCACC
forward	D257G	TGATCGGCGCC GGC CGGATCGCCGC
reverse	D257G	GCGGCGATCCG GCC GGCGCCGATCAG
forward	D257N	TGATCGGCGCC AAC CGGATCGCCGC
reverse	D257N	CGGCGATCCG GTT GGCGCCGATCAG

The mutagenic primers were purchased (Eurogentec) and the PCR reaction was carried out with the *pFu* polymerase in a *pFu* reaction buffer and 5 % DMSO. The mixture was allowed to complete 16 cycles of denaturation at 95 °C (1 min), annealing at 55 °C (45 s) and extension at 68 °C (4 mins). The PCR products were purified from a gel with DNA gel extraction kit (Roche). They were treated with the *Dpn* I restriction enzyme to digest the parental DNA, which differ from the PCR products by having an *N*-6 methylated guanidine cap (5'-G^{M6}-ATC-3').

The *Dpn* I treated DNA was transformed into *E. coli* DH10B competent cells and allowed to grow on an agar plate containing ampicillin. Colonies were picked and the plasmid was isolated for sequencing, to confirm a correct mutation.

The mutant plasmids were cut with *Nco* I and *Xho* I restriction enzymes and cloned into pEHISTEV using these restriction sites. The pEHISTEV constructs were transformed into *E. coli* BL21 Gold competent cells for protein over-expression.

The over-expression was carried out in LB media (1 L) at 25 °C for 24 h. The cells were induced with IPTG to a final concentration of 1 mM at OD$_{600}$ of 0.5-0.6. The cells were harvested by centrifugation (5000 rpm, 5180 *g*) and lysed in lysis buffer (50 ml) by sonication (Vibra Cells apparatus). The cell lysates were cleared by centrifugation (20000 rpm, 48,384 *g*) and the supernatant were loaded to a Ni-NTA column (3 ml) and the protein was purified by Ni-affinity chromatography.

The eluant was dialysed against 20 mM phosphate buffer pH 7.8 (5 L) at 4 °C, and then concentrated to around 10 mg/ml (Amicon concentrator, 10 kDa MWCO).

The purified proteins were analysed by mass spectrometry by an in-gel tryptic digest method[14] (Table 3.2). The mass spectra of the mutant proteins were identical to the native protein for all the signals expect those corresponding to the fragment containing the mutation. MS-MS was carried out on these signals and the mutation confirmed by the mass changes (Fig 3.5).

Table 3.2.   Site-directed mutagenesis mutants of the 5-FDRPi.

Mutation	trypsin digest fragment containing mutation	calc. MW
native	VLTH**C**NTGALVSGGEGTALAVVLAAHR aa173-199	2617.3878
C177S	VLTH**S**NTGALVSGGEGTALAVVLAAHR aa173-199	2601.4106
C177A	VLTH**A**NTGALVSGGEGTALAVVLAAHR aa173-199	2585.4157
native	AGVAHTLLPDGAAGSLFAAGEVDAVLIGA**D**R aa 228-258	2934.5318
D257N	AGVAHTLLPDGAAGSLFAAGEVDAVLIGA**N**R aa 228-258	2933.5478
D257G	AGVAHTLLPDGAAGSLFAAGEVDAVLIGA**G**R aa 228-258	2876.5264
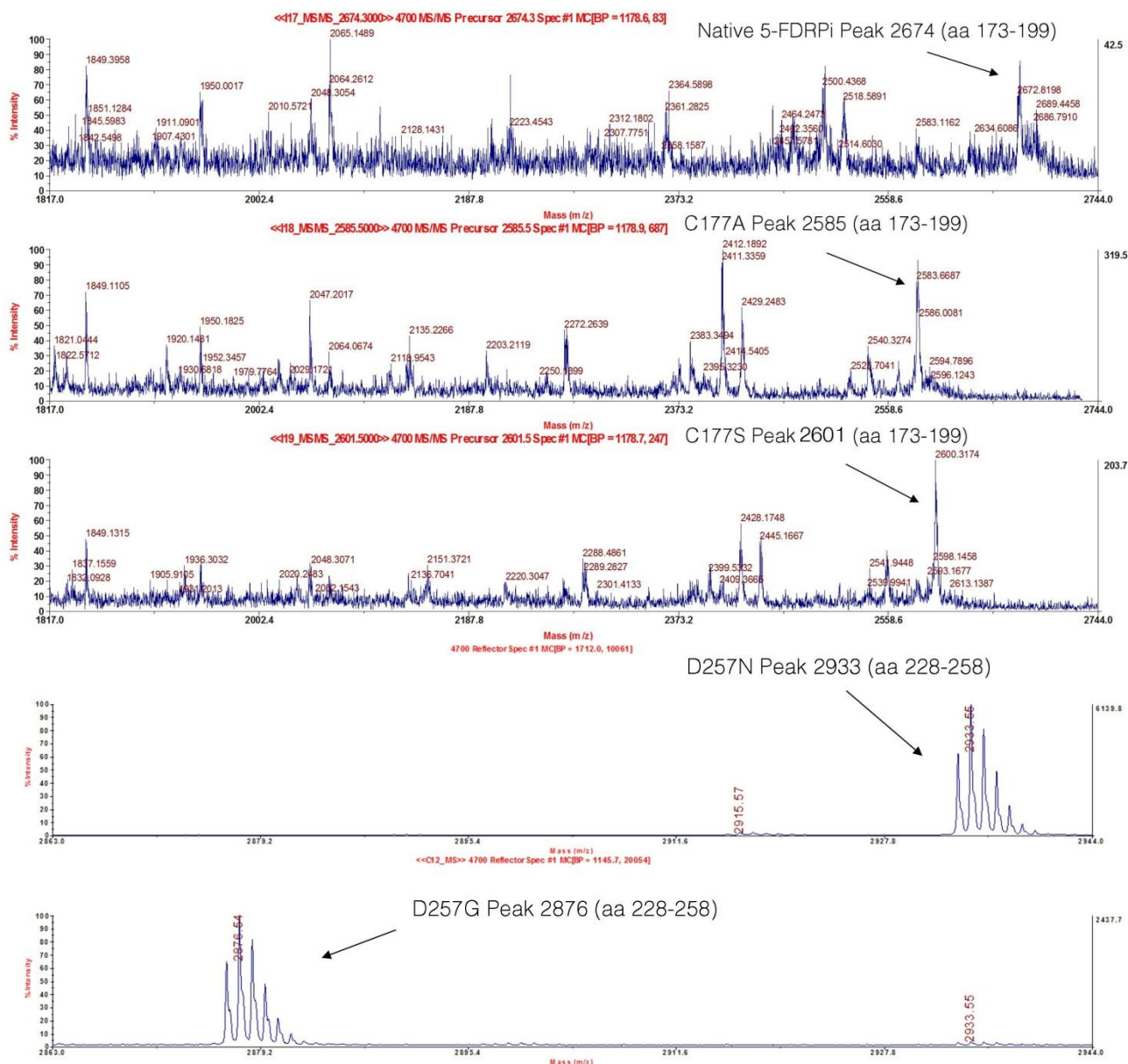
Fig. 3.5 Mass spectra from the tryptic-digest of the isomerase mutation proteins.

The four site-mutated isomerases were assayed for their activity. The purified isomerase mutation enzymes (0.1 mg) were separately incubated with synthetic 5'-FDA solution (20 mM, 7.2 μl), purified PNP enzyme (0.1 mg) in phosphate buffer (20 mM, pH 7.8, 720 μl). The reactions were carried out at 37 °C for 4 h, then D_2O (100 μl) was added to the mixture for NMR signal lock. The mixtures were analysed by $^{19}F\{^1H\}$ NMR. A control experiment was also carried out with native 5-FDRPi.

Only two peaks were observed in the NMR of the mutants. These corresponded to 5-FDRP and 5'-FDA. The third peak for 5-FDRulP was only observed in the control experiment with the wild type protein (*cf.* Fig. 2.9, page 71).

The assay demonstrated that both Cys177 and Asp257 are essential for isomerase activity. The activity could not be restored by replacement with similar amino acids. Deprotonation of Cys177 to a thiolate is predicted to be important to its activity. The pK_a of cysteine thiol is 8.0, however it was suggested that this Cys residue would form a hydrogen bond with an amide group from a neighbouring residue and this interaction would lower the pK_a to around 7,[4] and would be deprotonated at the optimal pH of 8 for the isomerase.[15] The alcohol of serine in C177S has a pK_a of around 15, and is not possible to deprotonate this alcohol under physiological conditions.

Asp257 is predicted to be a proton donor/acceptor. The normal pK_a of the carboxylic acid side chain of Asp is 3.8, but a highly conserved hydrophobic pocket surrounds this amino acid. When Asp is located in a hydrophobic environment, its pK_a has been observed to change to 7.5.[16] This would make it an ideal general acid and general base. Replacement of COOH with COONH$_2$ would retain hydrogen-bonding capabilities but this brings a significant change to the general acid/base property of the residue and as a result, the residue is no longer active.

3.3 The stereochemical course of the isomerase reaction

3.3.1 The study of the specificity of proton transfer in the MTRPi

The 1,2-hydride shift offers a working hypothesis for the mechanism of the isomerase. It is distinguished from the *cis*-enediol mechanism since the C-2 hydrogen is transferred to C-1 and is not exchanged with the solvents.

Imker and co-workers analysed the stereochemical outcome of the MTRPi reaction with D-[1-^{2}H]ribose 1-phosphate **62A** and D-[2-^{2}H]ribose 1-phosphate **62B** as substrates (Scheme 3.2).[17] The ^{1}H-NMR spectra of the reaction product **63A/63B** from the MTRPi reactions were recorded to try to evaluate deuterium incoporation, but the assignment of the signal for the C-1 protons of ribulose 1-phosphate was complicated by the α- and β- anomers of its hemiketal and a congested region where the signals appear (δ_H 3.6-3.8) (Fig 3.6).
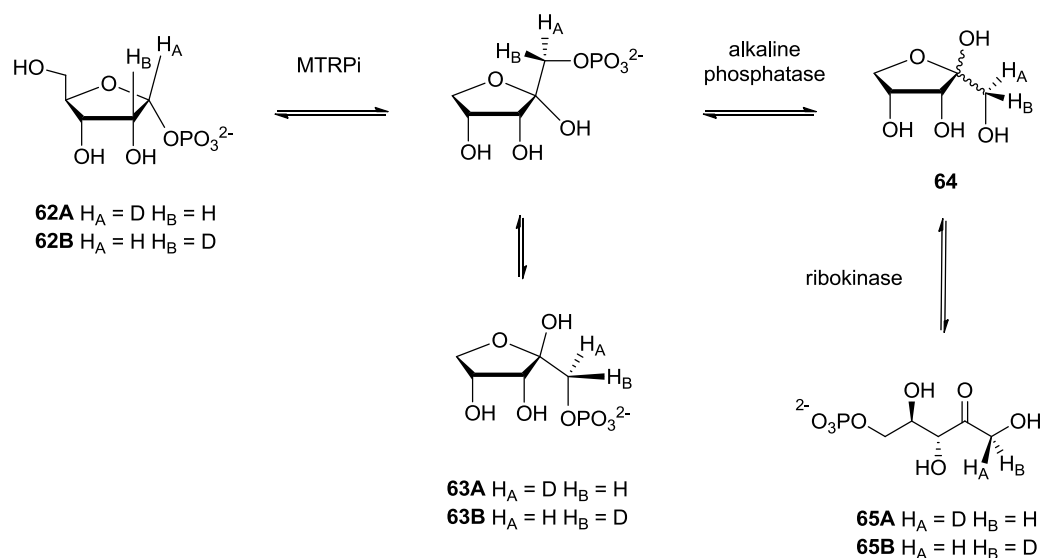
Scheme 3.2. Reaction sequence in the experiment by Imker *et al.*[17]
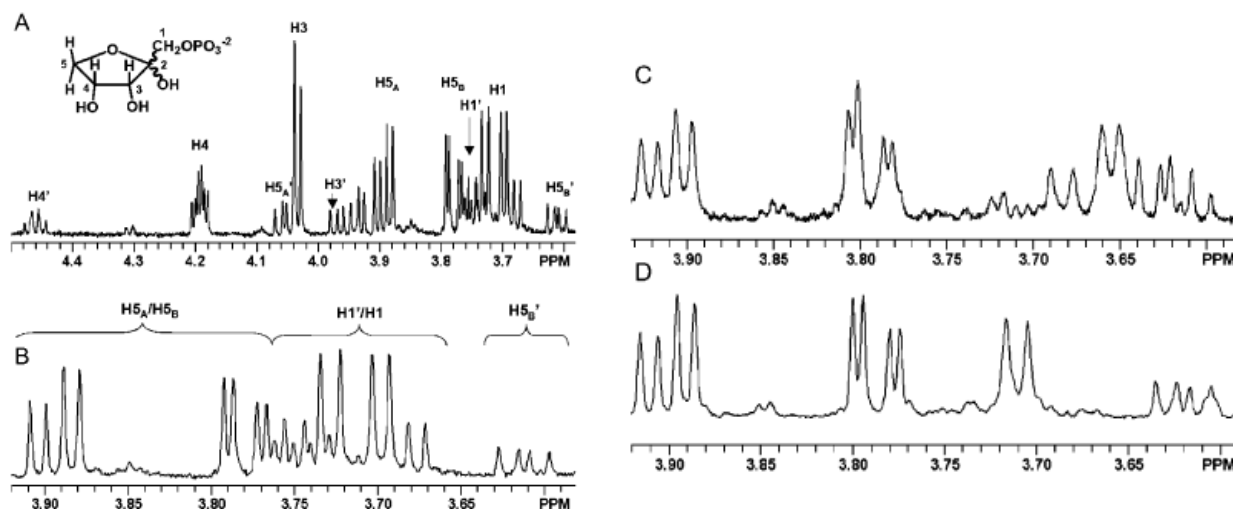
Fig 3.6. **(A)** ^{1}H NMR from unlabelled ribulose 1-phosphate.[17]

Partial spectra (δ_H 3.60 – 3.95) for:

(B) Unlabelled ribulose 1-phosphate

(C) (1S)-D-[1-^{2}H]Ribulose 1-phosphate **63A** from D-[1-^{2}H]ribose 1-phosphate **62A**.

(D) (1R)-D-[1-^{2}H]Ribulose 1-phosphate **63B** from D-[2-^{2}H]ribose 1-phosphate **62B**.

The problem was circumvented by converting ribulose 1-phosphate **64** to ribulose 5-phosphate **65** with the successive treatment of calf intestine phosphatase and D-ribokinase (over-expressed from *E. coli*) (Scheme 3.2). The C-1 pro-*S* and pro-*R* protons could be distinguished in the ^{1}H-NMR,[18] and from these spectra, they assigned the stereochemistry of the MTRPi reaction (Fig 3.7).
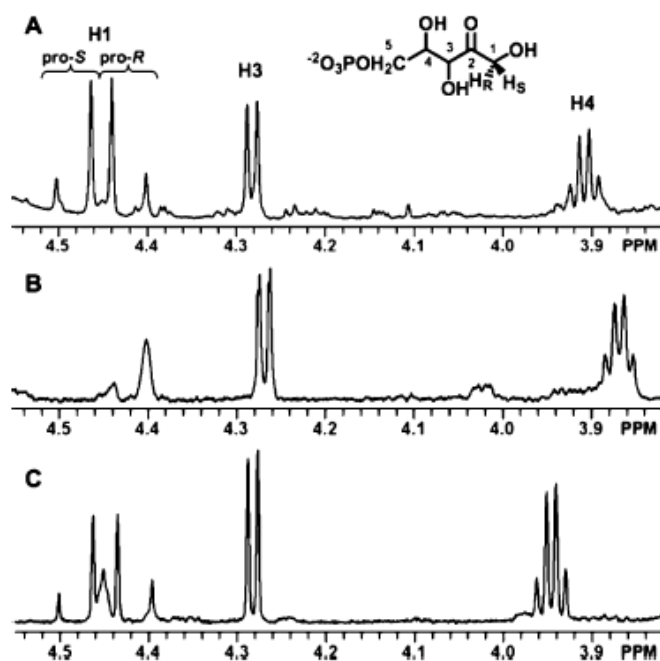
Fig 3.7. Partial ^{1}H NMR for: **(A)** unlabelled ribulose 5-phosphate. **(B)** $(1S)$-D-$[1$-^{2}H]ribulose 5-phosphate **65A** from D-$[1$-^{2}H]ribose 1-phosphate **62A**. **(C)** $(1R)$-D-$[1$-^{2}H]ribulose 5-phosphate **65B** from D-$[2$-^{2}H] ribose 1-phosphate **62B**.

The observation that the C-2 proton of ribose 1-phosphate becomes the C-1 pro-*R* proton of ribulose 1-phosphate in the MTRPi reaction provides strong evidence for a 1,2-hydride shift.

3.3.2 Study of the proton transfer in 5-FDRPi

An isotope-labelling experiment similar to that described by Imker *et al.* was carried out in order to study the hydride migration of 5-FDRPi. Since the PNP from *S. cattleya* is able to transform adenosine to ribose 1-phosphate,[19, 20] it was decided to prepare [2'-^{2}H]adenosine by chemical synthesis, as a labelled substrate for the PNP. This should generate a sample of D-[2-^{2}H]ribose 1-phosphate, a precursor that it is difficult to prepare by synthesis.[21] [2-^{2}H]Ribose 1-phosphate can then act as a substrate for 5-FDRPi and if a 1,2-hydride shift takes place, deuterium would become located at C-1 of the product ribulose 1-phosphate (Scheme 3.3).
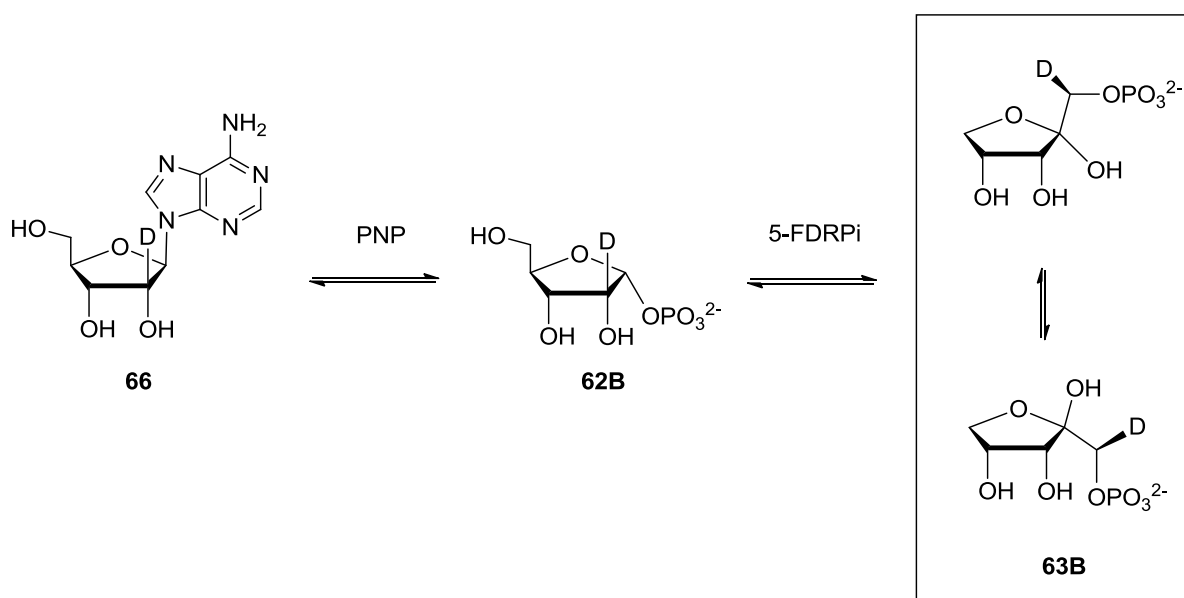
Scheme 3.3. Reaction sequence for deuterium labelling study from [2'-^{2}H]adenosine.

Since the D-ribokinase used by Imker *et al.* to convert ribose 1-phosphate into ribose 5-phosphate was not available to us, it was planned to locate the position of the deuterium by a GC-MS analysis instead of ^{1}H-NMR. Previously, 5-deoxy-5-fluoro-ribulose 1-phosphate was identified by GC-MS based on a mass fragment (m/z = 283) containing C-1[22] (*cf.* Scheme 1.11, page 23). It was envisaged that persilylated **62B** and **63B** would be separable by GC and if deuterium is present on these compounds, M+1 mass signals would be observed from their mass spectra.

3.4 Synthesis of [2'-²H]adenosine

The synthesis of [2'-²H]adenosine **66** has been reported several times in the literature[23, 24, 25] and Robins' synthesis[23] offered a suitable route to this compound.

[2'-²H]Adenosine **66** was prepared in five steps from adenosine **67** (Scheme 3.4). The key step involved the stereo-controlled introduction of deuterium *via* reduction of 2'-ketoadenosine **70**.
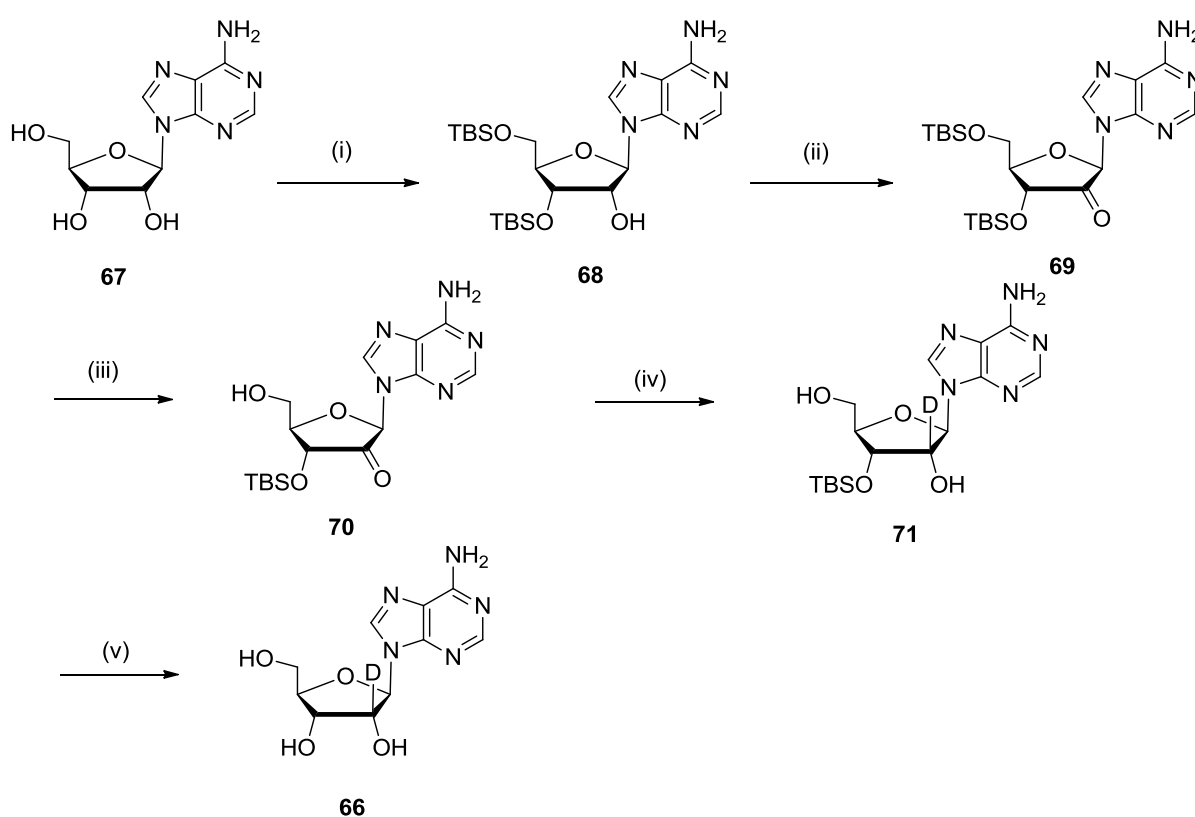
Scheme 3.4. Reagents and conditions: (i) TBSCl (2.2 eq), imidazole, DMF (25 %), (ii) CrO₃, Ac₂O, pyr (75 %), (iii) 90 % TFA (aq), 0 °C, 1 h (96 %), (iv) NaBD(OAc)₃, AcOH, EtOAc (88 %), (v) 1. NH₄F, MeOH, reflux, 2. Ambersep 900 OH (quant.).

The first step involved the protection of the 3' and 5' hydroxyl groups of adenosine **67** as TBS silyl ethers by treatment with TBSCl (2.2 eq.). The partial TBS silylation of adenosine invariably produces a mixture of products. The most reactive alcohol (5'-OH) is first to be silylated, followed by 2'-OH and 3'-OH groups. The conditions reported to give the highest yield of the 3',5'-bis-*O*-silyl ether was adopted.[26, 27] The reaction gave a mixture of four compounds: 5'-mono-*O*-silyl **72**, 2',5'-bis-*O*-silyl **73**, 3',5'-bis-*O*-silyl **68** and 2',3',5'-tris-*O*-silyl **74** adenosines, which were separable by silica gel chromatography. However since the 2',5'-bis-*O*-silyl **73** and 3',5'-bis-*O*-silyl **68** products eluted with some overlap, it was necessary to perform a second chromatography on this mixture in order to obtain a clean sample of **68**. Although the isolated yield was modest (25 %), it was convenient to scale up this first reaction as it uses readily available starting materials.
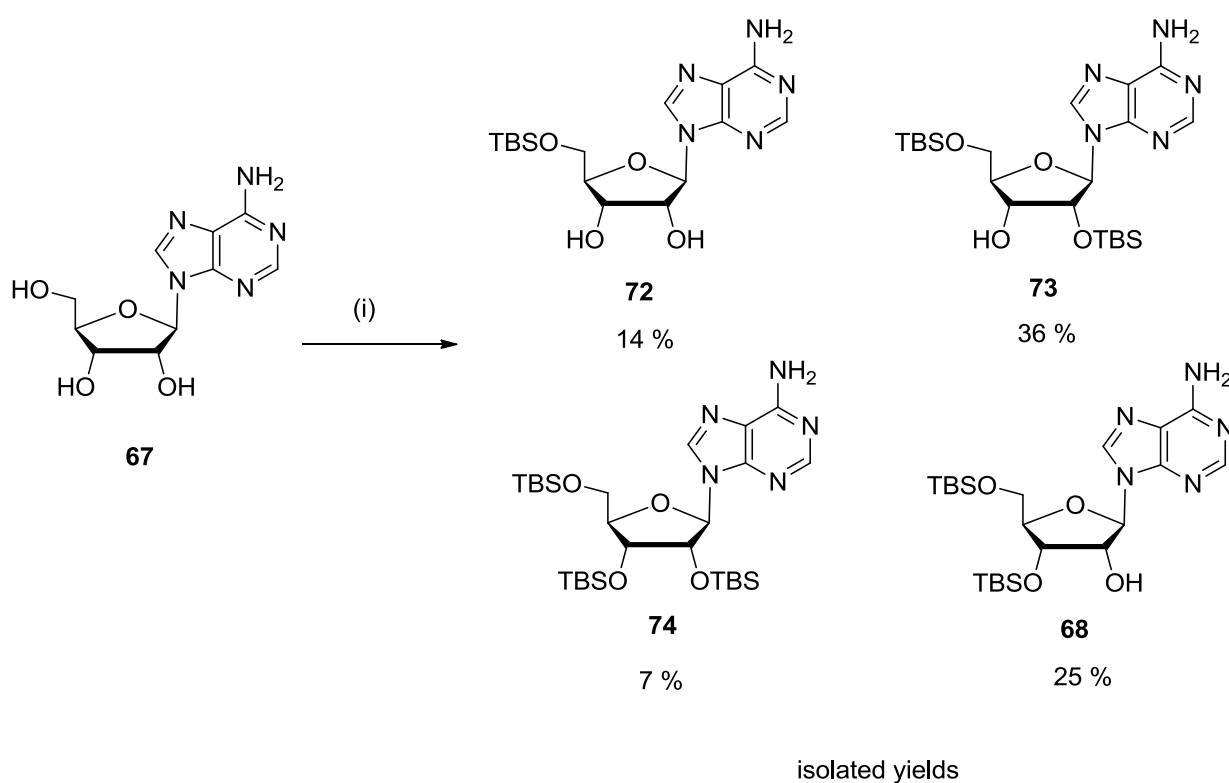
Scheme 3.5. Reagents and conditions: (i) TBSCl (2.2 eq), imidazole, DMF.

Oxidation of alcohol **68** was carried out using Garegg-Samuelsson conditions[28, 29] $(CrO_3/pyridine/Ac_2O)$.
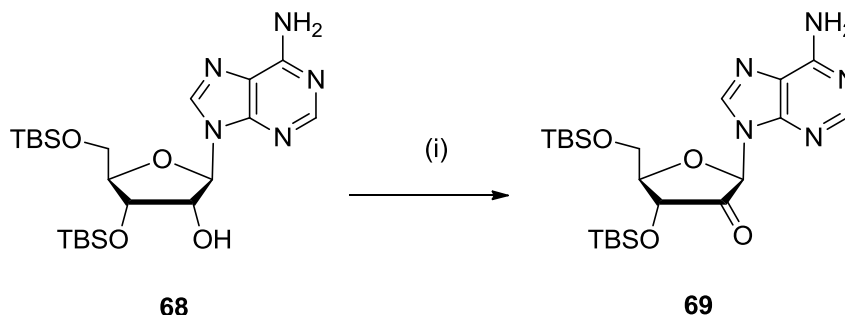
Scheme 3.6. Reagents and conditions: (i) CrO_3, Ac_2O, pyr (75 %)

The transformation of alcohol **68** to ketonucleoside **69** was accomplished in a 75 % yield. The active oxidant in this reaction is a chromium trioxide-pyridine complex and acetic anhydride is suggested to assist in the reduction of Cr (VI) to Cr (IV)[28] (Scheme 7).
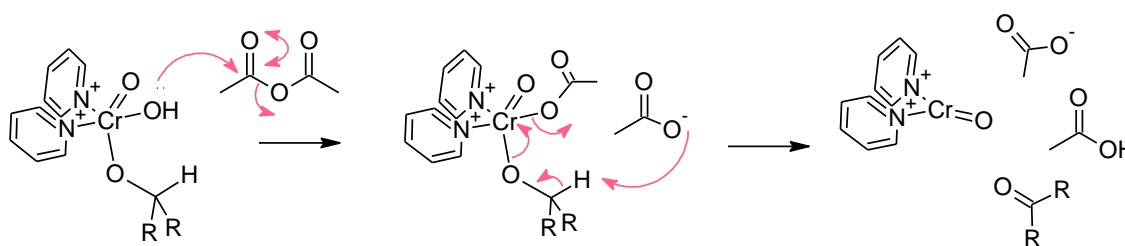
Scheme 3.7. Mechanism of alcohol oxidation by the Garegg-Samuelsson conditions.

Deuterium was then introduced stereoselectively to the 2'-ketone **69** in a two-step protocol involving reduction with a source of deuteride. $NaBD_4$ and $LiAlD_4$ preferentially attacked the bottom face of 2'-ketonucleoside **69**. This gave the undesired *arabino*-isomer as the major product.[29, 30] To override this selectivity, a coordinating reducing agent $NaBD(OAc)_3$ was employed as it anchors to the 5'-OH group to delivery deuteride from the top face (Fig. 3.8).
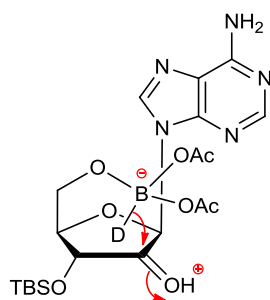
Fig 3.8. Triacetoxyborodeuteride reduction directs the deuteride to the top face of the ketone.

To provide the free 5' hydroxyl group for coordination of the reducing agent, the 5'-OTBS group was selectively hydrolysed by a short treatment of **69** with 90 % TFA at 0 °C. Under these conditions the 3'-OTBS group was not affected and gave **70** in 96 % yield. Treatment of **70** with $NaBD(OAc)_3$ gave **71** with excellent selectivity. The formation of *arabino*-product was not detected.
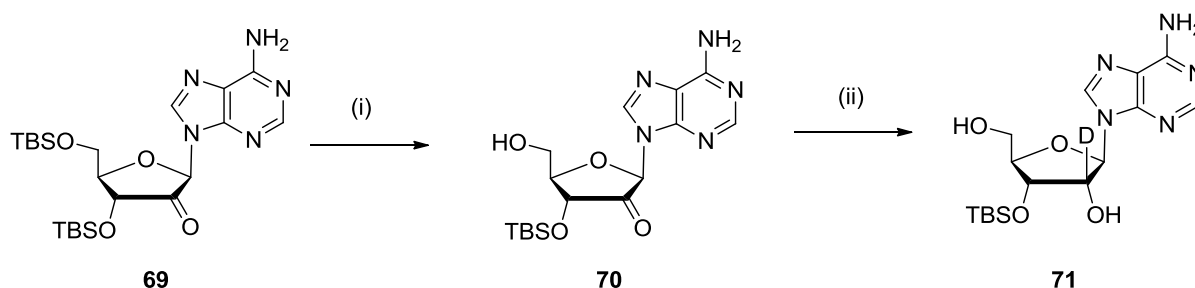
Scheme 3.8. Reagents and conditions: (i) 90 % TFA (aq), 0 °C, 1h (96 %) (ii) $NaBD(OAc)_3$, AcOH, EtOAc (88 %).

Finally the 3'-OTBS group on **71** was removed with ammonium fluoride in methanol under reflux. [2'-^{2}H]Adenosine **66** was extracted into the aqueous layer together with some excess ammonium fluoride. A scavenger resin (Ambersep 900 OH) was used to exchange excess fluoride with hydroxide ions and then the aqueous layer was lyophilised to give **66** as an off-white powder. The mass spectrum of [2-^{2}H]adenosine showed a clear M+1 ion compared with unlabelled adenosine as shown in Fig. 3.9. The ^{13}C-NMR spectrum of **66** was identical to pure adenosine **67**, except the signal for the 2'-carbon became a 1:1:1 triplet due to deuterium ($I = 1$) coupling to carbon (Fig. 3.10). There is the presence of unlabelled adenosine in the sample (2 % by NMR) as a minor product consistent with the 98 % deuteration level of NaBD$_4$ reagent used in the synthesis.
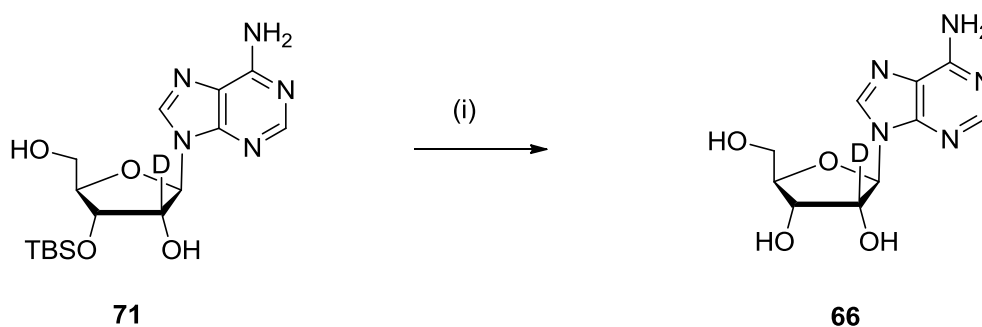
Scheme 3.9. Reagents and conditions: (i) 1. NH$_4$F/MeOH, reflux , 2. Ambersep 900 OH resin (quant.).
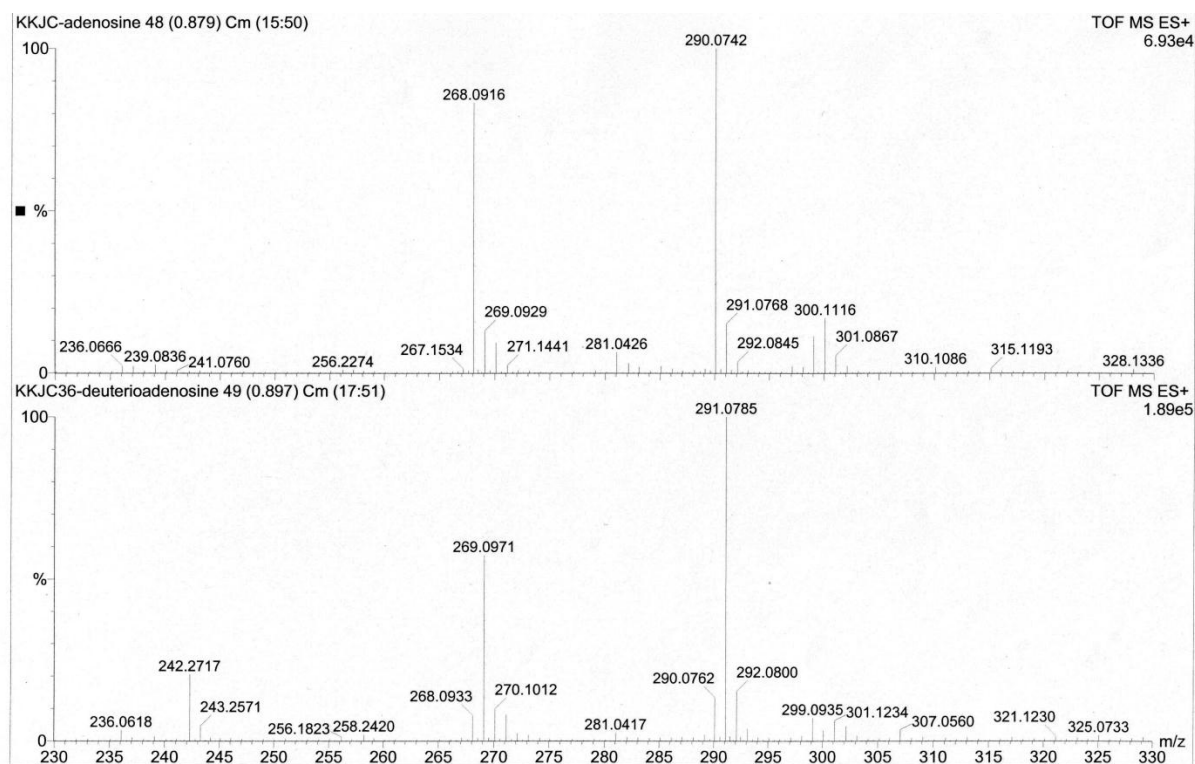
Fig. 3.9. ESI Mass spectra of normal adenosine **67** (**top**) and [2'-^{2}H]adenosine **66** (**bottom**), showing the [M+H]$^+$ (m/z = 268 and 269) and [M+Na]$^+$ (m/z = 290 and 291).
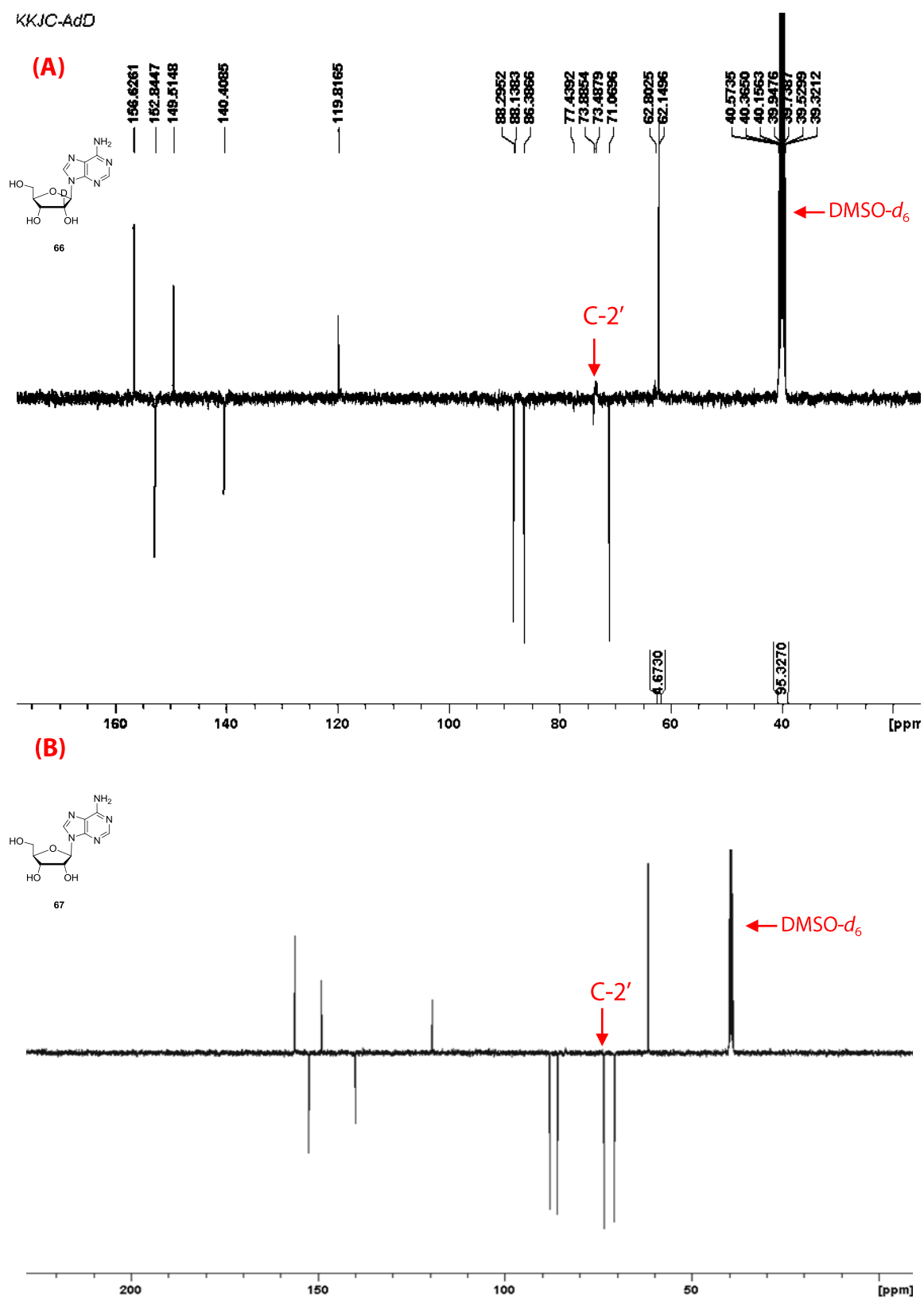
Fig 3.10. (A) ^{13}C NMR spectrum of synthetic [2'-^{2}H]adenosine **66**. (B) ^{13}C NMR spectrum of adenosine **67**.

3.5 Enzyme reactions with [2'-^{2}H]adenosine

Having prepared deuterium labelled substrate [2'-^{2}H]adenosine **66**, the enzymatic assay was then carried out.

PNP (*flB* from *S. cattleya*) was over-expressed in *E. coli* as a fusion protein with maltose binding protein (MBP) to improve its solubility.[31] Adenosine is not the natural substrate for PNP of the fluorometabolite pathway, which normally acts on 5'-FDA instead. It was envisaged that the PNP would accept **66** as surrogate substrate with only a minor perturbation due to a small secondary kinetic isotope effect.

The isomerase 5-FDRPi is the next enzyme on the fluorometabolite pathway. In this labelling experiment, it receives ribose 1-phosphate generated from the PNP reaction *in situ* as its substrate. When 5-FDRPi is included with PNP in the enzyme reaction mixture, D-ribose 1-phosphate is converted to D-ribulose 1-phosphate. If the deuterium is transferred by a specific 1,2-hydride migration, D-[1-^{2}H]ribulose 1-phosphate **63B** will be generated. Alternatively if the *cis*-enediol mechanism takes place, the deuterium label would be lost to the solvent (Scheme 3.1).

It is assumed that 5-FDRPi is able to accept the hydroxydefluoro substrate **62B**, since MTRPi is known to accept ribose 1-phosphate **67**.[17]

Incubation reactions were carried out with synthetic sample of [2'-^{2}H]adenosine **66** (50 mM final concentration) with PNP (0.1 mg) alone, or together with 5-FDRPi (0.1 mg), in phosphate buffer (pH 7.8, 10 mM, 5 ml) at 37 °C for 6 h. Unlabelled adenosine was used in a parallel experiment as a control.
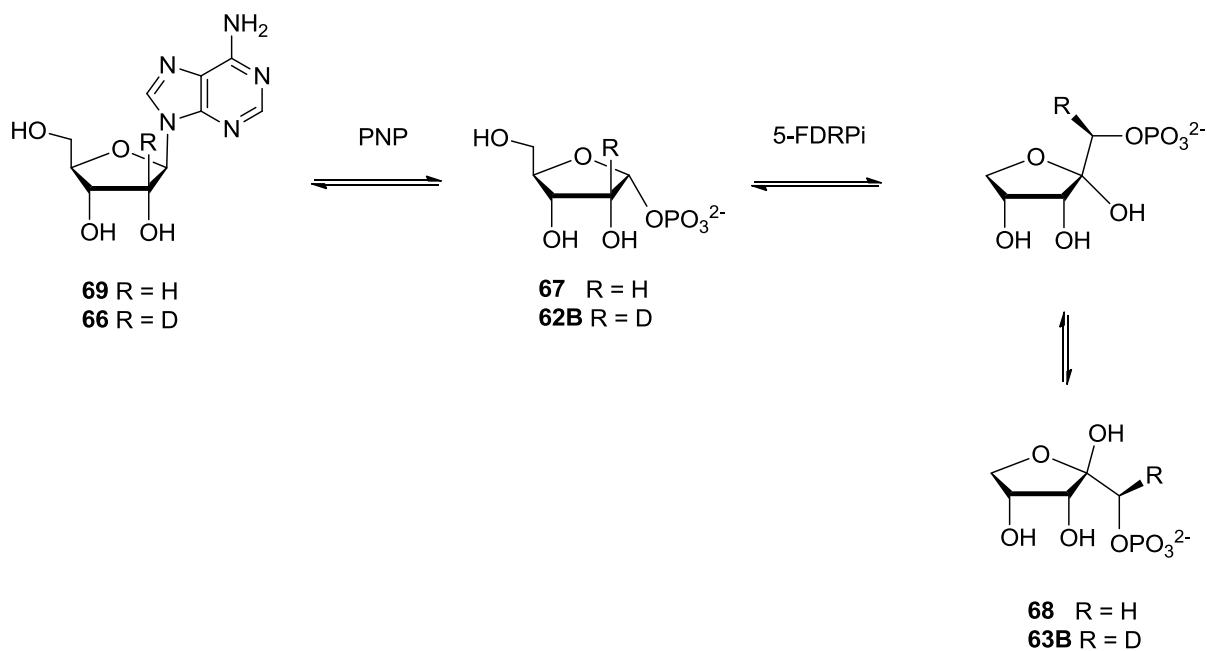
Scheme 3.10. Enzymatic conversion of [2'-^{2}H]adenosine to [2-^{2}H]ribose 1-phosphate and [1-^{2}H]ribulose 1-phosphate.

Analysis of the mixture to locate the deuterium label on the ribulose 1-phosphate formed was attempted by different methods. Despite different attempts, a conclusive result could not been reached. Details of the analytical methods used are described below.

3.6 Analysis of the deuterium labelling experiment

3.6.1 GC-MS analysis of the deuterium labelled products

In a previous study of the fluorometabolite products, a GC-MS method carried out by Dr John G. Hamilton (Queen's University of Belfast) was successful in separation of the MSTFA derivatised 5-deoxy-5-fluororibose 1-phosphate and 5-deoxy-5-fluororibulose 1-phosphate.[32] In that experiment, the two isomeric sugars were separated in GC with retention times of 9.7 mins and 10.8 mins. It was envisaged that this technique could also be applied to the analysis of the deuterium labelled ribose and ribulose phosphates.

To prepare the samples for GC-MS analysis, the reaction mixtures were filtered through protein concentrators (10 kDa MWCO) to remove PNP and 5-FDRPi proteins. The filtrates were freeze-dried and the residues were treated with neat MSFTA (1 ml) at 100 °C for 1 h. These samples were then sent for analysis at Queen's University of Belfast. In this experiment, adenosine **75/76** was detected with a retention time of 17.5 mins for both the labelled and unlabelled samples. The sugar phosphates **77/79** from the unlabelled adenosine reaction could be observed at retention times of 12.8 and 13.6 mins as two major peaks, however the sugar phosphates **78/80** could not be found in the labelled adenosine incubation experiment (Fig. 3.12).
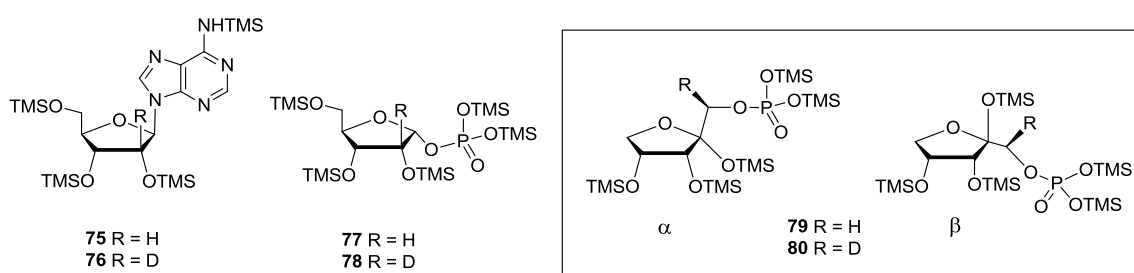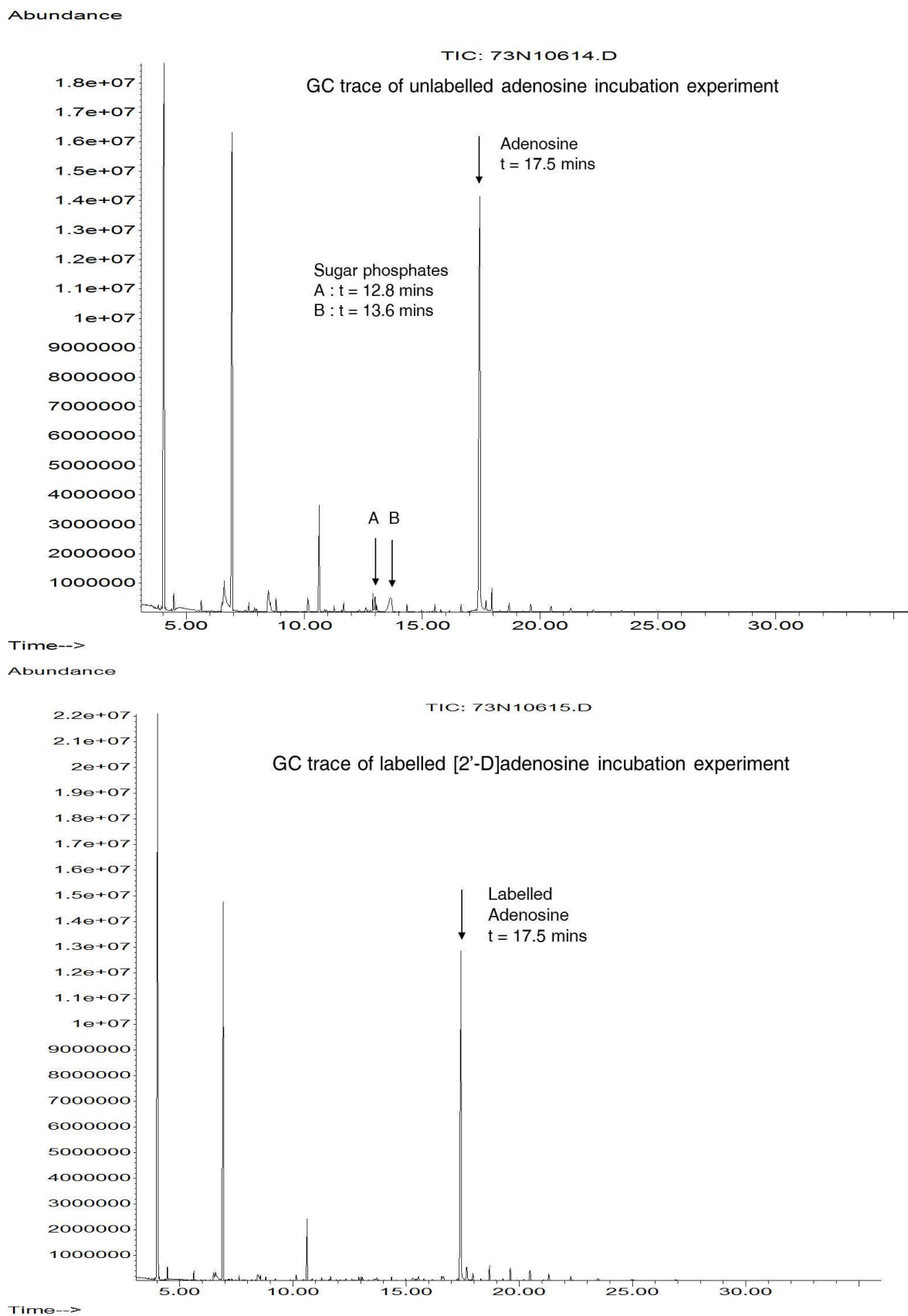

Fig 3.11. Persilylated enzyme reaction products.

Fig 3.12. (Top) GC-trace of unlabelled adenosine incubation experiment. **(Bottom)** GC-trace of labelled $[2'-^2H]$adenosine incubation experiment.

The mass spectrum (CIMS) of the peaks at retention times 12.8 and 13.6 mins from the unlabelled control experiment confirmed the identity of persilylated ribose/ribulose phosphates **77/79/81** (MW = 590) with $m/z = 619$ (M+29) and 631 (M+41), corresponding to $[M+C_2H_5]^+$ and $[M+C_3H_5]^+$ (Fig. 3.13).
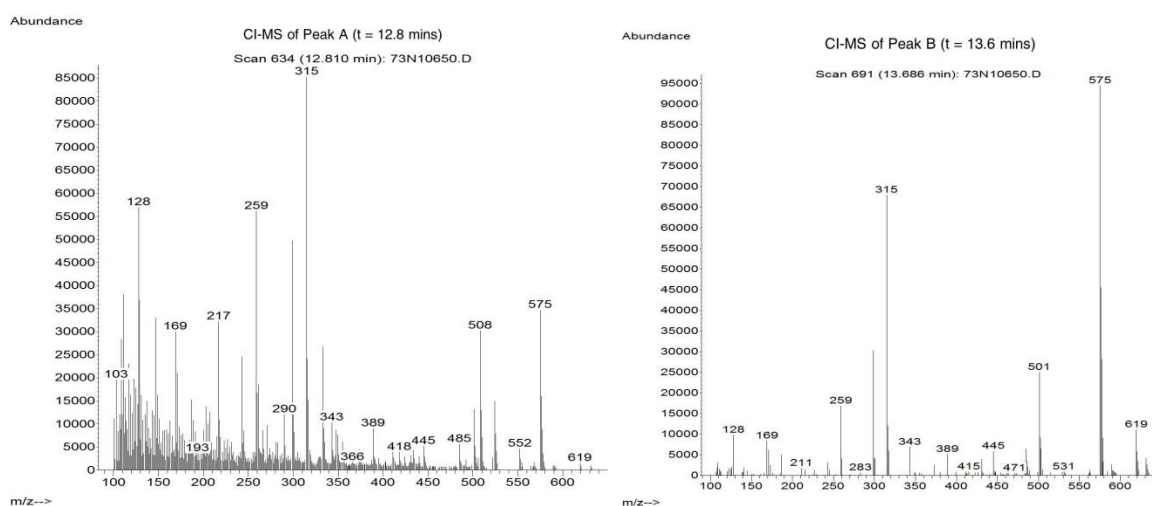

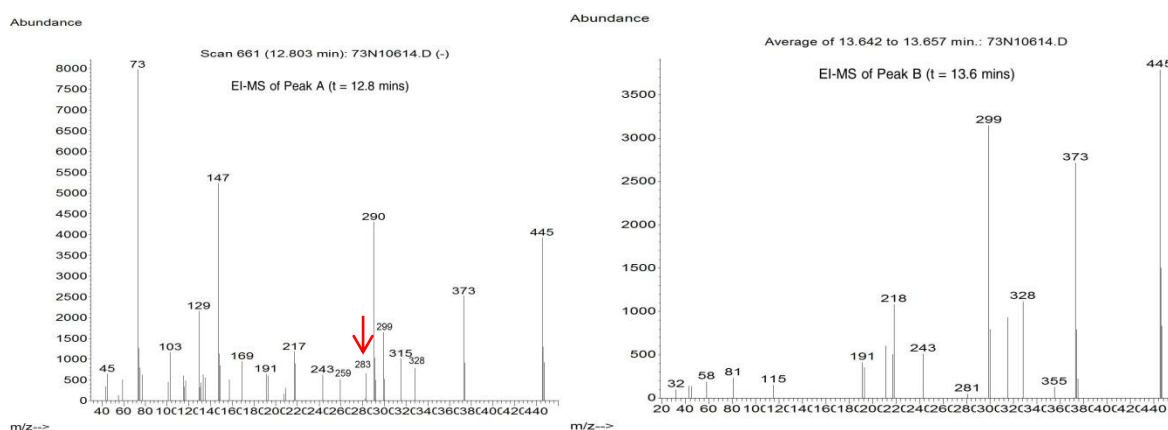**Fig 3.13.** CIMS spectra from unlabelled adenosine experiment. **Left:** peak A at t=12.8 mins

   Right: peak B at t=13.6 mins

77

persilylated ribose 1-phosphate
m/z = 509

81

+

79

persilylated ribulose 1-phosphate
m/z = 509
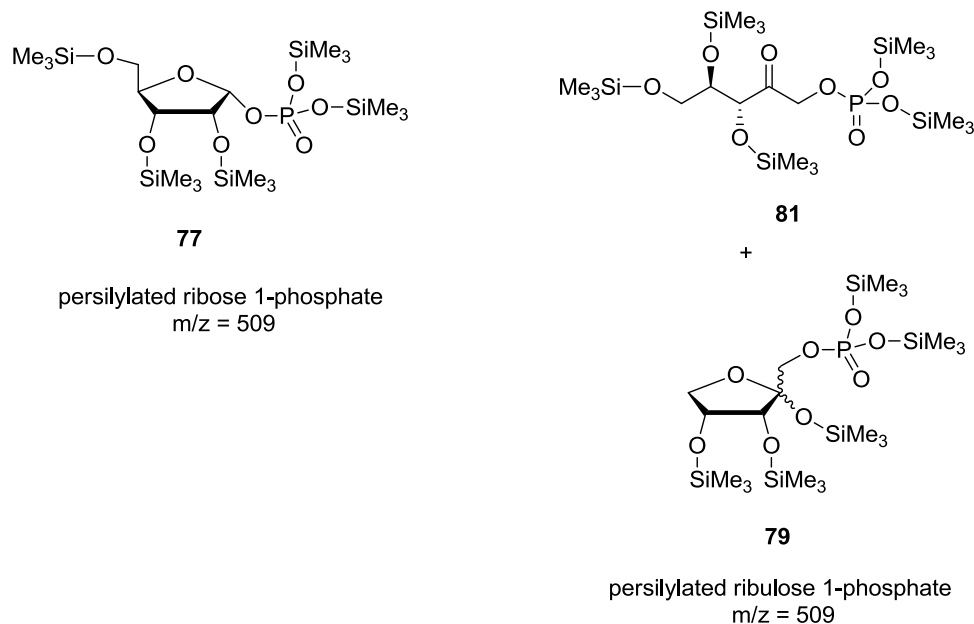
Fig 3.14. Possible structures giving rise to molecular ion peaks of 619 (M+29) and 631 (M+41).

The analysis of these GC signals at retention times 12.8 and 13.6 mins from the unlabelled control experiment by EIMS gave mass ion fragments that could be rationalised with both the structure of ribose 1-phosphate or ribulose 1-phosphate (Fig. 3.15, 3.16). The mass fragment of $m/z = 283$ was found as a minor signal in peak A (retention time = 12.8 mins) but it is unclear whether it had originated from fragmentation of the bond between C-2 and C-3 of chain form ribulose 1-phosphate, as such a mass could also be calculated for the structures shown in the red box in Fig. 3.16.

Fig 3.15. EIMS spectra from unlabelled adenosine experiment. **Left:** peak A at t=12.8 mins

   Right: peak B at t=13.6 mins
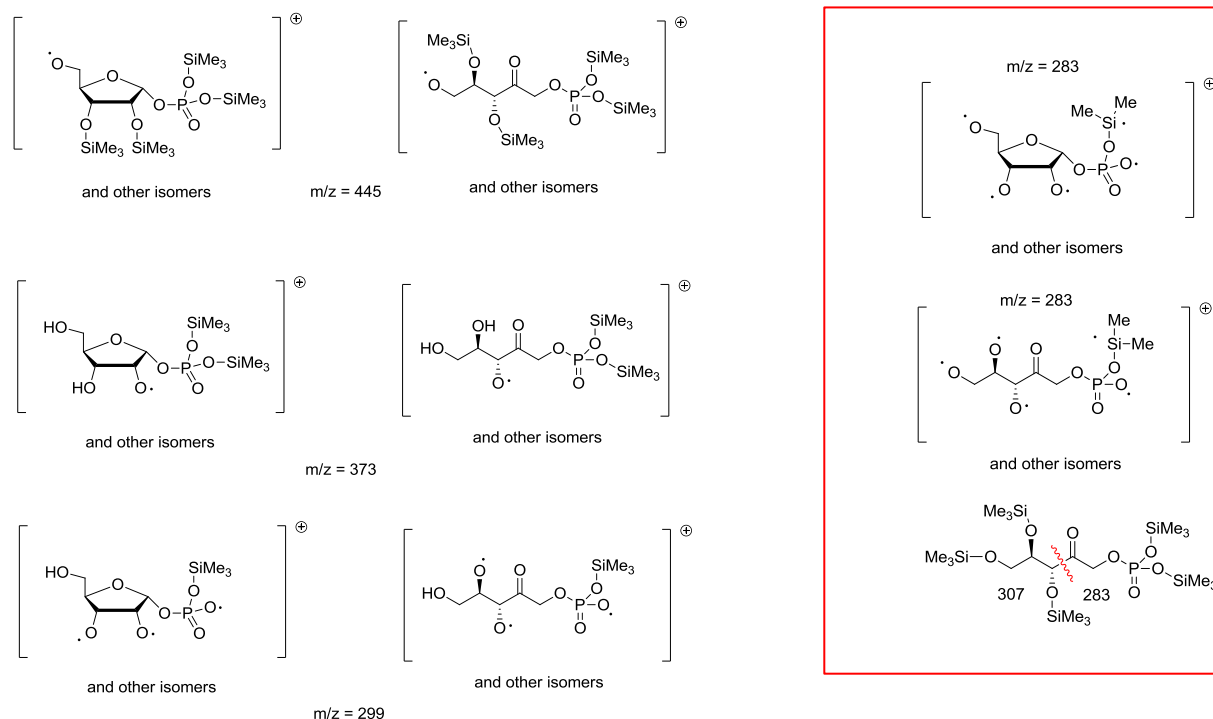
Fig 3.16. Possible fragmentation patterns to give rise to the major peaks in EIMS.

Analysis of the experiment using [2'-^{2}H]adenosine did not result in any identifiable sugar phosphate products. Since both the reaction using labelled or unlabelled adenosines were carried out under identical conditions, it is unclear why the sugar phosphates could not be found in the deuterium labelled experiment. A secondary kinetic isotope effect may reduce the rate at which 2'-deuteroadenosine was being turned over by PNP.

The GC-MS method is limited by its ability to distinguish the chemical identities of the two isomeric sugar phosphates as they have the same mass. The equilibrium between the furanose ring and linear open chain forms of ribulose 1-phosphate also further complicates structure assignment based on fragmentation patterns in EI mass spectrometry alone.

3.6.2 ^{13}C-NMR analysis of the deuterium labelled products

The ^{13}C NMR signal of a carbon with a deuterium attached is split into a triplet (^{13}C I = ½, ^{2}H I = 1) and usually diminishes into noise. It was envisaged that the deuterium label could be located by comparing the ^{13}C NMR spectra of the reaction products from both the labelled or unlabelled experiments.

However since ^{13}C only has a natural abundance of 1 %, a relatively large quantity of substrate (eg. 30 mg adenosine) and enzymes (eg. 30 mg) was required for each reaction to ensure sufficient products for detection by ^{13}C NMR.

Solutions of both adenosine **67** and [2'-^{2}H]adenosine **66** (30 mg each) in phosphate buffer (50 ml, pH 7.8) were added to PNP and isomerase (30 mg each) enzymes. After incubation at 37 °C for 6 hours, the mixtures were filtered through protein concentrators (10 kDa MWCO) to remove the enzymes and the filtrates were freeze-dried. The residues were extracted with dry d_6-DMSO, in which the inorganic phosphates were insoluble, and the DMSO solution was analysed by NMR.

However the sugar phosphates were not observed in the NMR spectrum suggesting that they may be insoluble in DMSO due to the charged phosphate groups.

3.6.3 Analysis of the deuterium labelled products by deoxofluorination and ^{19}F{^{1}H} NMR

Fluoromethyl groups ($-CH_2F$) give a ^{19}F-NMR signal in the -220 to -240 ppm region.[34] Replacement of one of the hydrogen atoms with deuterium on a fluoromethyl group ($-CHDF$) has been observed to cause the ^{19}F chemical shift to move upfield by approximately 0.5 ppm when compared with a CH_2F group in the same chemical environment. This has been used as a tool to observe the presence of deuterium on a fluorinated substrate in enzyme reactions.[35, 36] Therefore an experiment was devised to assess whether the deuterium label was present on C-1 of ribulose 1-phosphate by converting the C-1 of the sugar phosphate into a fluoromethyl group and then studying the fluorinated substrate by ^{19}F{^{1}H} NMR.

Treatment of ribulose 1-phosphate with a phosphatase would give the free alcohol. Deoxofluor^{TM} **82** treatment of this product would fluorinate the primary alcohol to generate a fluoromethyl group (Scheme 3.12).
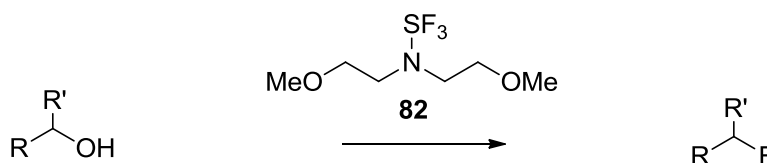
Scheme 3.11. General dehydroxyfluorination of an alcohol by Deoxofluor^{TM} **82**.
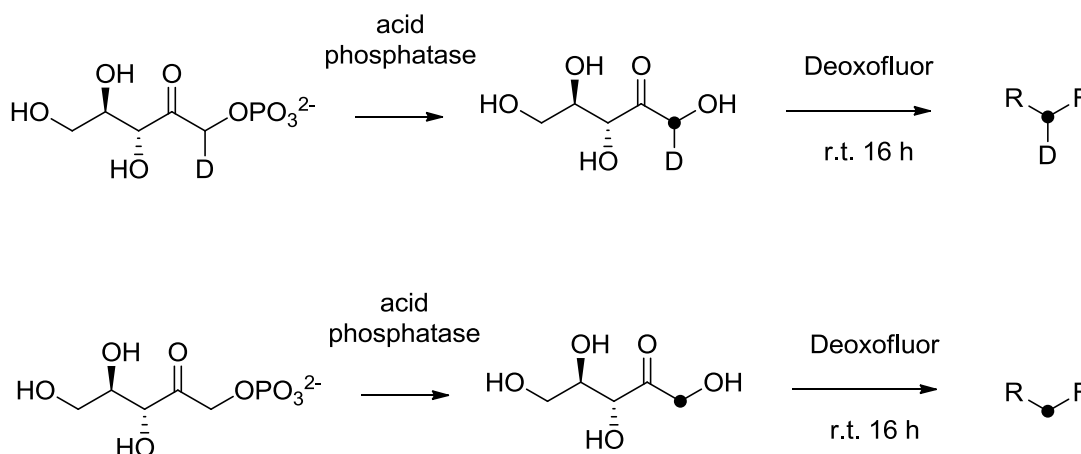
Scheme 3.12. Enzymatic dephosphorylation of ribulose 1-phosphate and fluorination to generate a fluoromethyl group.

Both unlabelled adenosine and [2'-^{2}H]adenosine were treated with PNP alone, or with PNP and isomerase. Then the enzymes were denatured by briefly heating to 100 °C and the precipitated enzymes were removed by centrifugation. The pH of the solution was adjusted to 5.6, optimal for acid phosphatase. The acid phosphatase (Sigma Aldrich) was added and was allowed to react for 2 h at 37 °C. Then the enzymes were again denatured and removed. The supernatant was lyophilised, and the dry residues were treated with an excess of Deoxofluor^{TM} in THF (50 % solution).

After the fluorination reaction, the reaction was worked up with sat. NaHCO$_3$ solution until netural, and extracted into CDCl$_3$. The CDCl$_3$ extract was separated and dried (MgSO$_4$) and analysed by ^{19}F$\{^1$H$\}$ NMR.

The results were inconclusive. The fluorine signals for the fluoromethyl group region (-220 to -240 ppm) were almost identical when comparing the labelled and unlabelled experiments. Changes in chemical shifts greater than 0.2 ppm was not observed which suggested that-CHDF signals were not present as products in these reactions (Fig. 3.17).
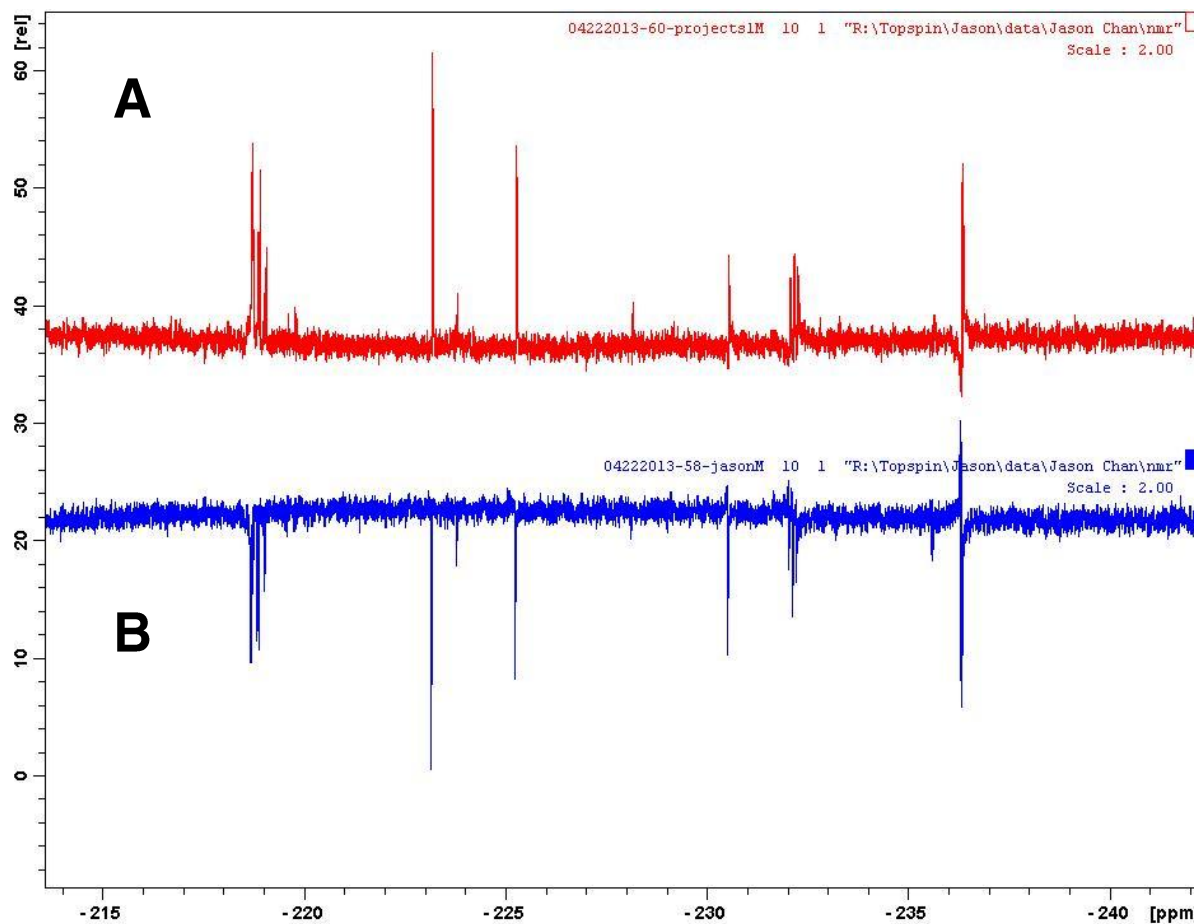
Fig 3.17. $^{19}F\{^1H\}$ NMR of fluorination on the lyophilised enzymatic reaction mixture for the region of fluoromethyl groups from -215 to -240 ppm. (**A**) The top spectra had unlabelled adenosine as substrate and (**B**) the bottom spectra had $[2'\text{-}^2H]$adenosine as a substrate.

To test the working hypothesis of hydride migration from C-2 of ribose 1-phosphate to C-1 of ribulose 1-phosphate against the *cis*-enediol mechanism, the reaction using unlabelled adenosine was repeated with normal water or with 50 % D_2O in H_2O as solvent. If the hydrogen transfer was not specific, deuterium from the solvent could be exchanged with the product and a CHDF group could be formed.

Once again, there was no observable change to the NMR shifts between the experiment conducted in H_2O or in 50 % D_2O in H_2O (Fig. 3.18).

It is unclear whether the C-1 of ribulose, if present in the mixture had been converted into any fluoromethyl groups.
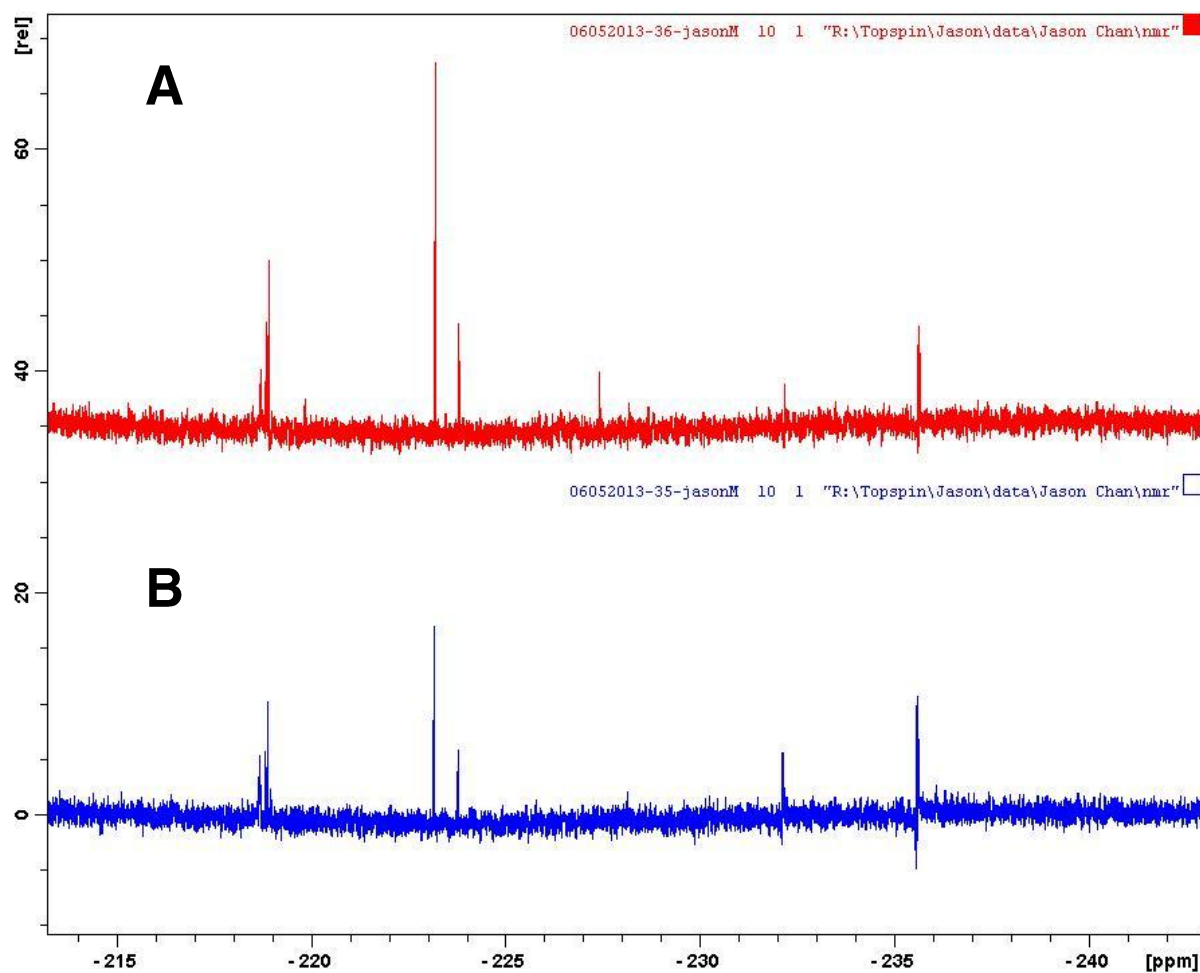

Fig 3.18. $^{19}F\{^{1}H\}$ NMR of fluorination on the lyophilised enzymatic reaction mixture for the region of fluoromethyl groups from -215 to -240 ppm. Both reactions used unlabelled adenosine but the reaction was carried out in (**A**) H_2O or in (**B**) 50 % D_2O in H_2O.

3.7 Conclusions

Structural analysis of 5-FDRPi from *S. cattleya* with the linear sequences of MTRPi from other species revealed they are highly conserved. In particular, the two catalytic residues Cys177 and Asp257 are always conserved. The predicted secondary structure of 5-FDRPi also resembles the crystallography-determined secondary structures of MTRPi from *B. subtilis* and yeast closely. Based on these similarities, it is reasonable to assume that 5-FDRPi is an enzyme that resembles MTRPi in its overall protein folding and the arrangement of its active site.

The isomerase becomes catalytically inactive when Cys177 or Asp257 are mutated. This observation further attests to their roles in catalysis and the need for these residues to remain highly conserved across MTRP isomerases.

A 2'-deuterium labelled adenosine **66** was prepared by chemical synthesis for monitoring the stereochemical course of the isomerase reaction, since a specific migration of deuterium will lend support to the working hypothesis of the hydride shift mechanism. Different analytical techniques have been explored but in the end, the analyses did not provide sufficient data to identify the deuterium label after enzyme reactions. A possible explanation is that PNP and/or 5-FDRPi turn over unlabelled adenosine easier than [2'-^{2}H]adenosine, due to the kinetic isotope effect.

3.8 Chapter 3 References

1. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* 1990, **215**, 403-410.

2. W. Gish, D. J. States, *Nature Genet.,* 1993, **3**, 266-272.

3. K. M. Flowers, S. R. Kimball, R. C. Feldhoff, A. G. Hinnebusch, L. S. Jefferson, *Proc. Natl. Acad. Sci. USA,* 1995, **10**, 4274-4278.

4. H. Tamura, Y. Saito, H. Ashida, T. Inoue, Y. Kai, A. Yokoya, H. Matsumura, *Protein Sci.,* 2008, **17**, 126-135.

5. F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C.V. Brushi, A. Danchin, *Nature,* 1997, **390**, 249-256.

6. M. Fischbach, P. Godfrey, D. Ward, S. Young, C. D. Kodira, Q. Zeng, M. Koehrsen, L. Alvarado, A. M. Berlin, D. Borenstein, Z. Chen, R. Engels, E. Freedman, M. Gellesch, J. Goldberg, A. Griggs, S. Gujja, D. I. Heiman, B. Birren, 2009, *Genome Sequence submitted to the EMBL/GenBank/DDBJ databases.*

7. H. Bussey, R. K. Storms, A. Ahmed, K. Albermann, E. Allen, W. Ansorge, R. Araujo A. Aparicio, B. G. Barrell, K. Badcock, V. Benes, D. Botstein, S. Bowman, M. Brueckner, J. Carpenter, J. M. Cherry, E. Chung, C. M. Churcher, J. Hani, *Nature,* 1997, **387**, 103-105.

8. F. Sievers, A. Wilm, D. G. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Mol. Syst. Biol.,* 2011, **7**, 539.

9. M. Bumann, S. Djafarzadeh, A. E. Oberholzer, P. Bigler, M. Altmann, H. Trachsel, U. Baumann, *J. Biol. Chem.,* 2004, **279**, 37087-37094.

10. http://www.clustal.org, University College Dublin.

11. B. Rost, C. Sander, *J. Mol. Biol.*, 1993, **232**, 584-599.

12. P. Carter, *J. Biochem.*, 1986, **237**, 1-7.

13. H. Liu, J. H. Naismith, *BMC Biotechnology*, 2008, **8**, 91.

14. A. Shevchenko, H. Tomas, J. Havli, J. V. Olsen, M. Mann, *Nature Protocols*, 2007, 2856-2860.

15. T. Tamura, M. Wada, N. Esaki, K. Soda, *J. Bacteriol.*, 1995, **177**, 2265.

16. K. Langsetmo, J. A. Fuchs, C. Woodward, *Biochemistry*, 1991, **30**, 7603-7609.

17. H.J. Imker, A.A. Fedorov, E.V. Fedorov, S.C. Almo, J.A. Gerlt, *Biochemistry*, 2007, **46**, 4077-4089.

18. W. S. Yew, J. Akana, E. L. Wise, I. Rayment, J. A. Gerlt, *Biochemistry*, 2005, **44**, 1807-1815.

19. R. P. McGlinchley, *Ph.D. Thesis*, University of St Andrews, 2006.

20. G. M. Tener, R. S. Wright, H. G. Khorana, *J. Am. Chem. Soc.*, 1956, **78**, 506-507.

21. G. M. Tener, R. S. Wright, H. G. Khorana, *J. Am. Chem. Soc.*, 1957, **79**, 441-443.

22. M. Onega, R.P. McGlinchey, H. Deng, J.T.G. Hamilton, D. O'Hagan, *Bioorg. Chem.*, 2007, **35**, 375-385.

23. M.J. Robins, S. Sarker, V. Samano, S.F. Wnuk, *Tetrahedron*, 1997, **53**, 447-456.

24. M. E. Perlman, *Nucleosides and Nucleotides*, 1993, **12**, 73-82.

25. A. Földesi, M. K. Kundu, Z. Dinya, J. Chattopadhyaya, *Helv. Chim. Acta*, 2004, **87**, 742-757.

26. K. K. Ogilvie, S.L. Beaucage, A. L. Schifman, N. Y. Theriault, K. L. Sadana, *Can. J. Chem.*, 1978, **56**, 2768-2780.

27. V. Samano, M. J. Robins, *J. Org. Chem.*, 1991, **56**, 7108-7113.

28. P.J. Garegg, B. Samuelsson, *Carbohydr. Res.*, 1978, **67**, 267-270.

29. A. Földesi, T. V. Maltseva, Z. Dinya, J. Chattopadhyaya, *Tetrahedron*, 1998, **54**, 14487-14514.

30. F. Hansske, D. Madej, M. J. Robins, *Tetrahedron*, 1984, **40**, 125-135.

31. H. Deng, S.M. Cross, R.P. McGlinchey, J.T.G. Hamilton, D. O'Hagan, *Chem. Biol.*, 2008, **15**, 1268-1276.

32. M. Onega, *Ph.D. Thesis*, University of St Andrews, 2009.

33. A. Sekowska, A. Danchin, *BMC Microbiol.*, 2002, **25**, 8.

34. W. R. Dolbier, Jr., *Guide to Fluorine NMR for Organic Chemists*, Wiley, 2009.

35. R. Keck, H. Haas, J. Retev, *FEBS Lett.*, 1980, **114**, 287-290.

36. D. H. Kim, G. W. Tucker-Kellogg, W. J. Lees, C. T. Walsh, *Biochemistry*, 1996, **35**, 5435-5440.

4. The total synthesis of β-ethynyl-L-serine and its *in vivo* detection by the click reaction

4.1 *Streptomyces cattleya* and β-ethynyl-L-serine

β-Ethynyl-L-serine **4** was isolated from a culture of *S. cattleya* in 1986 by Sanada *et al.* in Okazaki, Japan.[1] The same research group also reported the isolation of fluoroacetate and 4-fluorothreonine from extracts of *S. cattleya*.[2] β-Ethynyl-L-serine **4** was a known natural product, of the fungus *Sclerotium rolfsii*. The fungus had contaminated the feedstock of domestic chickens, leading to food poisoning and the acetylenic amino acid was identified as the toxin responsible (LD_{50} 150 mg/kg).[3]

In early investigations into the biosynthetic pathway of fluoroacetate and 4-fluoro-L-threonine, a hypothesis developed that β-ethynyl-L-serine **4** is a precursor to 4-fluoro-L-threonine.[4] When *S. cattleya* is growing in a defined medium, it produces β-ethynyl-L-serine after six days (at *ca.* 0.05 mM) and the accumulated concentration increases gradually to *ca.* 0.2 mM by day 28. The production level of 4-fluoro-L-threonine, at *ca.* 0.8 mM on day 28, is approximately four times higher than that of β-ethynyl-L-serine (Fig 4.1).[5] The low accumulated titre of β-ethynyl-L-serine is likely due to the decomposition of the unstable amino acid during the fermentation experiment.
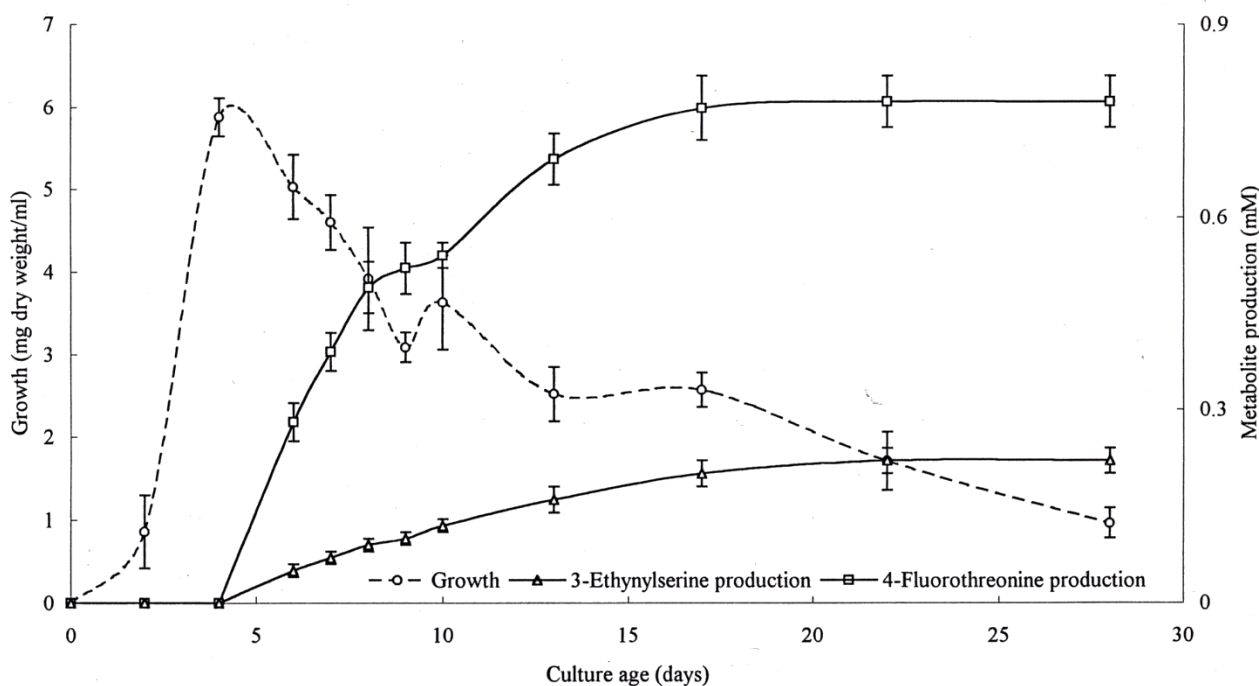
Fig. 4.1. The growth and production of secondary metabolite (β-ethynylserine and 4-fluorothreonine) of *S. cattleya* in a defined medium. Image taken from *PhD thesis,* J.T.G. Hamilton.[3]

Feeding experiments were conducted with $[1,2\text{-}^{13}C_2]$glycine, $[1\text{-}^{13}C_2]$pyruvate, $[2\text{-}^{13}C_2]$pyruvate and $[3\text{-}^{13}C_2]$pyruvate. In all cases there was no ^{13}C incorporation into β-ethynylserine. On the other hand, 4-fluorothreonine was significantly labelled by the pyruvate precursors in the same experiments. It was therefore concluded that the two amino acids do not originate from a common biosynthetic pathway.[5] A biosynthetic hypothesis for the observed labelling pattern into 4-fluorothreonine is presented in Scheme 4.1.[6,7,8]

The interest at that time was focused on the biosynthetic origin of the fluorometabolites, and therefore further studies into the biosynthesis of β-ethynylserine **4** were not pursued.
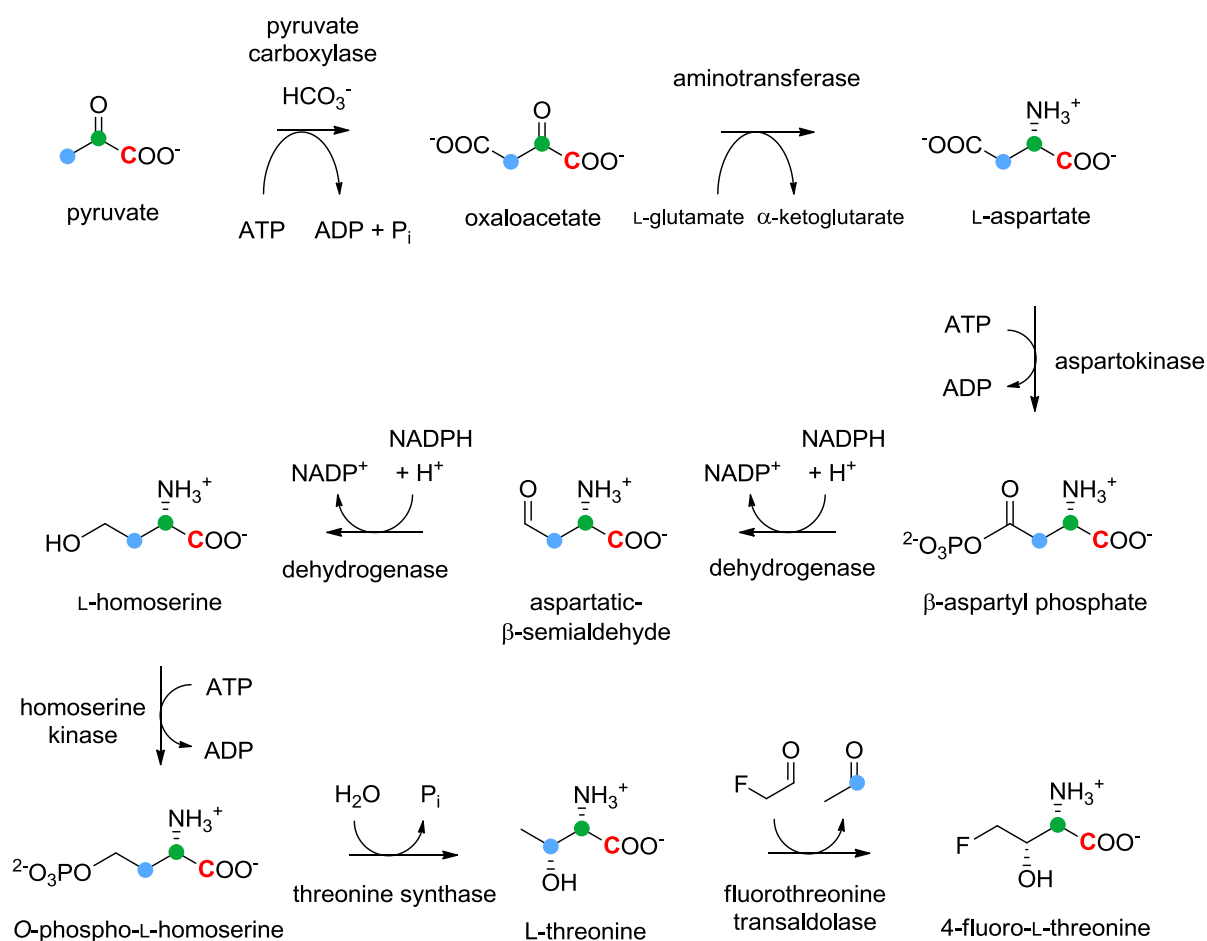
Scheme 4.1. Rationale for ^{13}C-labelled pyruvate incorporation into fluorothreonine in *S. cattleya*.[6,7,8]

A target of this project was to develop an enantioselective synthesis of β-ethynyl-L-serine to be used as a reference compound. This reference would be used to develop a click reaction procedure for the assay of the amino acid in extracts of of *S. cattleya* using HPLC/LCMS. It is envisaged that the analytical procedure will be applied for the future analysis of isotope-labelling studies, such that the biosynthesis of the amino acid can be explored in more details.

4.2 Previous syntheses of β-ethynyl-L-serine and derivatives

β-Ethynyl-L-serine is a small molecule with two stereogenic centres and four different functional groups. A synthesis of racemic β-ethynylserine has been reported.[3] Also an enantioselective synthesis of a protected form of β-ethynyl-L-serine has also been carried out.[9] Currently the total synthesis of this natural product in its free form had not been reported.

There are two obvious ways to disconnect the molecule by retrosynthetic analysis. These involve cleavage to either an acetylene and a serine aldehyde synthon (Disconnection A) or alternatively to glycine and propiolaldehyde (Disconnection B). In both cases, a strategy to control the stereochemistry can be envisaged. In the former case, the asymmetry could be derived from enantiopure serine which is oxidised to an aldehyde followed by diastereoselective addition of an acetylene group. In the latter case, enantioselectivity could be achieved by attaching a chiral auxiliary to the glycine before a diastereoselective coupling to propiolaldehyde (Scheme 2).
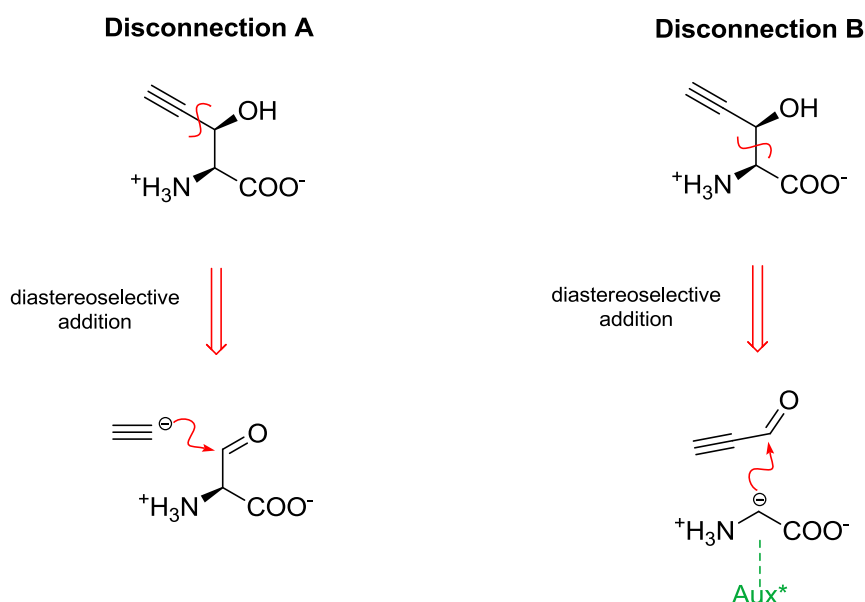
Scheme 4.2. Two possible disconnections for the retrosynthesis of β-ethynyl-L-serine.

4.2.1 Racemic synthesis of β-ethynylserine

Potgieter and co-workers[3] (University of Pretoria, South Africa) first isolated the toxic component from the fungus *Sclerotium rolfsii*. Chemical analysis showed that the compound was likely to be an α-amino acid (ninhydrin staining) and it precipitated as a cuprous acetylide with ammoniacal cuprous chloride. An empirical formula of $C_5H_7NO_3$ was obtained from a molecular mass of 129; IR absorptions were measured at 3264 cm^{-1} (asymmetric -C≡CH) and 2134 cm^{-1} (symmetric -C≡C-). Hydrogenation of the compound gave D,L-*threo*-β-hydroxynorvaline. Based on these data, along with the ^1H and ^{13}C NMR spectra, a correct structure was assigned to the isolated toxin. The absolute configuration was assigned on the basis that the compound was oxidised by an L-amino acid oxidase but not by a D-amino acid oxidase.[3]

A racemic sample of a mixture of diastereomeric β-ethynylserine was prepared (Scheme 4.3) to help confirm the structure assignment.

The synthesis was based on disconnection B (Scheme 4.2), by a condensation of copper glycinate **83** with propiolaldehyde **84** (Scheme 4.3).

racemic mixture of both diastereomers
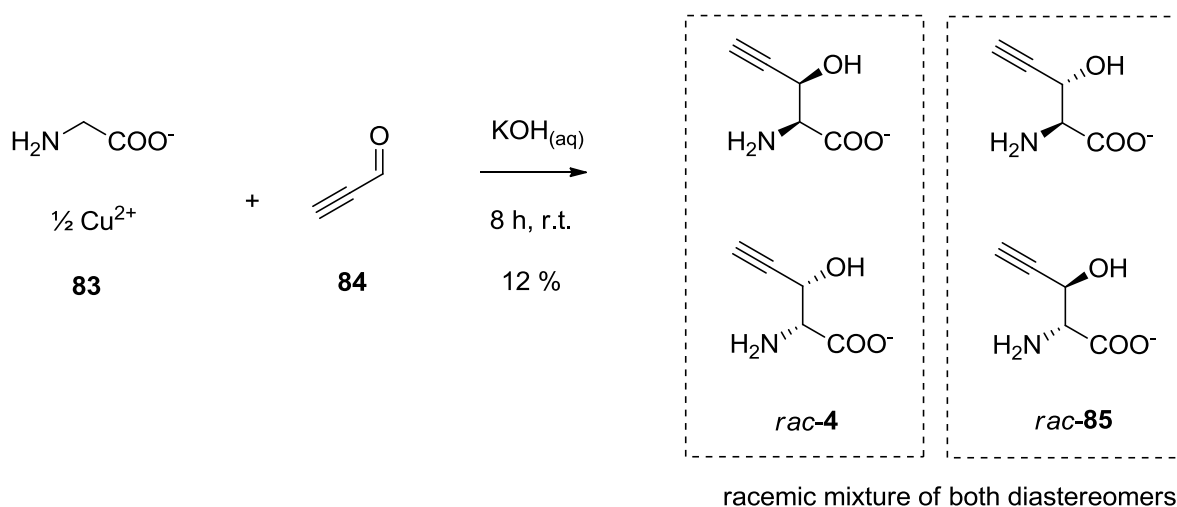
Scheme 4.3. The racemic synthesis of β-ethynylserine as a mixture of the two diastereomers by Potgieter.[3]

The compound rapidly decomposed under neutral or basic aqueous conditions where it turned into a brown solution.

Although the synthesis gave a mixture of four stereoisomers, it was accomplished in a single step, and provided a synthetic sample of the amino acid to secure a structure elucidation of the natural product.
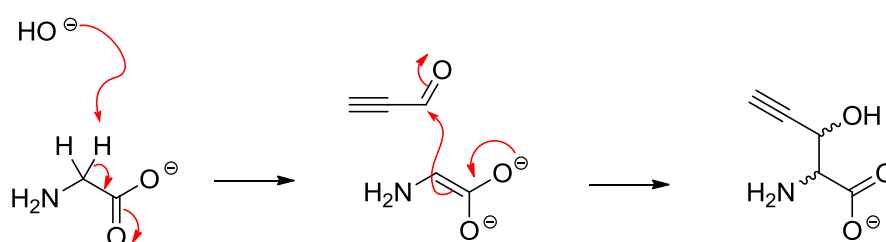
Scheme 4.4. Reaction mechanism of glycinate addition to propiolaldehyde.

4.2.2 Synthesis of a protected form of β-ethynyl-L-serine

LctM is a dehydratase involved in the biosynthesis of the lantibiotic lacticin 481 by *Lactococcus lactis.*[10, 11] LctM dehydrates hydroxyl group-bearing amino acids (Ser, Thr) into olefinic dehydroamino acids.[12] The dehydration can only occur on amino acid residues at specific positions of a short peptide chain extending from the C-terminus of a leader peptide sequence, that is recognised by the enzyme LctM (Scheme 4.5).
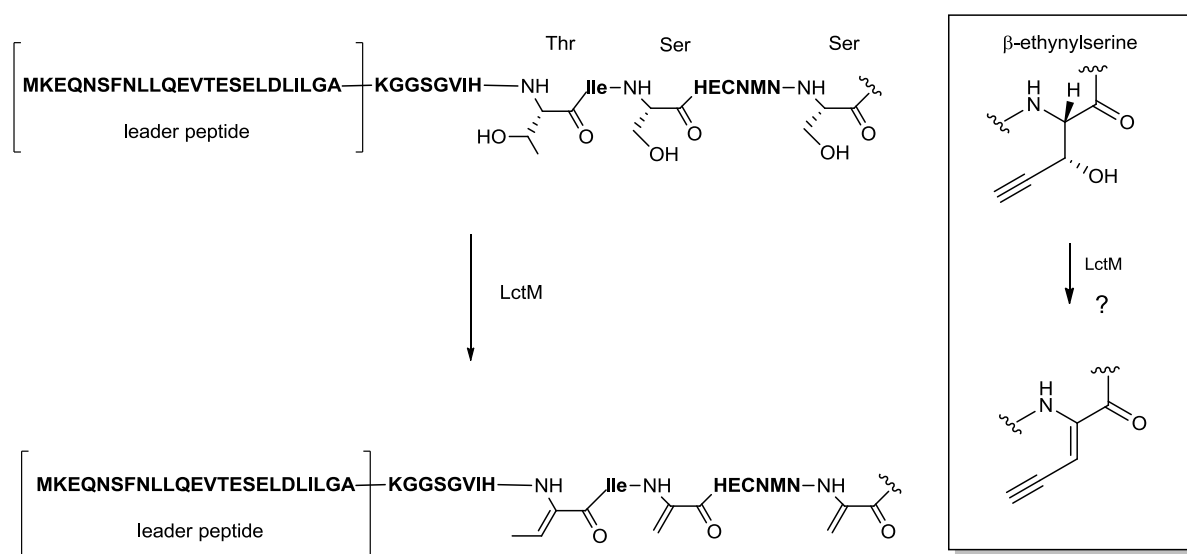

Scheme 4.5. LctM catalyse the dehydration of β-hydroxyamino acids.

Zhang and van der Donk conducted an investigation into the substrate specificity of LctM.[9] They prepared short synthetic peptides where Ser or Thr was replaced by unnatural amino acids, and these were linked to the leader peptide for enzymatic assay. β-Ethynylserine was among the eight different β-hydroxyamino acids used to probe the specificity of LctM. They were all prepared in suitably protected forms for Fmoc-based solid phase peptide synthesis.

β-Ethynyl-L-serine **86** was prepared with Fmoc on NH_2 and a *tert*-butyl ether on OH, the carboxylic acid was left free for coupling chemistry.

86
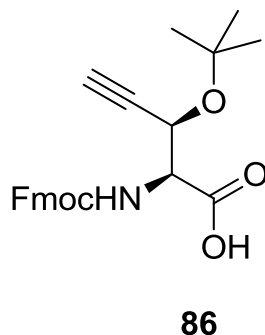
The synthetic approach was based on disconnection A (Scheme 4.2), and involved a key coupling reaction between D-Garner's aldehyde **87** and an organocopper(I) acetylide (Scheme 4.6). This coupling reaction was developed earlier by Herold and proceeded with high (20:1) diastereoselectivity.[13]
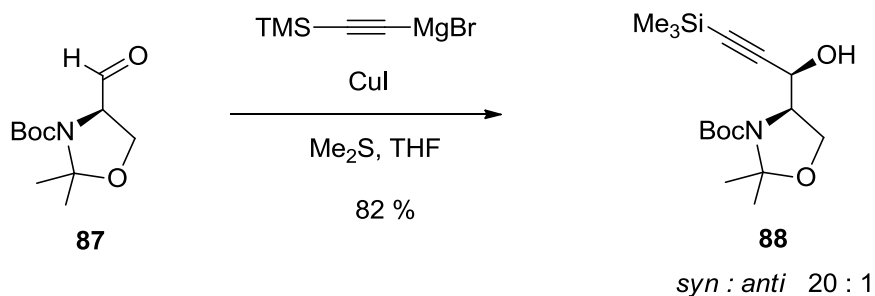

Scheme 4.6. The key coupling reaction in the synthesis of **86**.[4]

The diastereoselectivity of the reaction can be rationalised by chelation control, where the aldehyde and the Boc-carbamide C=O co-ordinate a magnesium ion. This stabilisation promoted an *anti*-Felkin-Anh orientation and the use of a soft

organocopper(I) nucleophile, ensured the reaction was under thermodynamic rather than kinetic control. Herold had shown that a lithium acetylide gives an equally high selectivity, but for the *anti*-product (Fig 4.2).[13,14]
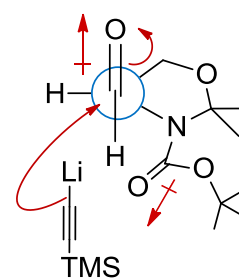
Organocopper(I) acetylide
in the presence of Mg^{2+}

Organolithium acetylide
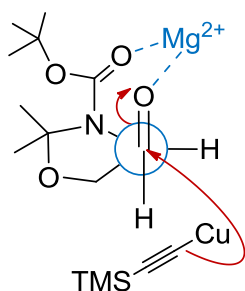
Chelation control
The conformation is held by a Mg^{2+} ion.

Felkin-Anh model
The conformation minimises the overall dipole.

Fig 4.2. Rationale for the diastereoisomeric control of the coupling reaction.

The synthesis was completed in 8 steps and 62 % overall yield from D-Garner's aldehyde by protecting group manipulations and an oxidation of the primary alcohol to the free carboxylic acid (Scheme 4.7). After the coupling reaction to form **88**, all protecting groups were removed in two steps giving amino alcohol **90**, on which protecting groups were installed sequentially. The primary alcohol was first protected as a bulky TBDPS silyl ether. Then an O-*t*-Bu ether was installed on the free secondary OH group. This strategy allowed the primary OH to be released by treatment with HF-pyridine such that a Jones oxidation could be performed to give the protected ethynylserine **86**. This synthesis illustrated in Scheme 4.7 formed the basis of our synthetic strategy.
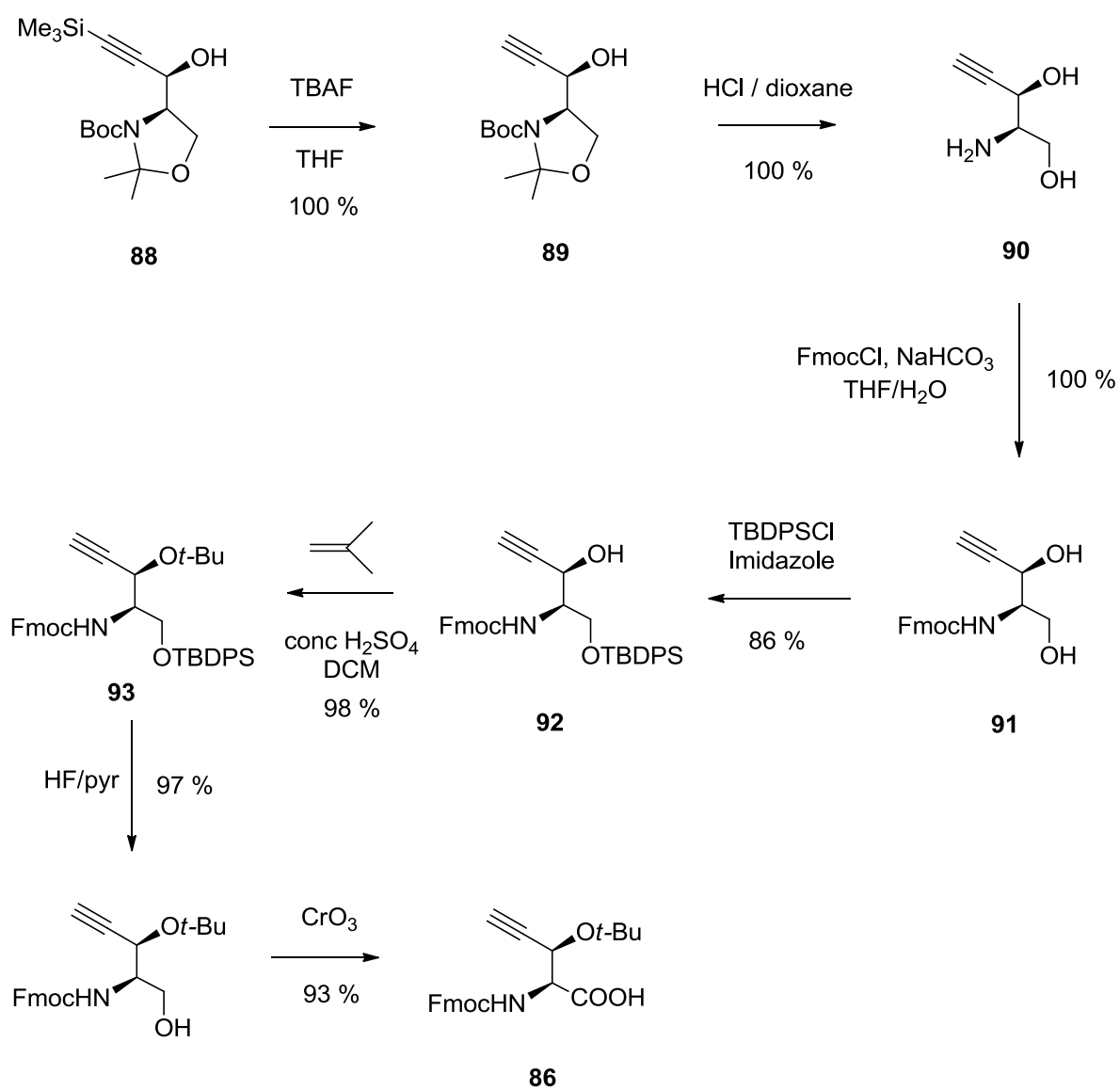
Scheme 4.7. The synthesis of protected ethynylserine **86** by van der Donk and Zhang.[9]

4.3   An approach to an asymmetric synthesis of β-ethynyl-L-serine using the Belokon nickel(II) complex of glycine

Belokon developed a versatile asymmetric synthesis for β-hydroxy-α-amino acids using a chiral glycine equivalent, which has been reacted with various aldehydes and ketones.[15, 16] Glycine **95** is attached to the proline-based chiral auxiliary (S)-2-[N-(N-benzylprolyl)amino]benzophenone [(S)-BPB] **96** *via* formation of the Schiff base and this is complexed with a nickel(II) ion to form a highly stable square-planar complex, known as Ni-(S)-BPB-Gly **97**.

glycine
95

(S)-BPB
96

Ni-(S)-BPB-Gly
97
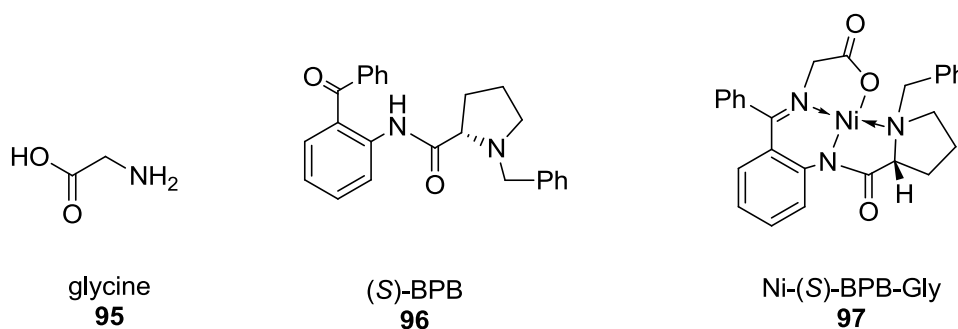
The methylene protons of glycine in complex **97** can be deprotonated and many carbon electrophiles, mainly aldehydes and ketones have been coupled with the complex to produce α-amino acids with excellent diastereoselectivity for the *threo*-product and with excellent enantioselectivity (often >90 %)[15] (Scheme 4.8). It was envisaged that a coupling between propiolaldehyde and Ni-(S)-BPB-Gly may offer a useful route to β-ethynyl-L-serine.
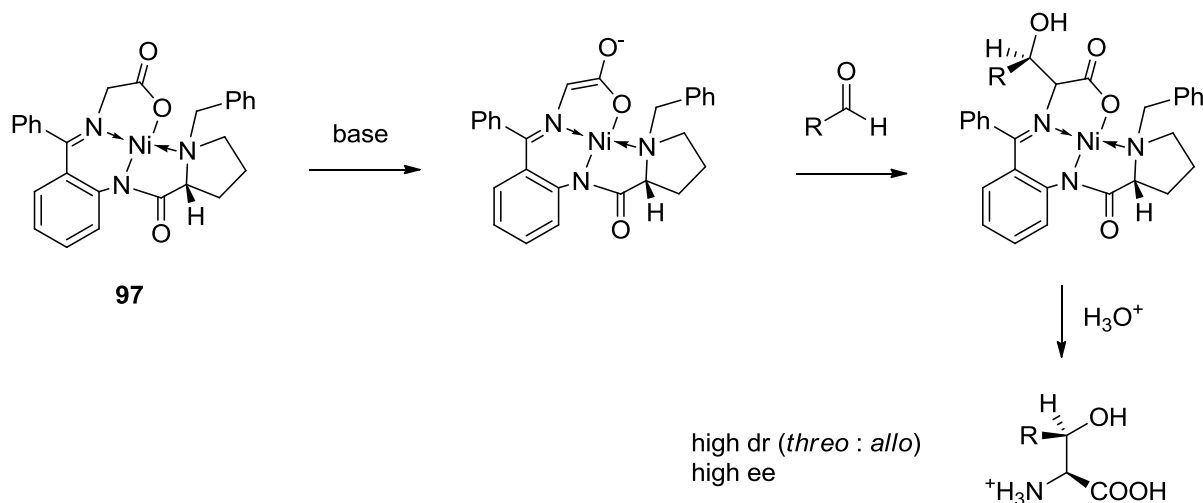
Scheme 4.8. Ni-(S)-BPB-Gly reacts with an aldehyde to synthesise α-amino acids.

While most saturated aliphatic aldehydes and ketones reacted well to produce the expected β-hydroxy-α-amino acids in good to excellent yields and stereocontrol,[15] conjugated α,β-unsaturated aldehydes and ketones do not give the obvious 1,2-addition product. Instead the conjugate 1,4-addition product forms exclusively and this had been exploited as a practical route to substituted prolines (Scheme 4.9).[17]
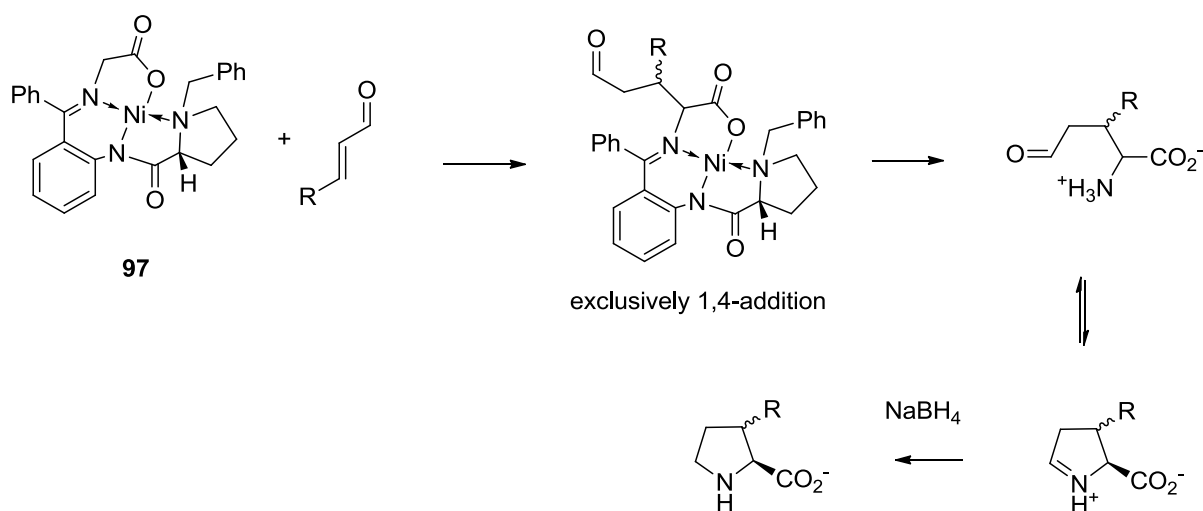
Scheme 4.9. Using α,β-unsaturated aldehydes to synthesise substituted prolines.[17]

The Ni-(S)-BPB-Gly complex was prepared following the published procedure.[18] Firstly L-proline **98** was benzylated with benzyl chloride and potassium hydroxide to generate *N*-benzyl L-proline **99**, which was then coupled to *o*-aminobenzophenone **100** to give the chiral auxiliary (S)-BPB **96**.
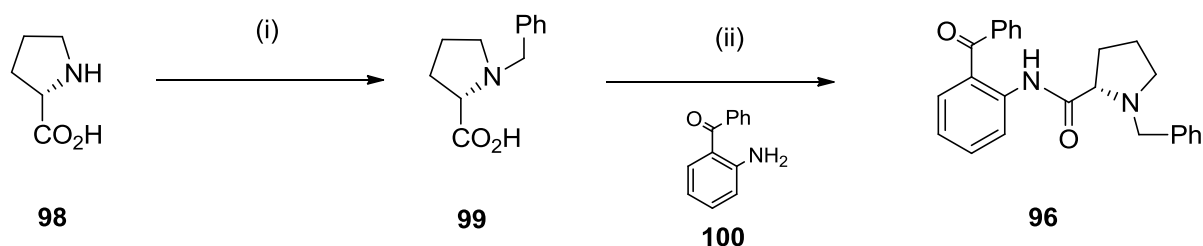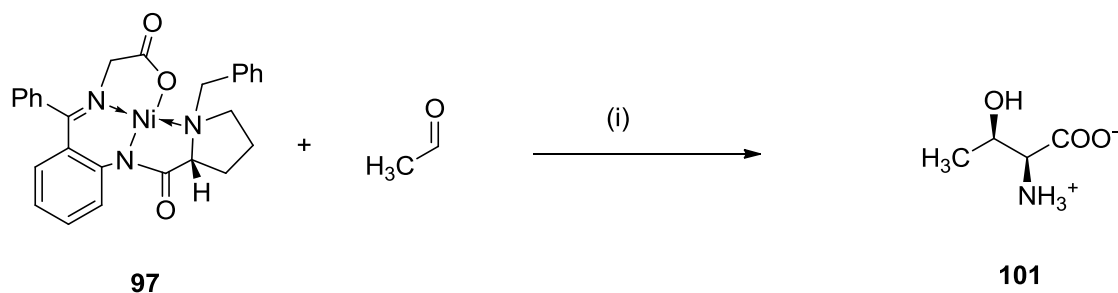

Scheme 4.10.

Reagents and conditions: (i) BnCl, KOH, *i*-PrOH, 40 °C (63 %) (ii) *o*-aminobenzophenone **100**, MsCl, *N*-methylimidazole, CH_2Cl_2 (88 %).

Finally (S)-BPB **96** was reacted with nickel(II) nitrate and glycine to give the complex Ni-(S)-BPB-Gly **97**, which was obtained as bright red crystals after recrystallisation.
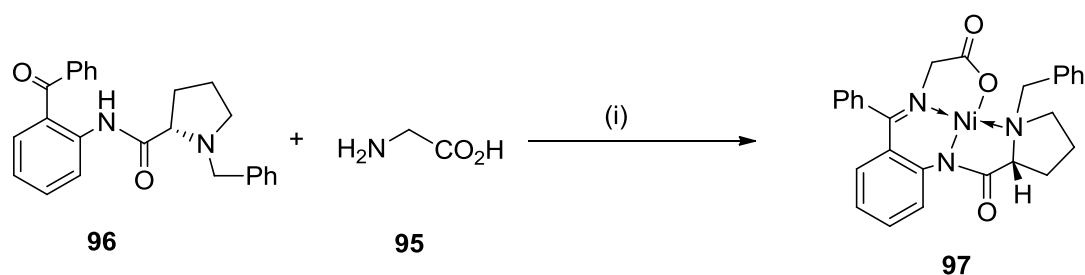
Scheme 4.11. Reagents and conditions: (i) $Ni(NO_3)_2$, NaOH, MeOH, 55 °C (84 %).

As a model reaction, the complex was reacted with acetaldehyde using sodium methoxide as the base, and a product **101** identical to L-threonine by TLC was obtained after quenching the reaction with an acid.

Scheme 4.12. Reagents and conditions: (i) 1. NaOMe, MeOH, 1 min., then 5 % AcOH (aq).

Propiolaldehyde is not a commercially available substrate. It is susceptible to self-polymerisation and must be freshly prepared immediately prior to use from propargyl alcohol.[19] However, the TMS-protected propiolaldehyde **102** is more stable and is commercially available. The TMS-acetylene group can be deprotected relatively easily by treatment with KOH or NaOEt. This deprotection may even be completed *in situ* in an excess of sodium methoxide.

A series of test reactions were carried out using **102** and the products were analysed by both RP-TLC with ninhydrin staining and by NMR (Scheme 4.13, Table 4.1).
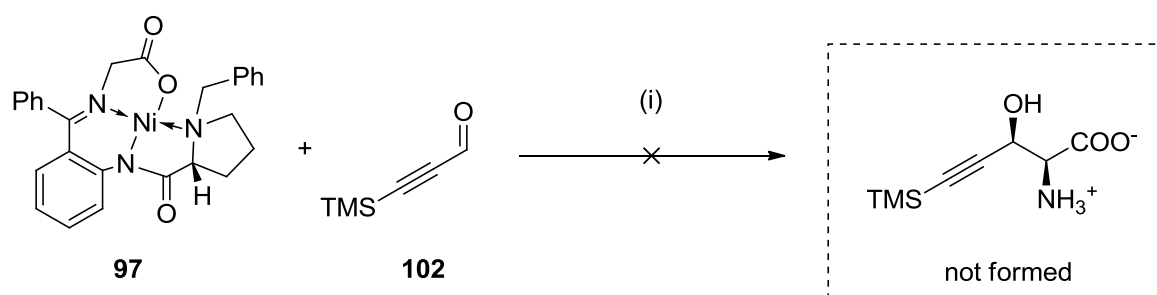
Scheme 4.13. Reagents and conditions: (i) 1. NaOMe, MeOH, then 5 % AcOH (aq).

Table 4.1. Reaction conditions tested for Ni-(S)-BPB-Gly coupling reactions.

Ni-(S)-BPB-Gly	aldehyde	base	reaction time/temp	Result
1 eq.	acetaldehyde 1.5 eq.	NaOMe 3.5 eq.	1 min / r.t.	L-threonine formed after work-up
1 eq.	**102** 1.5 eq.	NaOMe 3.5 eq.	1 min / r.t.	complex mixture, incomplete reaction
1 eq.	**102** 2.5 eq.	NaOMe 3.5 eq.	1 min / r.t.	complex mixture, SM consumed
1 eq.	**102** 2.5 eq.	NaOMe 10 eq.	1 min / r.t.	complex mixture, SM consumed
1 eq.	**102** 2.5 eq.	NaOMe 3.5 eq.	1 min / -78 °C	recovered SM

The reactions with **102** mostly gave complex mixtures. Side reactions such as conjugate addition and decomposition may have occurred. The strongly basic conditions may of course lead to the degradation of ethynylserine immediately on formation. This approach was not pursued.

4.4 The synthesis of β-ethynyl-L-serine from D-serine

The synthesis of β-ethynyl-L-serine **4** was completed using a stepwise approach from D-serine similar to that used by Zhang and van der Donk.[9] Since the protected form **86** of ethynylserine was not required, the synthesis was optimised for delivery of the free amino acid **4**.

4.4.1 Preparation of D-Garner's aldehyde from D-serine

Garner's aldehyde[20] has been shown to be a synthetically useful chiral building block[21] and it is relatively easy to prepare from either L- or D-serine in 3 to 5 steps.[22,23,24] The OH and NH$_2$ groups of serine are protected as an oxazolidine. This 5-membered ring gives the molecule conformational rigidity and steric bulk that imparts facial selectivity to nucleophilic reactions on the aldehyde. Our synthesis of Garner's aldehyde followed the published procedures by McKillop *et al.*[23] and Moriwake *et al.*[24]
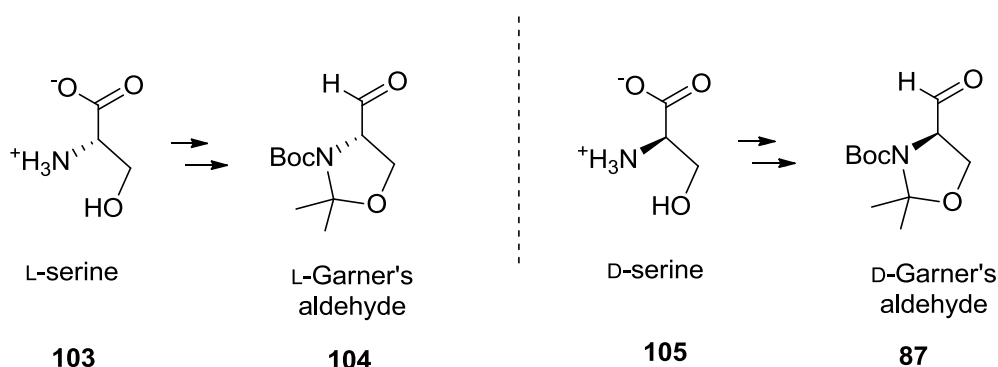

Fig. 4.3. D/L-serine can be converted into the synthetically versatile D/L-Garner's aldehyde.

D-Garner's aldehyde **87** was prepared in three steps from D-serine methyl ester **106**, which was purchased as its hydrochloride salt. The free amine was protected using Boc$_2$O, and then the oxazolidine **108** was formed by condensation with acetone and 2,2-dimethoxypropane. Lastly a diisobutylaluminium hydride (DIBAL) reduction of the methyl ester gave aldehyde **87**.

Boc-protection of **106** proved to be straight-forward and gave **107** in an excellent yield (92 %) and purity.
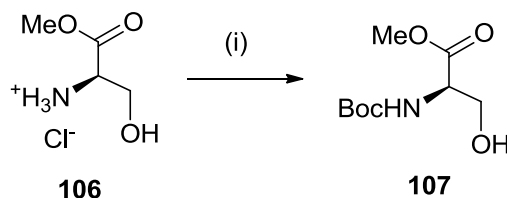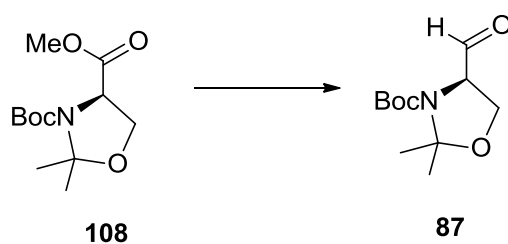

Scheme 4.14. Reagents and conditions: (i) Boc$_2$O, Et$_3$N, CH$_2$Cl$_2$ (92 %).

The formation of oxazolidine **108** from **107** was most efficient (96 %) when BF$_3$-etherate was used as a Lewis-acid catalyst. A low loading of the catalyst sufficed (20 μl per 1 g **107**, 3.5 mol %) and the reaction was allowed to proceed for 24 h at 0 °C. Experience showed that the 2,2-dimethoxypropane must be dried over sodium and freshly distilled prior to use. When the level of BF$_3$-etherate was increased or when the reaction was carried out at room temperature, a substantial increase in side-products was observed. The use of *p*-toluenesulfonic acid as the catalyst under Dean-Stark conditions also led to the generation of side-products.
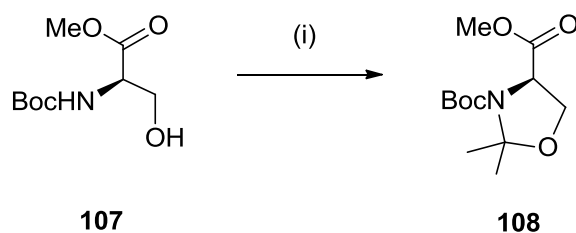
Scheme 4.15. Reagents and conditions: (i) Acetone, dimethoxypropane, BF$_3$.OEt$_2$, 4 °C (83-96 %).

DIBAL reduction of the methyl ester was carried out at -78 °C, and was then quenched with methanol, keeping the temperature at -78 °C. Under these conditions, very little of the over-reduction product was observed. The over-reduced side product could be readily separated by column chromatography. The reaction typically gave between 63-85 % yield after chromatographic purification.

Scheme 4.16. Reagents and conditions: DIBAL/Toluene, -78 °C (63 – 85 %).

This preparation of D-Garner's aldehyde was performed on a 10 g scale. It was obtained as a colourless syrup and was not very stable. Decomposition was observed after storage for 1 month at -20 °C, therefore it was prepared freshly before use and used in the following few days.

4.4.2 The addition reaction of the organocopper(I) acetylide reagent to D-Garner's aldehyde

With D-Garner's aldehyde in hand, the acetylene unit was then inserted using the stereoselective coupling reaction described by Herold.[13]

Scheme 4.17. Reagents and conditions: (i) CuI, Me₂S, THF (70 %).

The reagent in this procedure is an organocopper(I) acetylide **109**. This was prepared in two steps from ethynyltrimethylsilane **110** where the terminus of the acetylene was protected using a trimethylsilane moiety. Silane **110** was added to a solution of EtMgBr in THF and the mixture was refluxed for 2 h to pre-prepare a solution of the Gringard reagent **111**. This was then treated with CuI/Me₂S in THF to give the organocopper(I) reagent **109**.
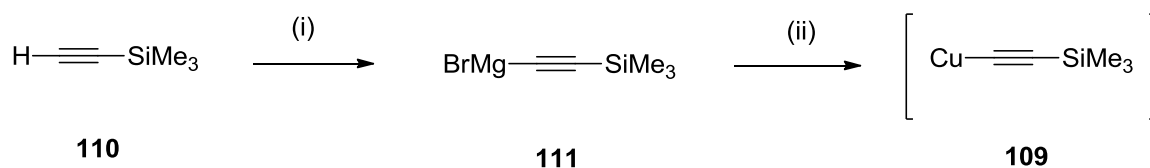
Scheme 4.18. Reagents and conditions: (i) EtMgBr, THF, reflux (ii) CuI, Me₂S, -30 °C.

A THF solution of D-Garner's aldehyde was then added slowly into the organometallic reagent at -78 °C, and then the reaction was left to proceed at -30 °C for 30 mins. After

work-up, the aldol addition product **88** was obtained in 70 % yield after column chromatography. The minor diastereoisomer usually formed at around 5 % (by ^{1}H-NMR), but was readily separated during the chromatography step, resulting in purification of the aldol product **88** as a single diastereoisomer.

The success of this reaction is very dependent upon a good preparation of the organometallic reagent **109**. In particular, the complete solvation of the CuI in Me$_2$S/THF was critical for good diastereoselectivity. This could be achieved by stirring the white powder in a large excess of Me$_2$S/THF at room temperature until a clear pale green solution was obtained, before cooling the mixture for the addition of the Grignard reagent. When the reaction was carried out by addition of CuI to the reaction at -30 °C, a poor diastereoselectivity was obtained and most of the CuI remained insoluble in the mixture.

## 4.4.2.1 Using Carreira's reaction for the aldol reaction

The Carreira alkynylation was explored in order to simplify this protocol. Carreira's reaction is an asymmetric aldol reaction that couples an aldehyde with acetylenes directly with zinc triflate as a catalyst and using N-methylephedrine as a chiral auxiliary. It has been widely used for the asymmetric preparation of chiral propargyl alcohols.[25, 26]
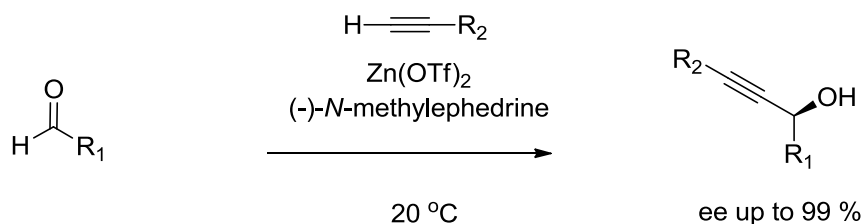
Scheme 4.19. General reaction scheme of the Carreria asymmetric alkynylation.[25, 26]

Success by this approach would allow for the direct coupling of D-Garner's aldehyde to ethynyltrimethylsilane, without the need to pre-prepare the organometallic reagents. Furthermore the enantioselectivity (diastereoselectivity in our case) of this reaction can be changed by using either the $(+)$- or the $(-)$-enantiomer of N-methylephedrine.

Both D-Garner's aldehyde and the acyclic aldehyde **112** were explored as substrates. However there was no reaction in either case, even when the reaction was heated to 60 °C overnight. Starting material was recovered in both cases.
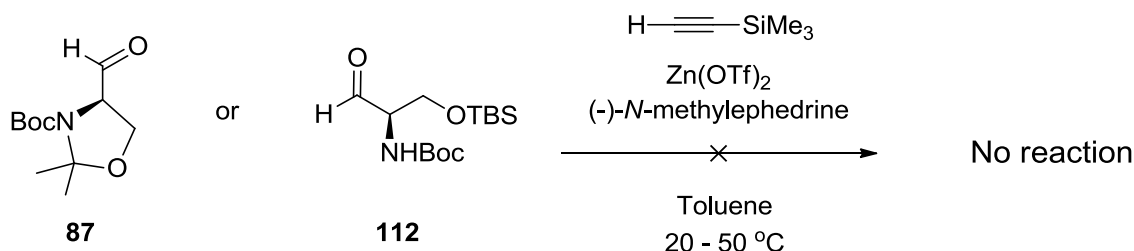
Scheme 4.20. Attempted acetylene coupling reactions under Carreira reaction conditions.

Acyclic aldehyde **112** was prepared as a substrate for this reaction. Aldehyde **112** was generated from the protected serine **107** as illustrated in Scheme 4.21.
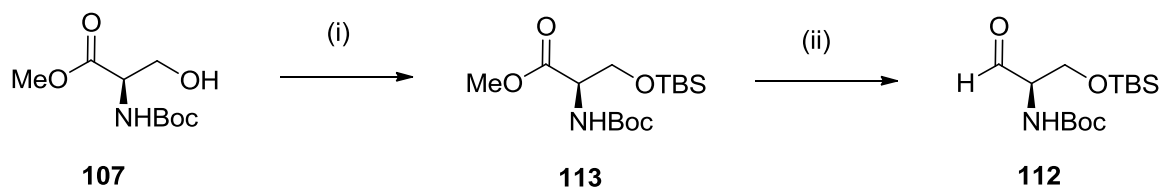
Scheme 4.21. Reagents and conditions: (i) TBSOTf, pyr (77 %) (ii) DIBAL/toluene (quant.).

The mechanism of the Carreira's reaction is believed to involve a pre-formed reactive scaffold of Zn^{2+} coordinated to both the acetylene and N-methylephedrine.[26] The lack of reactivity for **87** and **112** may be due to a competition between the substrate aldehyde and N-methylephedrine **114** for coordination to Zn^{2+}, since both of these compounds have a similar layout of N and O atoms (Fig. 4.4). There are no literature examples where such amino alcohol containing aldehydes were successful substrates for this reaction. Therefore this reaction was not further pursued.
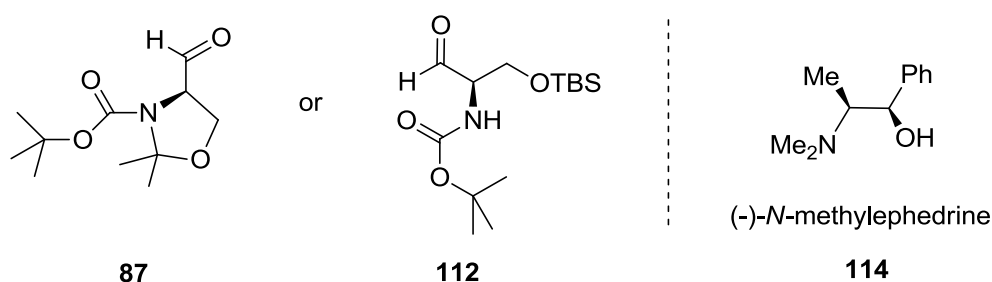

Fig 4.4. Comparison of the structure similarities of **87** and **112** with N-methylephedrine.

4.4.3 Synthesis to β-ethynyl-L-serine from the aldol addition product

The synthesis towards β-ethynylserine **4** is continued from the copper-mediated acetylene addition to D-Garner's aldehyde. In order to convert the aldol product **88** to **4**, the primary alcohol requires to be oxidised to the free carboxylic acid.

In this synthetic sequence, the original carboxylate group of D-serine was reduced to an aldehyde which then became the primary alcohol of the product **4**; and the original primary alcohol of D-serine was oxidised to become the carboxylate of **4**.

The particular connectivity of serine means that the two functional groups can swap their oxidation levels and invert the configuration without changing the chemical identity of the amino acid (Fig. 4.5).
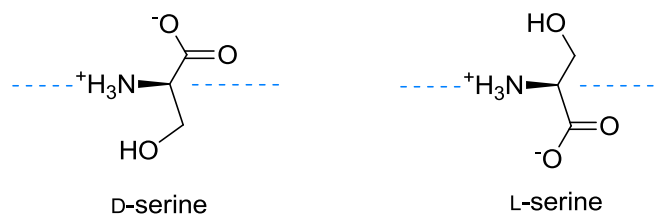
Fig 4.5. Symmetry within the serine molecule.

In order to carry out this oxidation and retaining selectivity, protecting groups are required on the secondary alcohol and the amino group of **88**, but the primary alcohol needs to be freed from the protecting group.

The protecting groups increase lipophilicity and facilitate reaction work-ups and the purification of the intermediates. Therefore careful consideration was taken regarding the selection of protecting groups such that they could be orthogonally manipulated.

A strategy was taken to first protect the secondary alcohol of **88** generated after acetylene addition. Then the oxazolidine ring could be opened to expose the free primary alcohol, which could then be oxidised to a carboxylic acid. Global deprotection would release the final amino acid **4** (Scheme 4.22).
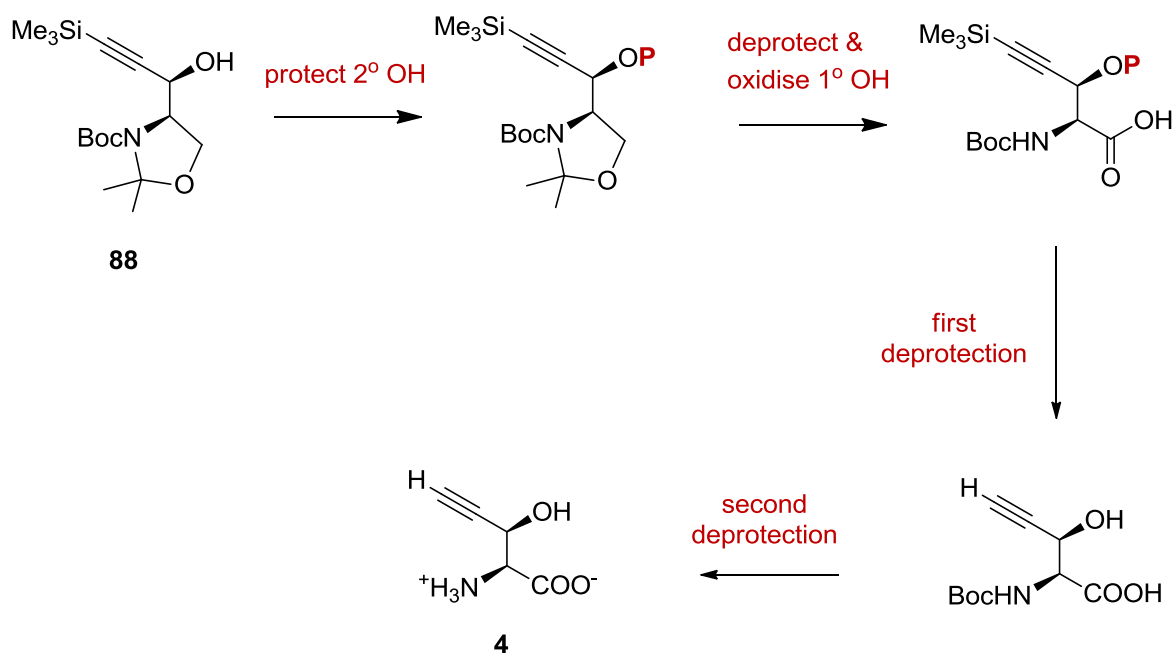
Scheme 4.22. Synthesis strategy for the final steps toward β-ethynylserine **4**.

4.4.3.1   Protection of the secondary alcohol **88**

The newly generated secondary hydroxyl group of **88** required to be protected leaving the primary alcohol exposed for a selective oxidation. A stable protecting group was required. In addition, a UV-active group was desirable to allow visualisation of products by TLC.

Initially the use of phenoxyacetate ester **116** was envisaged. The electron-poor bulky ester has a better stability under acidic conditions than most esters but is easy to remove with ammonia.[27] Therefore alcohol **88** was treated with phenoxyacetyl chloride **115** and triethylamine to give ester **116** in 82 % yield and as colourless needles.
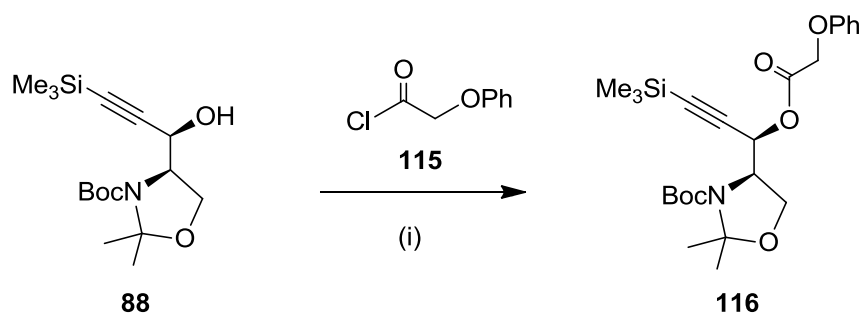
Scheme 4.23. Reagents and conditions: (i) Phenoxyacetyl chloride **115**, Et$_3$N, CH$_2$Cl$_2$ (82 %).

An absolute X-ray crystal structure of **116** was obtained confirming its absolute and relative configuration (Fig. 4.6). The structure had the desired stereochemistry.
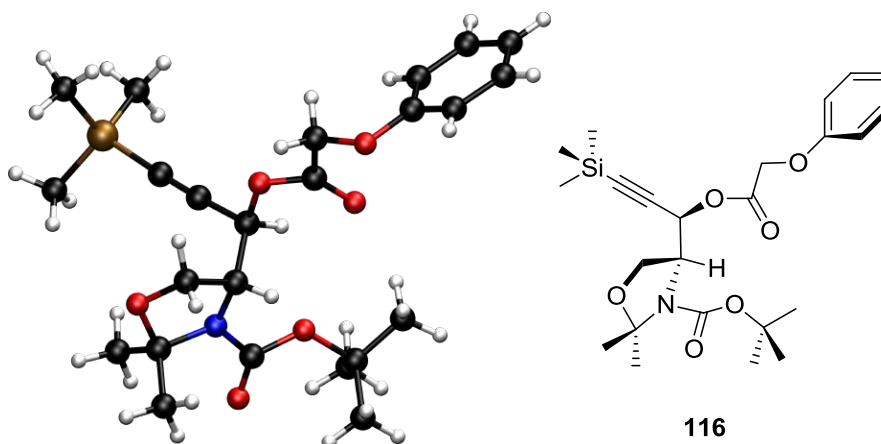
Fig 4.6. Single crystal X-ray structure of **116** with an absolute configuration of (2*R*, 3*R*)-**116**.

To generate amino alcohol **117,** ester **116** was treated with 90 % TFA. ^1H and ^{13}C NMR indicated that the phenoxyacetate group was intact. This amino alcohol was then subjected to a Jones oxidation. A single compound was purified which had a new carbonyl as judged by ^{13}C-NMR. It was initially thought that the carboxylic acid **118** was obtained, however upon closer inspection of the ^1H-^{13}C HSQC NMR, the product was in fact ketone **120.** This ketone clearly arises from oxidation of the secondary OH of **119** where the phenoxyacetate ester had migrated onto the primary alcohol.
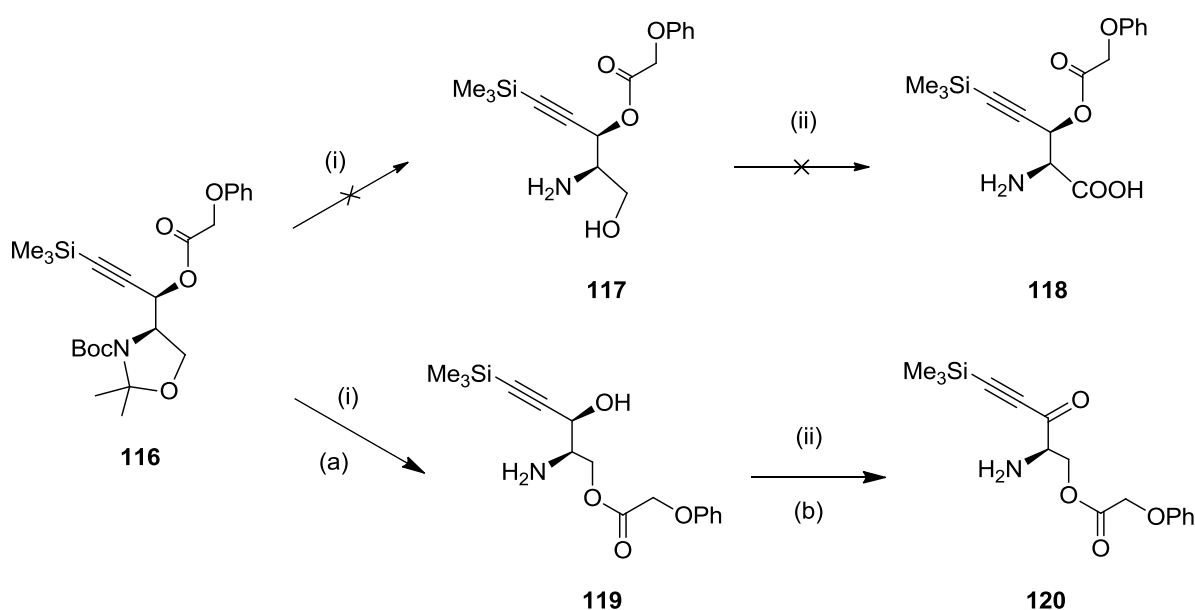
Scheme 4.24. Reagents and conditions: (i) 90 % TFA (aq) (a. 91 %) (ii) CrO_3/H_2SO_4 (aq), acetone (b. 84 %).

The migration of the ester probably occurred as a result of the strongly acidic deprotection conditions with TFA through a six-membered ring transition state (Fig 4.7) and the bulky ester group resting on the more thermodynamically stable primary position. It was not obvious that the migration had happened by studying the 1D ^1H and ^{13}C NMR spectra of **119** alone.
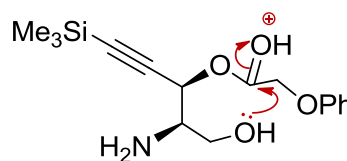
Fig 4.7. Phenoxyacetate group migration in **119**.

4.4.3.2 Protection of the secondary alcohol as a TBDPS ether

The ester group migration called for a more acid stable protecting group. For this, the *tert*-butyldiphenylsilyl (TBDPS) ether group was chosen. It possesses a very bulky silyl group but unlike most silyl ethers it is quite stable to strong acids. The typical procedure for the removal of the TBDPS group involves treatment with KOH in methanol or HF-pyridine.[28]

Initially the formation of the TBDPS ether was not very efficient (yields 30-50 %). Commercially available TBDPSCl was used as the silylating reagent and either pyridine or imidazole was used as the base. The reaction was rather slow and it was necessary to use 5 equivalents of the silyl chloride to force the reaction to completion. Removal of the side product TBDPSOH proved difficult by chromatography.
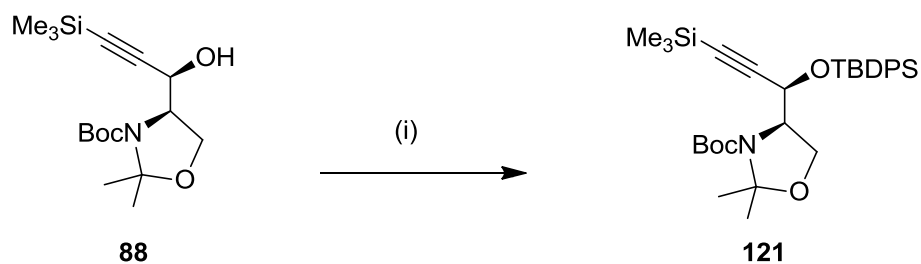
Scheme 4.25. Reagents and conditions: (i) TBDPSCl (5 eq.), imidazole or pyridine (30-50 %).

This problem was overcome by using the more reactive TBDPS triflate **122** as the silylating reagent. A smooth silylation was achieved using 1.5 equivalents of **122**. The work-up was straightforward, giving the silyl ether in 75 % yield as an oil.
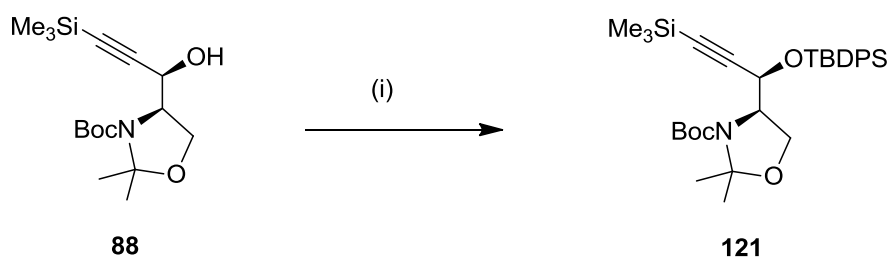
Scheme 4.26. Reagents and conditions: (i) TBDPSOTf **122** (1.5 eq), Et$_3$N, CH$_2$Cl$_2$ (75 %).

However TBDPSOTf is not commercially available and had to be prepared by a two-step reaction.[29] TBDPSCl **123** was first reacted with phenyllithium to give *tert*-butyltriphenylsilane **124** as a white solid. The silane was then reacted with 1 eq. of triflic acid to generate TBDPSOTf **122** in a DCM solution and this was used immediately for the silylation of alcohol **88**.
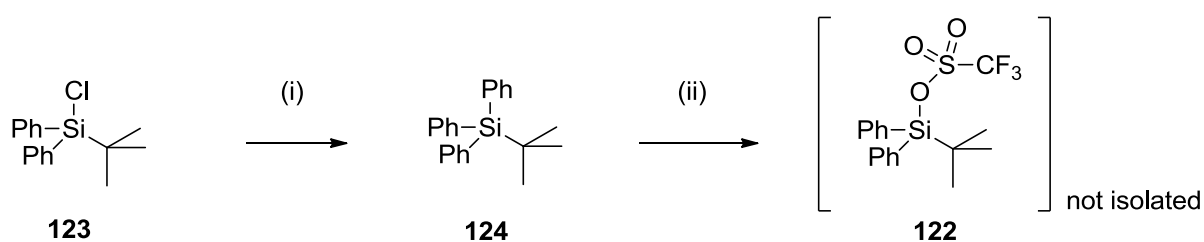
Scheme 4.27. Reagents and conditions: (i) PhLi, *t*-Bu$_2$O (95 %), (ii) CF$_3$SO$_3$H, CH$_2$Cl$_2$.

The TBDPS silyl ether remained intact during the subsequent deprotection with 90 % TFA. Re-protection of the amino group with Boc$_2$O also proceeded smoothly with both reactions giving excellent yields.
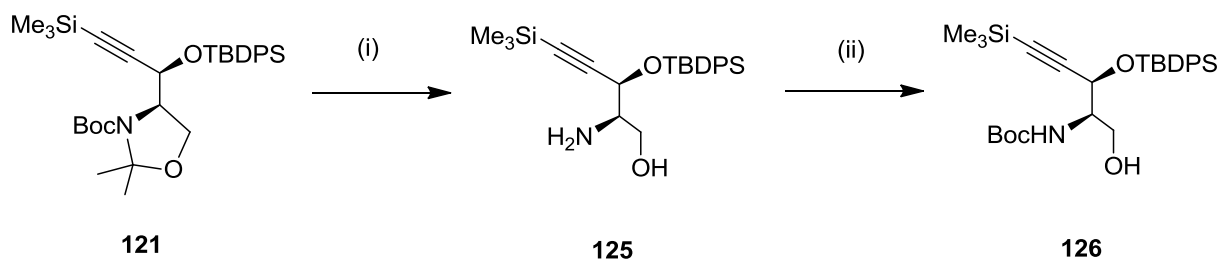
Scheme 4.28. Reagents and conditions: (i) 90 % TFA (aq) (99 %), (ii) Boc$_2$O, Et$_3$N, CH$_2$Cl$_2$ (94 %).

4.4.3.3 Oxidation of primary alcohol **126** into carboxylic acid

Now with the protecting group manipulation achieved, oxidation of the free primary alcohol was explored.

The oxidation of alcohol **126** was carried out in acetone with 2.5 equivalents of Jones reagent[30] (CrO_3 in dilute H_2SO_4). The reaction went smoothly and complete conversion was observed after 3 hours at room temperature. At this point, the appearance of the reaction mixture was observed to change from cloudy to clear with dark green precipitates.
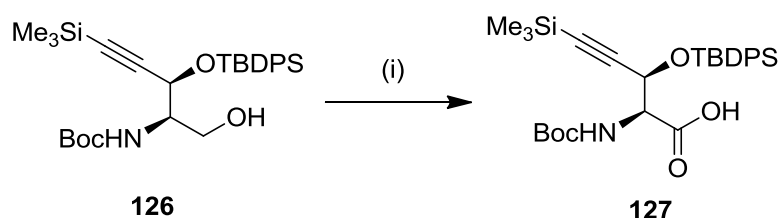
Scheme 4.29. Reagents and conditions: (i) CrO_3/H_2SO_4 (aq), acetone (93 %).

Carboxylic acid **127** was readily purified by chromatography and was obtained in 93 % as a colourless oil that crystallised on standing. Recrystallisation from hexane/ether gave **127** as colourless crystalline prisms.

An absolute structure of **127** was obtained by X-ray crystallography, and the absolute configuration was determined to be (2*S*, 3*R*) as shown in Figure 4.8. This assignment is consistent with that based on relative configuration, since the (2*S*) chiral centre had originated from D-serine. This crystal structure shows the complete skeleton of β-ethynyl-L-serine in a protected form and confirmed the desired stereochemistry and connectivity are in place.
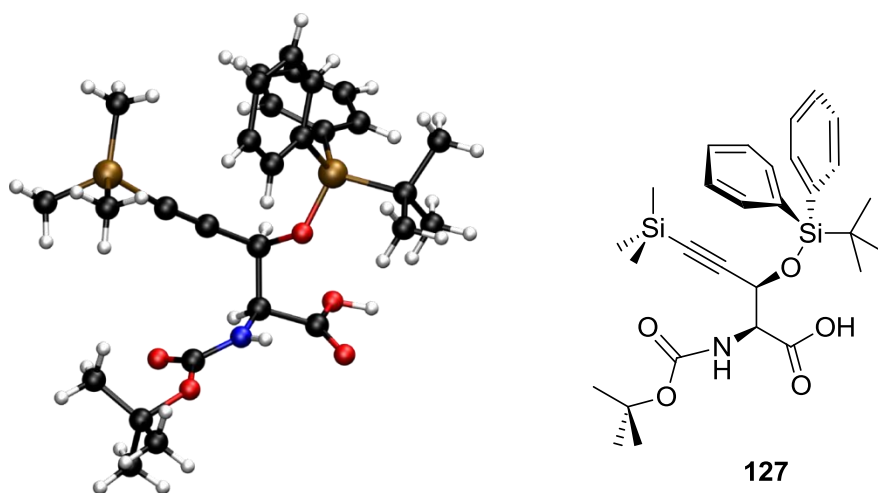
Fig 4.8. Single crystal X-ray structure of **127** with absolute configuration assigned as (2*S*, 3*R*)-**127**.

4.4.3.4 Final deprotection to the free amino acid **4**

The deprotection of carboxylic acid **127** had to be carefully planned, because after protecting group removal, the product is zwitterionic and water-soluble. Any ionic contaminants will be difficult to remove.
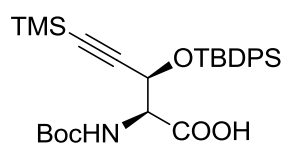
127

The silyl groups could clearly be removed by either a fluoride source (HF or TBAF) or strong base (KOH in MeOH). The Boc-group is generally removed by strong acids such as HCl or TFA.

Removal of both silyl groups was carried out in a single step. Potassium hydroxide was chosen in favour of a fluoride ion source since a simple work-up could be carried out using cation exchange (H^+) resins to neutralise any excess OH^- and scavenge K^+ ion.

Thus carboxylic acid **127** was treated with a methanolic solution of KOH until the protecting groups were cleaved (by TLC). The mixture was diluted with methanol and then neutralised with Dowex 50WX4-400 (H^+ cation exchange resin) and the resins were filtered off, and the solvents evaporated.

The residues containing **128** were stirred in conc. HCl to remove the last Boc-protecting group, then the mixture was diluted with water, and washed thoroughly with ether to remove any organic residues related to silyl group deprotection.

The aqueous layer containing the amino acid and small excess of hydrochloric acid was then freeze-dried. This gave β-ethynyl-L-serine hydrochloride as a light yellow powder. The NMR spectra data (Fig. 4.9, 4.10) matches that reported for this compound when isolated from *S. cattleya*.[1] A high resolution mass spectra also confirms the identity of this compound (Figure 4.11). The measured optical rotation of this synthetic sample of **4** was -69.6° (c = 0.05, EtOH). The value reported for a sample of **4** isolated from *S. cattleya* was -71.5° (c = 1, H_2O).[1]
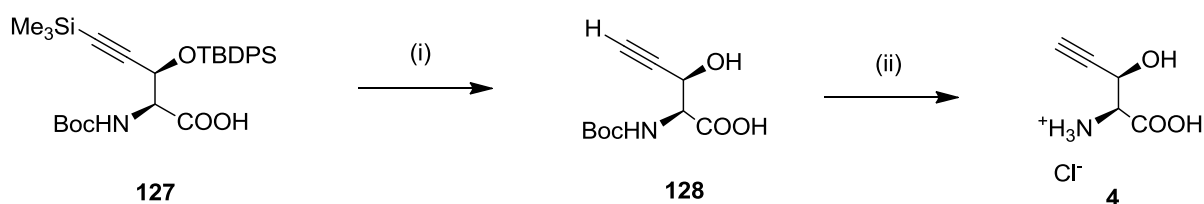
Scheme 4.30. Reagents and conditions: (i) 1. KOH/MeOH 2. Dowex H^+ resin, (ii) conc. HCl (85-92 % two steps)

The product proved to be unstable and it decomposed within a few days in storage, turning into a brown colour. This is consistent with the low titre recovered from isolation experiments with *S. cattleya and S. rolfsii*.[1,3]
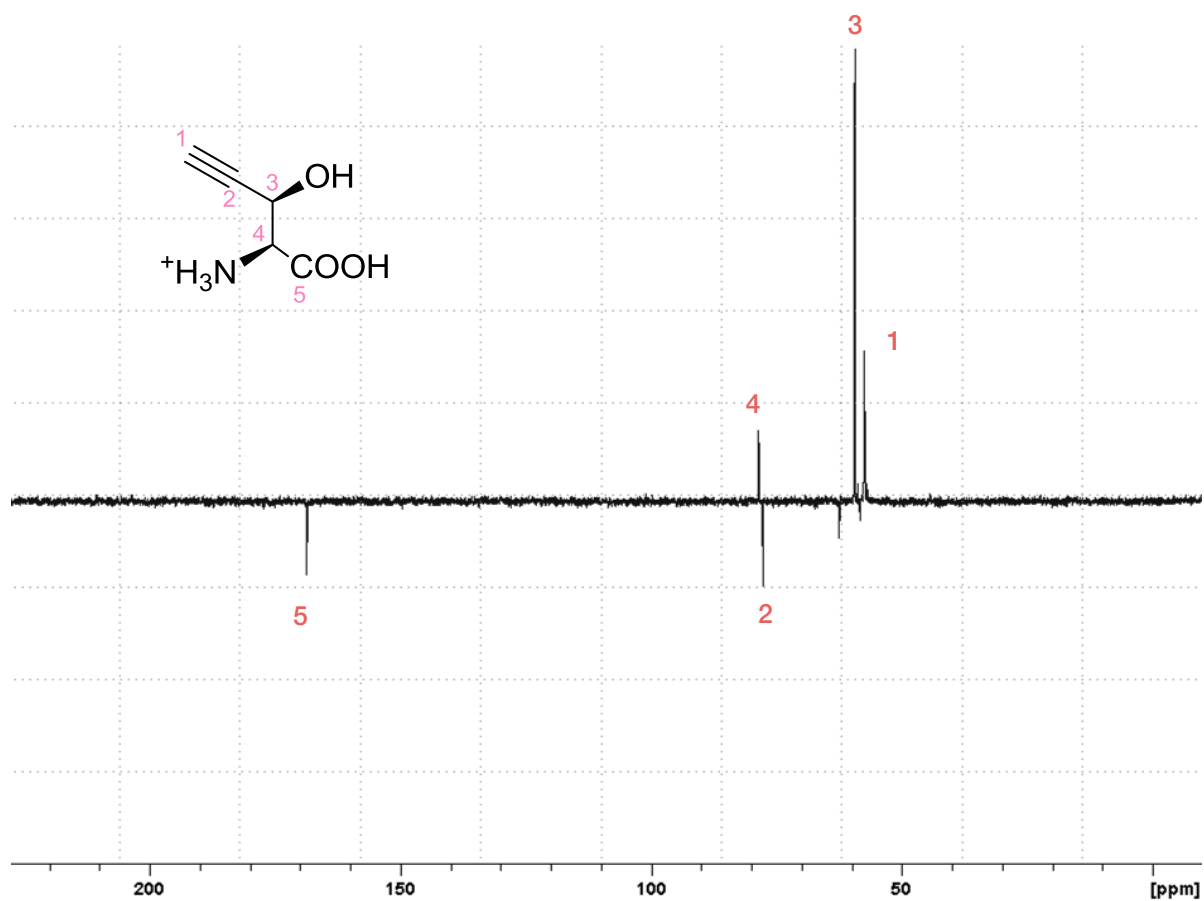
Fig 4.9, *above.* ^{13}C NMR (DEPT, D$_2$O) of **4.**   **Fig 4.10,** *below.* ^{1}H NMR (D$_2$O) of **4.**
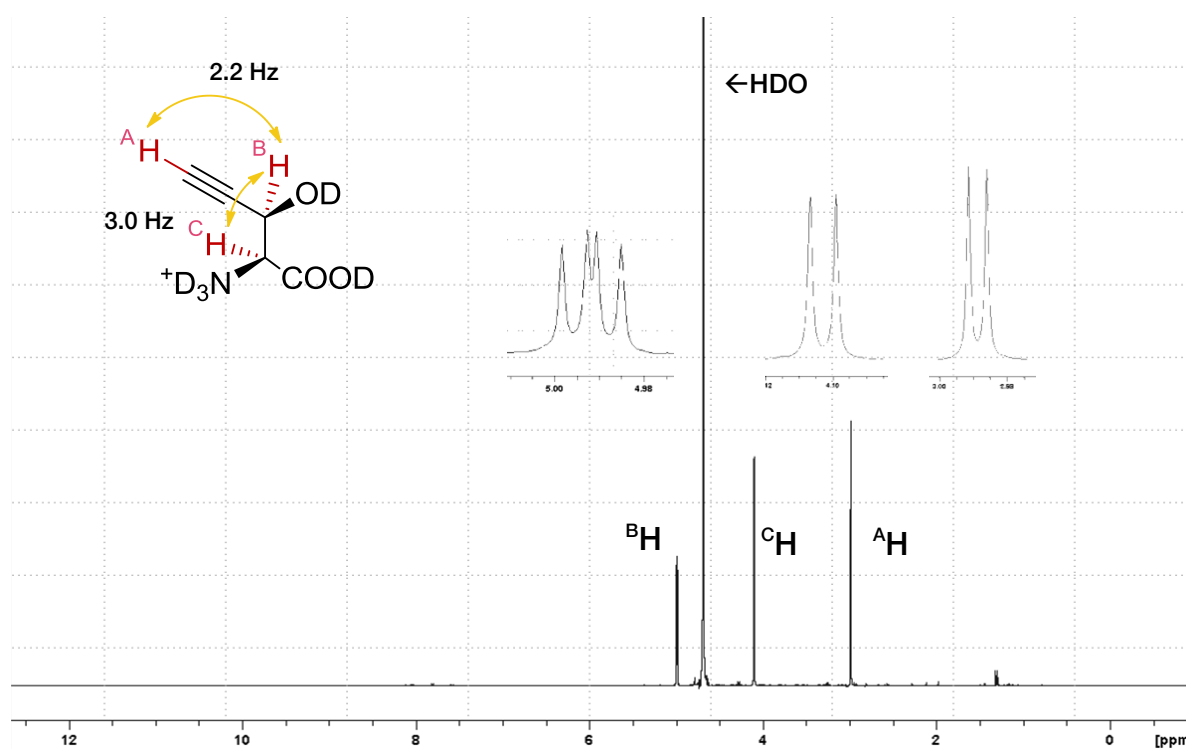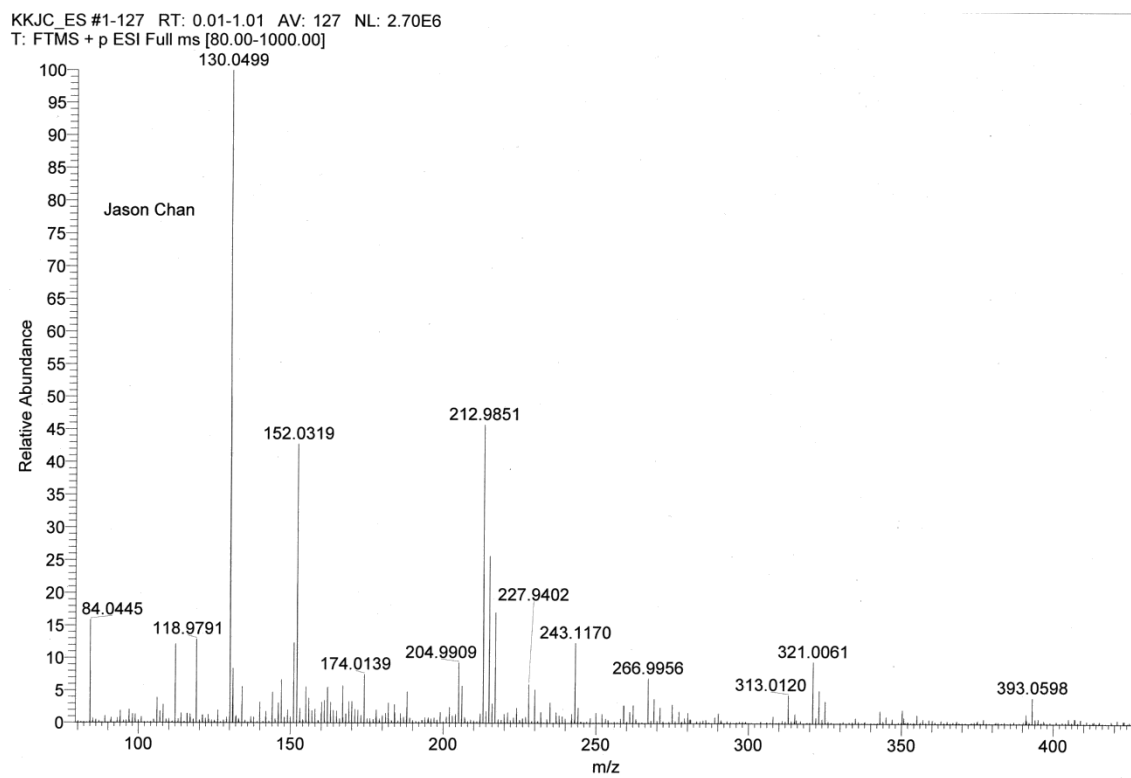
Fig 4.11. HRMS (ESI) of β-ethynylserine **4**.

$[M+H]^+$ calculated $C_5H_8O_3N$ = 130.0499, found 130.0499.

4.5 Using the click reaction as a method to analyse the amino acid from the fermentation media of *S. cattleya*

A reaction was developed to specifically identify β-ethynyl-L-serine in the fermentation media of *S. cattleya* and to convert it into a more stable compound for analysis. The free amino acid is prone to decomposition during the isolation procedures from the fermentation extract, leading to a low recovery. Such a derivatisation can be useful in later studies of the isotope labelling patterns of this natural product.

A Hüisgen 1,3-dipolar cycloaddition reaction,[30] or 'Click' reaction between the alkyne and an azide would be selective for this compound since any other alkynes are unlikely to be present in the fermentation extract. This reaction also converts the unstable acetylenic amino acid into a stable triazole-containing amino acid. Secondly if an UV-active azide is used, it will allow the amino acid to be detected through UV monitoring by HPLC or LC-MS.
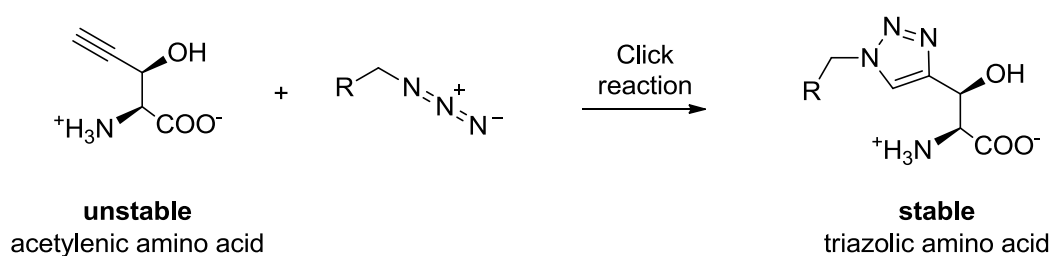
Scheme 4.31. Proposed Click reaction to derivatise the acetylenic amino acid for analytical studies.

α-Azido-*p*-toluic acid **129** was chosen for this task. This azide can be easily prepared from the commercially available bromocarboxylic acid **130**. The carboxylic acid function has a two-fold purpose. Firstly it can be made water-soluble by forming a salt,

and secondly it can be attached if desired to other molecules or solid supports by amide-formation.

α-Azido-p-toluic acid **129** was prepared by refluxing α-bromo-p-toluic acid **130** and sodium azide in methanol for 2 hours. The azidotoluic acid **129** was obtained as a stable white solid in excellent yield.
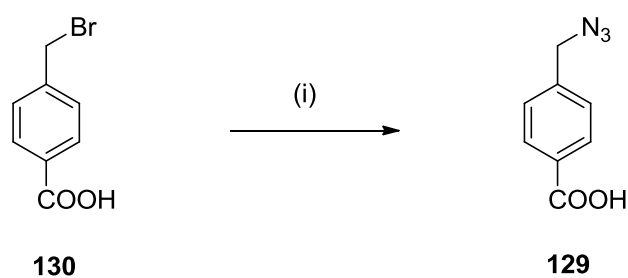
Scheme 4.32. Reagents and conditions: (i) NaN$_3$, MeOH, reflux, 2 h (96 %).

4.5.1 Click reaction of the synthetic amino acid

At the outset, the click reaction between azide **129** and the synthetic sample of β-ethynyl-L-serine was explored under standard click reaction conditions (1 mol % $CuSO_4$ and 10 mol % sodium ascorbate in 50 % DMSO). However this was unsuccessful. Varying the solvents (DMF, EtOH) and increasing catalyst loading did not result in a successful reaction. It was not clear why the reaction did not take place, however other catalysts were explored.

An efficient catalyst for the click reaction, tris-(benzyltriazolylmethyl)amine (TBTA) **131** has been reported.[32] TBTA coordinates copper(II) ions and the Cu-TBTA complex can be reduced by sodium ascorbate to the catalytically active copper(I) oxidation state. The stabilised Cu(I) species is believed to promote the efficient click reaction. Such reactions are typically carried out in 50 % aqueous DMSO in which the Cu-TBTA complex is soluble. In this case, when Cu-TBTA was used as the catalyst, the click reaction proved successful (Scheme 4.33). The products **132/133** were identified by MS, but the regioisomers were not separable even by C18-RP column chromatography (Fig. 4.11).
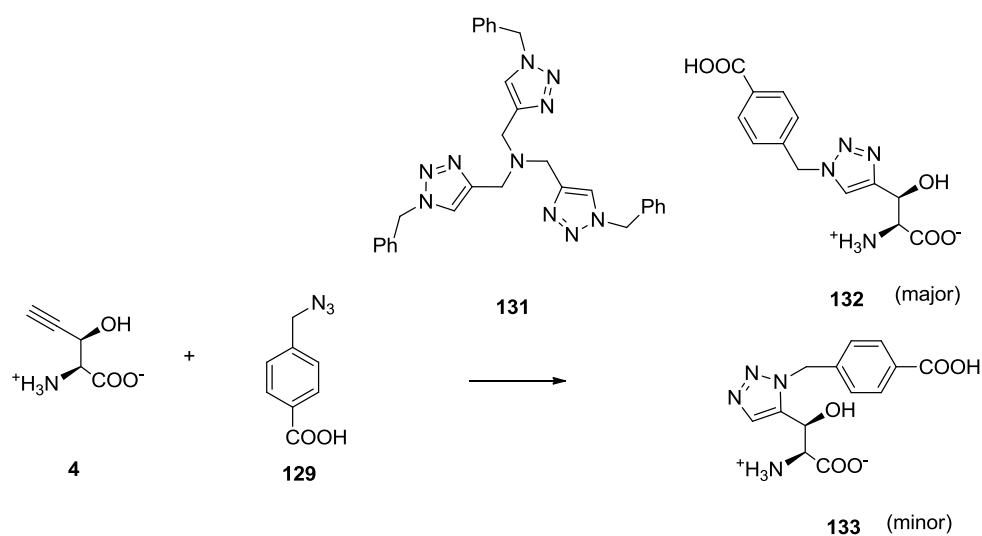
Scheme 4.33. Reagents and conditions: $CuSO_4$ (0.05 eq.), TBTA **131** (0.05 eq.), sodium ascorbate (0.1 eq.), DMSO/H_2O (1 : 1).
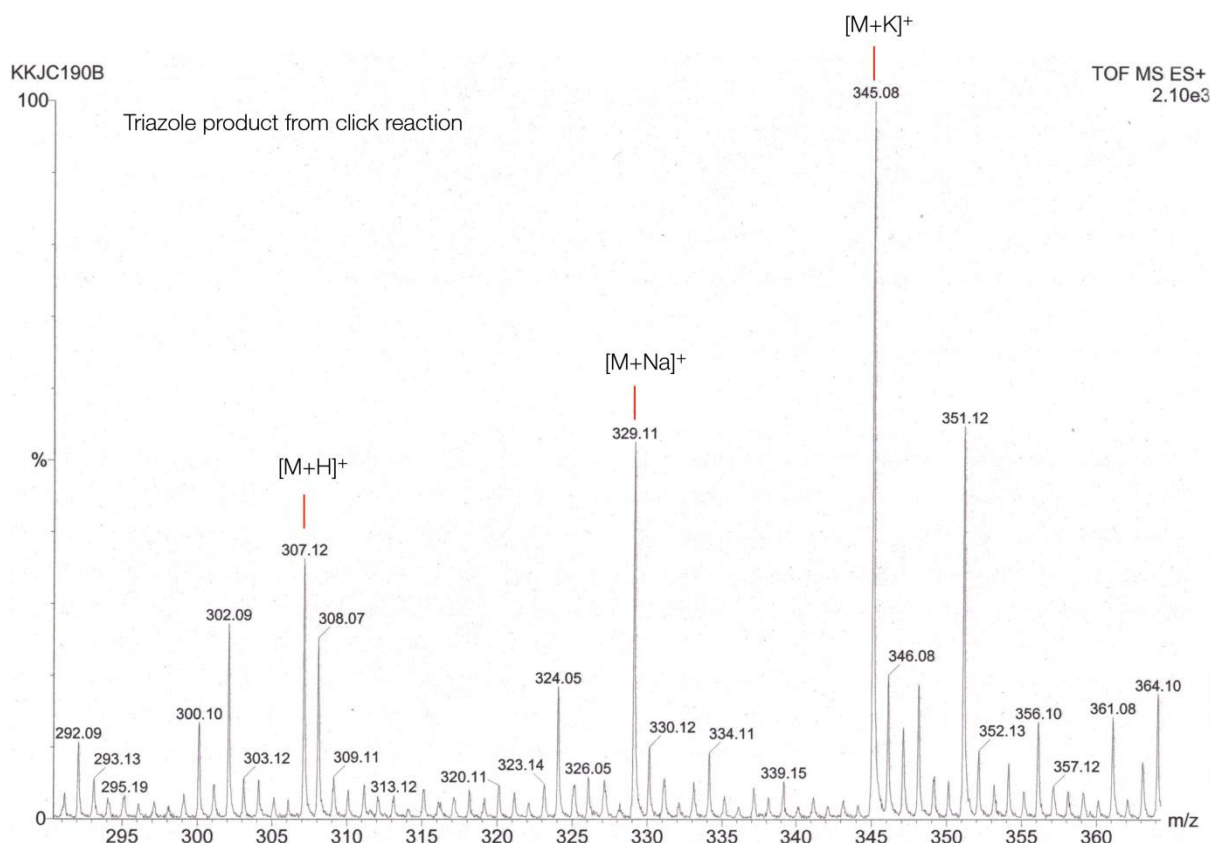
Fig 4.11. Mass spectra of the triazole product **132/133** after RP-C18 column chromatography, showing $[M+H]^+$, $[M+Na]^+$ and $[M+K]^+$.

HPLC conditions (Table 4.2) were optimised for this analysis and the triazole products **132/133** (m/z $[M+H]^+$ = 307) eluted as two peaks, at approximately 14.8 mins and 15.7 mins. They had a weak UV-absorbance peak at 240 nm (Fig 4.12). Mass analysis of the two peaks demonstrated that the signals were consistent with the triazole regioisomers (Fig 4.13). An analytical sample of **132** was prepared by a careful semi-preparative HPLC separation and its ^{1}H-NMR was recorded (Fig. 4.14).

Table 4.2. HPLC conditions optimised for assay of triazoles **132/133**.

Time (mins)	% mobile phase A (95 % MeCN, 4 % H$_2$O, 1 % formic acid)	% mobile phase B (5 % MeCN, 94 % H$_2$O, 1 % formic acid)
0 (start)	0	100
20	20	80
22	0	100
30 (end)	0	100

Flow Rate = 0.4 ml/min

Column = RP- C18 Kinetex 5u XB-C18 100A (Phenomenex) 150mm x 4.60 mm diameter

Fig . 4.12 LCMS of the synthetic triazole reaction mixture. The compounds **132/133** eluted at 14.86 and 15.74 mins. **Top:** UV trace **Bottom:** Mass ion current trace.
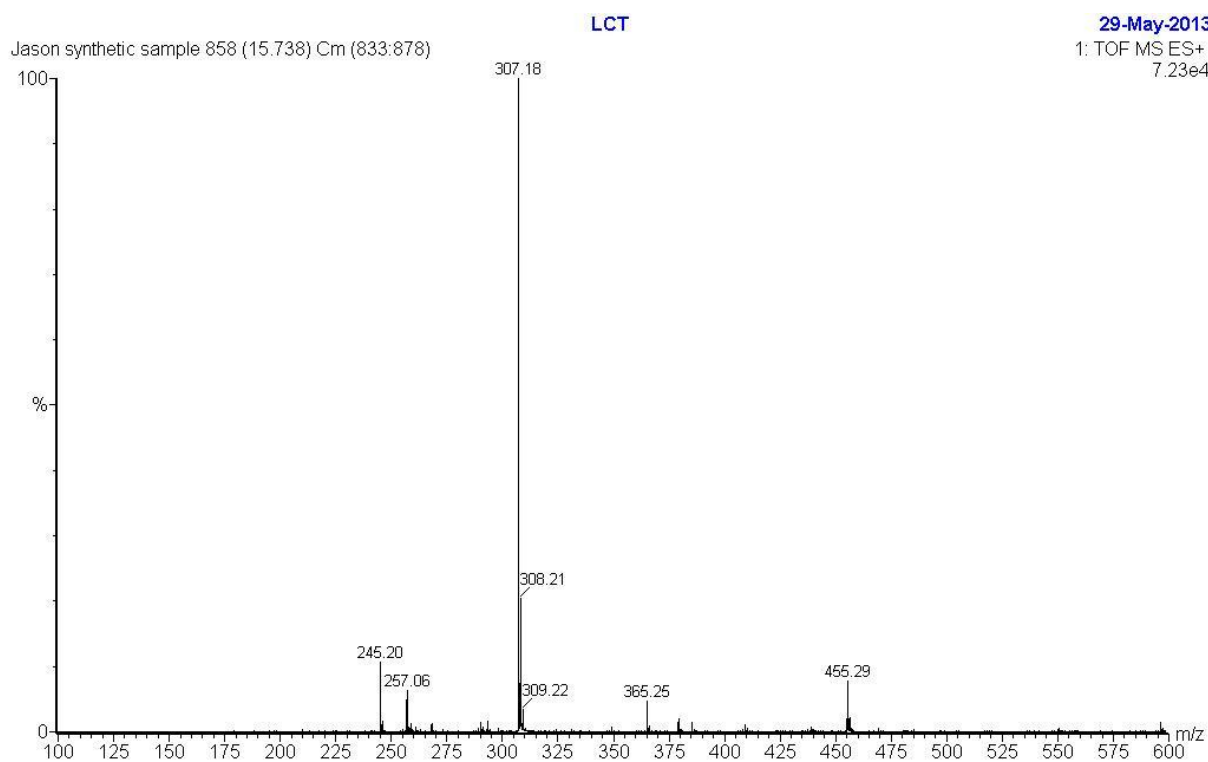
Fig. 4.13 The mass spectrum from the peak at time = 15.74 mins from the LCMS experiment, showing the compounds **132/133** as the major component being eluted with m/z = 307 for the $[M+H]^+$.
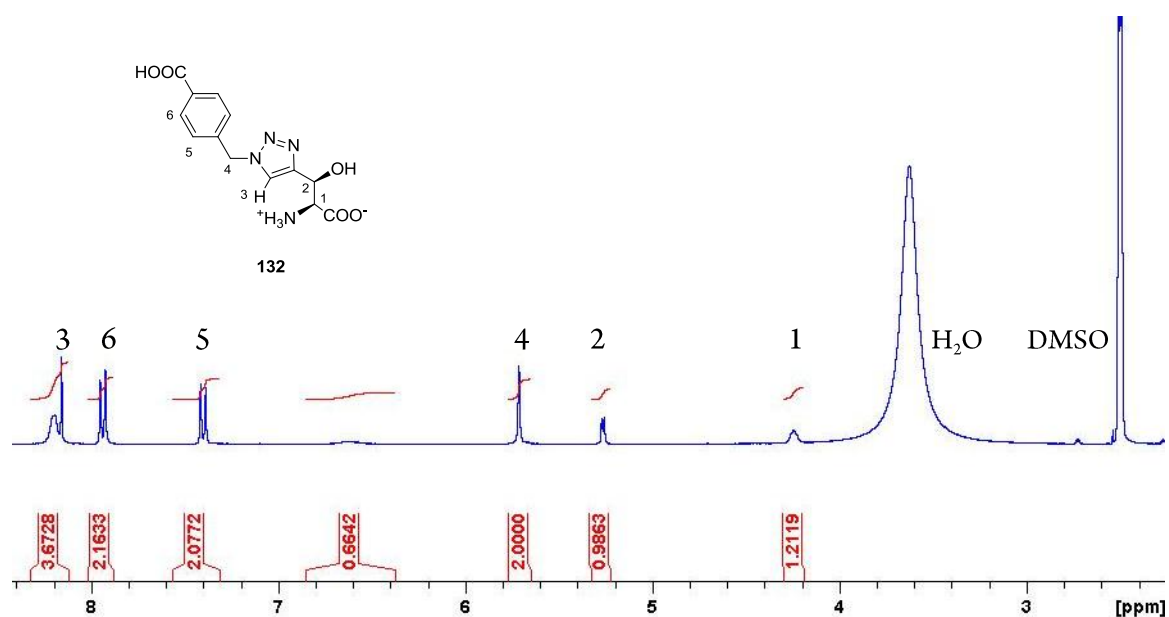
Fig. 4.14 ^{1}H-NMR of triazole **132** obtained by semi-preparative HPLC seperation.

4.5.2 Assay for β-ethynyl-L-serine produced by *S. cattleya*

Having developed the HPLC conditions for the detection of the triazole products **132/133**, the analysis was carried out on the fermentation media of *S. cattleya*. From previous studies, it is known that *S. cattleya* begins its secondary metabolism after 6 days and then the culture will accumulate β-ethynyl-L-serine to a final concentration of *ca.* 0.2 mM after 28 days. It was envisaged that the amino acid could be concentrated by freeze-drying the supernatant of the culture after 28 days and then the click reaction could be explored to derivatise the amino acid present.

A defined fermentation medium[5] was prepared for the optimal growth of *S. cattleya* which was supplemented with potassium fluoride (3 mM). The media was sterilised and the purple spores of *S. cattleya* were used to innoculate this media in sterile flasks. The spores were allowed to grow into a mycelial mass at 30 °C shaking at 150 rpm.

The fermentation cultures were worked up after 30 days and the culture was centrifuged to leave a clear supernatant. The supernatant tested positive for the production of the two fluorometabolites, fluoroacetate and 4-fluorothreonine by ^{19}F-NMR, indicating a healthy secondary metabolism.

The supernatant (400 ml) was lyophilised to give a pale brown powder (approx. 5 g). The powder was suspended in 50 % aqueous DMSO (5 ml) to form a slurry and a solution of 4-azidotoluic acid (200 mg) in DMSO (1 ml) was added. A solution of Cu-TBTA (10 mM, 750 μl) and sodium ascorbate (100 mg) was then added and the mixture was stirred for 16 hours at room temperature. The reaction mixture was then analysed by LC-MS, using the synthetic sample of **132/133** as a reference.

The click reaction mixture carried out with the 30-day extract had two peaks with m/z = 307. They eluted with similar retention times (+/- 0.2 mins) to the reference samples of **132/133** (Fig. 4.15). The mass ion current trace was clear but the UV-absorption was relatively weak. Mass spectra of the two peaks showed masses corresponding to **132/133** (Fig. 4.16). Control experiments omitting one of the two reactants (β-ethynylserine or α-azido-p-toluic acid) were also conducted and in each case, the two peaks belonging to **132/133** were absent in these mixtures (Fig. 4.17).
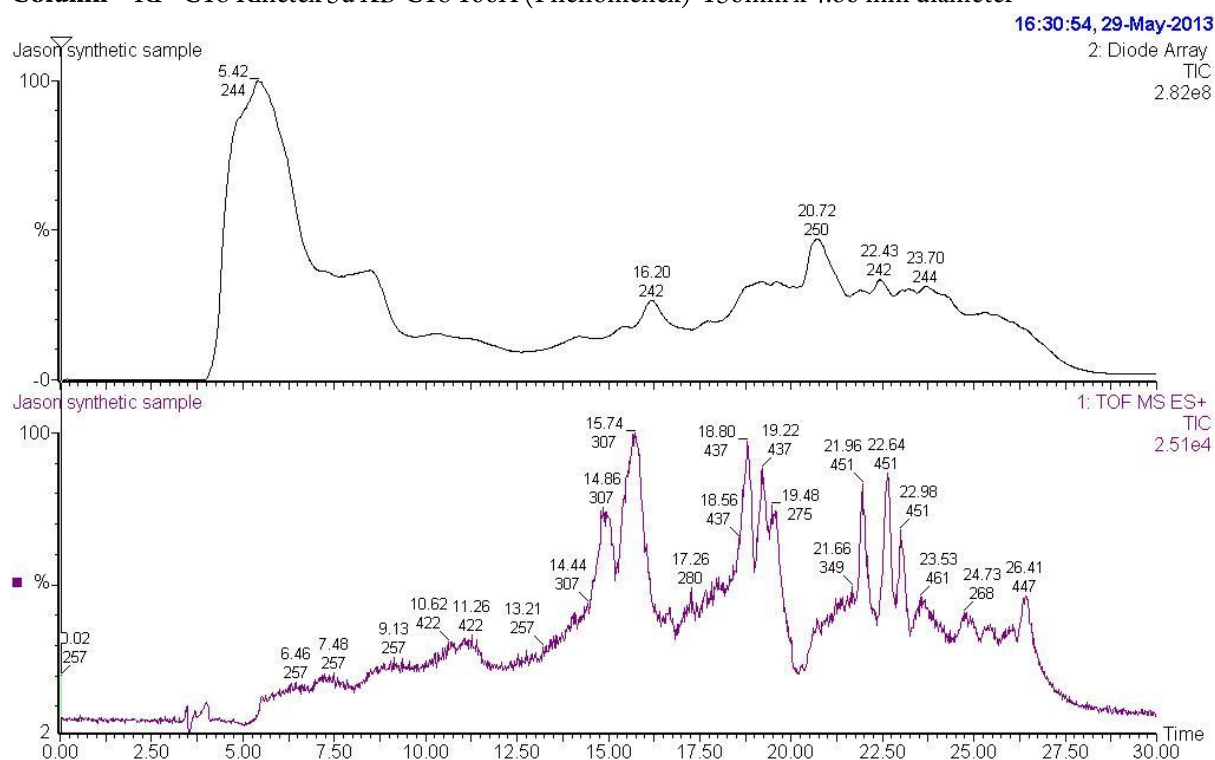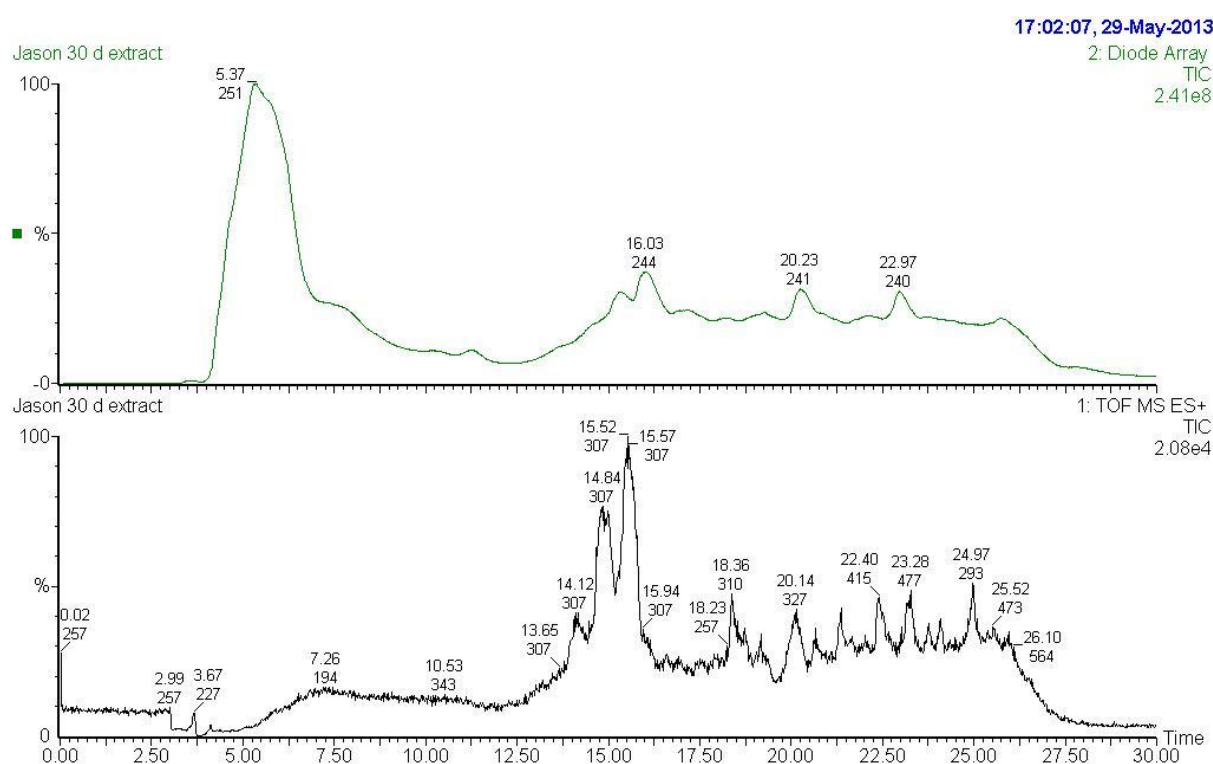

Fig . 4.15 LC-MS of the 30 days extract click reaction mixture. The compounds **132/133** eluted at 14.84 and 15.52 mins. **Top:** UV trace **Bottom:** Mass ion current trace.

**Fig. 4.16** Mass spectra at retention time = 15.5 mins of the 30 days supernatant reaction. The triazoles **132/133** with m/z = 307 is the major component eluting at this time.

Fig. 4.17 Mass ion current traces. From top to bottom: **1.** Control experiment without ethynylserine. **2.** Control experiment without α-azido-*p*-toluic acid. **3.** Click reaction with synthetic ethynylserine **4.** Click reaction with 30 days supernatant reaction.

The click reaction method has proven to be a useful protocol for the detection of β-ethynyl-L-serine **4** from extracts of *S. cattleya.* The conversion of **4** into more stable triazoles **132/133** circumvents degradation of the acetylenic amino acid during work-up. All of the atoms of **4** are conserved in derivatisation to **132/133**, and therefore incorporation of isotopic labels (eg. ^{13}C, ^{2}H, ^{15}N) from biosynthetic feeding experiments should be apparent from LC-MS analyses. The protocol is relatively simple to perform and all of the required reagents are easy to prepare or are commercially available.

4.6 Conclusion

In conclusion, a total synthesis of β-ethynyl-L-serine **4** has been accomplished from D-serine methyl ester *via* D-Garner's aldehyde.

D-Garner's aldehyde **87** was prepared from D-serine methyl ester **106** in 3 steps with an overall yield of 48 – 75 % (Scheme 4.34).
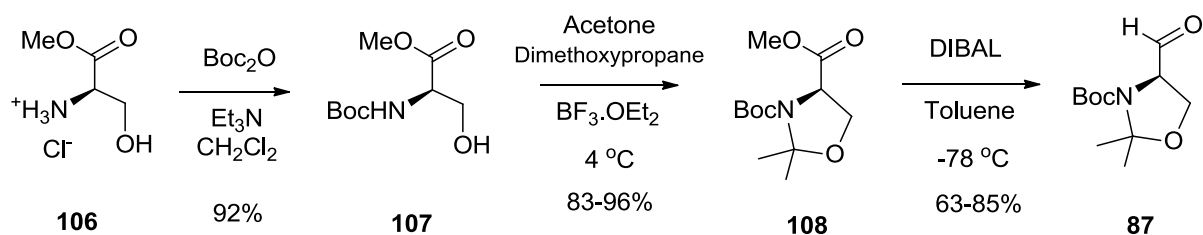
Scheme 4.34. Synthesis of D-Garner's aldehyde **87** from **106**.

β-Ethynyl-L-serine hydrochloride **4** was prepared from D-Garner's aldehyde in 7 steps with an overall yield of 42 % (Scheme 4.35).
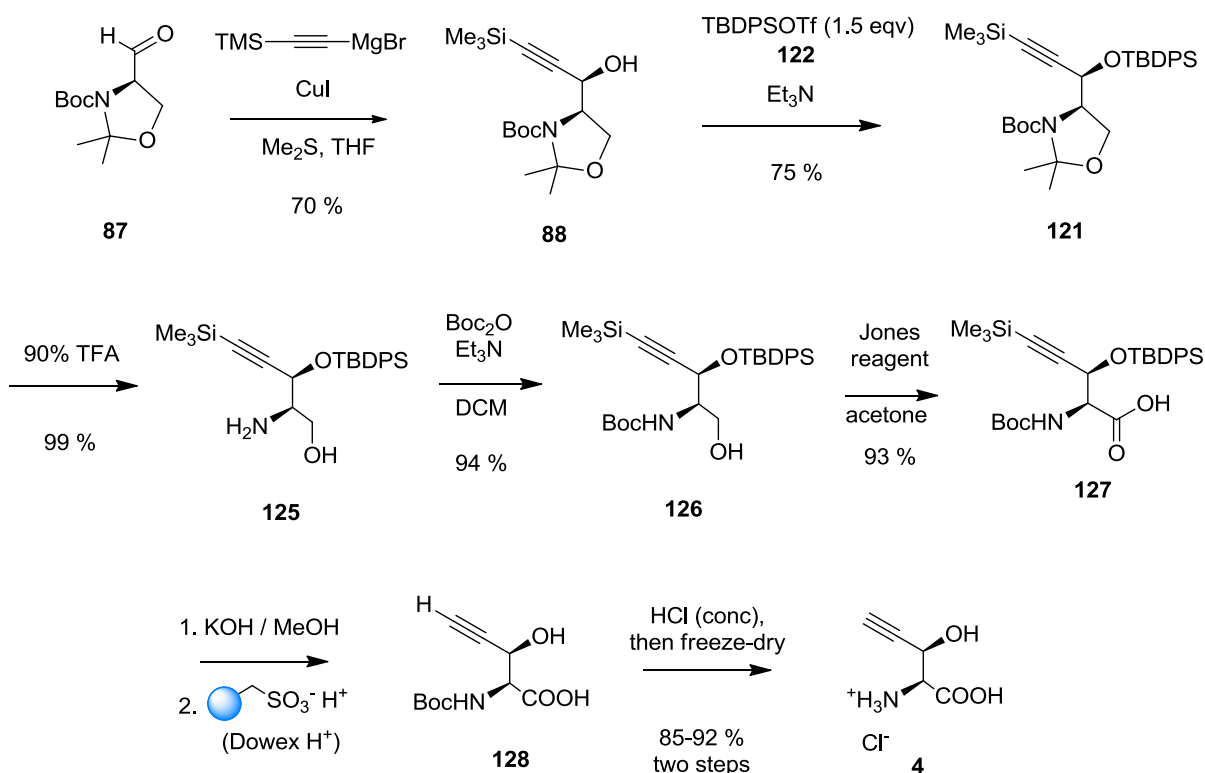
Scheme 4.35. Synthesis of β-ethynyl-L-serine **4** from **87**.

This is the first reported synthesis of the single enantiomer of the free amino acid of β-ethynyl-L-serine **4**. The first synthesis of *rac*-**4** and its diastereoisomers as a mixture by Potgieter *et al.* was a very simple coupling reaction but there was no control of stereochemistry and the yield was low (12 %, mixture of 4 stereoisomers).[3] The synthesis of Fmoc and *tert*-butyl ether protected β-ethynyl-L-serine **86** for Fmoc-based peptide synthesis by Zhang and van der Donk was stereoselective and was achieved in excellent yield (62 % in 8 steps).[9] The success of their synthesis was largely due to the careful choice and placement of protecting groups.

Our synthesis and Zhang's synthesis chose the TBDPS silyl ether protecting group for its stability to acids, and this was crucial to the success of both syntheses. In our synthesis, the TBDPS group was installed on the secondary alcohol prior to an acid deprotection of the *N*-Boc and oxazolidine groups. Zhang and van der Donk installed it on the primary alcohol, to allow an acid-catalysed addition of isobutylene to the alcohol. Installation of the bulky TBDPS group on the secondary alcohol is less efficient than on the primary alcohol. While their synthesis to the protected amino acid had eight steps from D-Garner's aldehyde, our route was optimised for delivery of the free amino acid in seven steps.

A UV-active azide α-azido-*p*-toluic acid **129** was prepared for a click reaction with **4**. The click reaction proved to be efficient using Cu-TBTA and ascorbate catalysts, generating two regioisomeric triazoles **132/133** (Ratio *approx.* 3 : 2 by LC-MS) (Scheme 4.36).
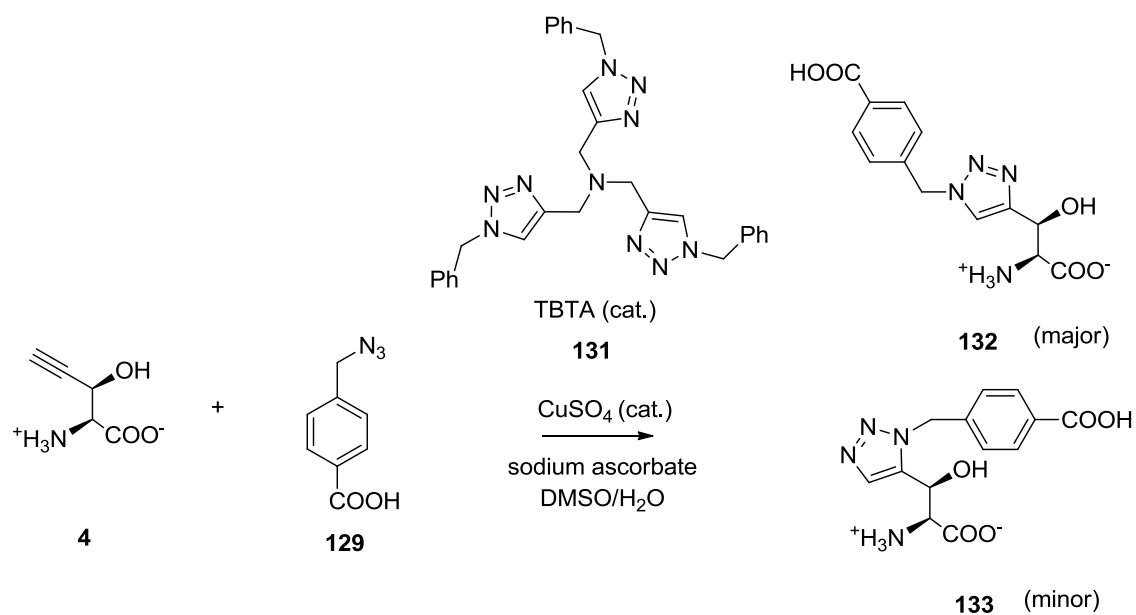
Scheme 4.36. Click reaction for the assay of ethynylserine.

A LC-MS procedure was developed to detect the triazoles **132/133** from the fermentation culture. This protocol can reasonably be applicable for later studies on the biosynthetic origin of this acetylenic amino acid **4** from the soil bacteria *S. cattleya*.

4.7 Chapter 4 References

1. M. Sanada, T. Miyano, S. Iwadare, *J. Antibiotics* 1986, **39**, 304 – 305

2. M. Sanada, T. Miyano, S. Iwadare, J.M. Williamson, B.H.J. Arison, L. Smith, A.W. Douglas, J.M. Liesch, E. Inamine, *J. Antibiotics*, 1986, **39**, 259-265.

3. H. C. Potgieter, N. M. J. Vermeulen, D. J. J. Potgieter, H. F. Strauss, *Phytochemistry* 1977, **16**, 1757 – 1759

4. D. O'Hagan, D. B. Harper, *Nat. Prod. Rep.*, 1994, **11**, 123-133.

5. J. T. G. Hamilton, *Ph. D. thesis*, Queen's University of Belfast, 1998.

6. S. Jitrapakdee, M. St Maurice, I. Rayment, W. W. Cleland, J. C. Wallace, P. V. Attwood, *J. Biochem.*, 2008, **413**, 369-387.

7. R. E. Viola, *Acc. Chem. Res.*, 2001, **34**, 339-349.

8. 8.M. M. Kaplan, M. Flavin, *J. Biol. Chem.*, 1965, **240**, 3928-3933.

9. X. Zhang, W. A. van der Donk, *J. Am. Chem. Soc.*, 2007, **129**, 2212-2213.

10. A. Rincé, A. Dufour, S. Le Pogam, D. Thuault, C. M. Bourgeois, J.-P. Le Pennec, *Appl. Environ. Microbiol.*, 1994, **60**, 1652-1657.

11. P. Uguen, J. P. le Pennec, A. Dufour, *J. Bacteriol.*, 2000, **182**, 5262-5266.

12. H. W. van den Hooven, F. M. lagerwerf, W. Heerma,, J. Haverkamp, J-C Piard, C. W. hibers, R. J. Siezen, O. P. Kuipers, H. S. Rollema, *FEBS Lett.*, 1996, **391**, 317-322.

13. P. Herold, *Helv. Chim. Acta*, 1988, **71**, 354-362

14. F. D'Aniello, A. Mann, M. Taddei, C-G Wermuth, *Tetrahedron Lett.*, 1994, **35**, 7775-7778.

15. Yu. N. Belokon, A. G. Bulychev, S. V. Vitt, Yu. T. Struchkov, A. S. Batsanov, T. V. Timofeeva, V. A. Tsyryapkin, M. G. Ryzhov, L. A. Lysova, *J. Am. Chem. Soc.*, 1985, **107**, 4252-4259.

16. V. A. Soloshonok, D. V. Avilov, V. P. Kukhar, V. I. Tararov, K. A. Kochetkov, T. F. Saveleva, T. D. Churkina, N. S. Ikonnikov, S. A. Orlova, A. P. Pysarevsky, Y. T. Stnichkov, N. I. Raevsky, Y. N. Belokon, *Tetrahedron: Asymmetry*, 1995, **6**, 1741-1756.

17. Y. N. Belokoń, A. G. Bulychev, V. A. Pavlov, E. B. Fedorova, V. A. Tsyryapkin, V. A. Bakhmutov and V. M. Belikov, *Perkin Trans I*, **1988**, 2075-2083.

18. H. Ueki, T. K. Ellis, C. H. Martin, T. U. Boettiger, S. B. Bolene, V. A. Soloshonok, *J. Org. Chem.*, 2003, **68**, 7104-7107.

19. J. C. Sauer, *Org. Syn. Coll.*, 1963, **4**, 813.

20. P. Garner, *Tetrahedron Lett.*, 1984, **25**, 5855.

21. X. Liang, J. Andersch, M. Bols, *Perkin Trans I*, **2001**, 2136-2157.

22. P. Garner, J. M. Park, *J. Org. Chem.*, 1987, **52**, 2361.

23. A. McKillop, R. J. K. Taylor, R. J. Watson, N. Lewis, *Synthesis*, **1994**, 31-33

24. T. Moriwake, S. Hamano, S. Saito, S. Torri, *Chem. Lett.* **1987**, 2085

25. D. E. Frantz, R. Fassler, E. M. Carreira, *J. Am. Chem. Soc.*, 2000, **122**, 1806-1807

26. D. Boyall, D. Frantz, E. M. Carreira, *Org. Lett.*, 2002, **4**, 2605-2606.

27. C. B. Reese, J. C. M. Stewart, *Tetrahedron Lett.*, 1968, **9**, 4273-4276.

28. T. W. Greene, P. G. M. Wuts, *Protective Groups in Organic Synthesis 3rd ed.*, Wiley & Sons, 141-143.

29. P. Dimopoulos, J. George, D. A. Tocher, S. Manaviazar, K. J. Hale, *Org. Lett.*, 2005, **7**, 5377-5380.

30. K. Bowden, I. M. Heilbron, E. R. H. Jones, *J. Chem. Soc.*, 1946, 39-45.

31. R. Huisgen, G. Szeimies, L. Mobius, *Chem. Ber.*, 1967, **100**, 2494.

32. T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Org. Lett.*, 2004, **6**, 2853-2855.

5. Experimental Section

5.1 Chemical Synthesis

5.1.1 General Methods

Reagents were obtained from commercial sources (Acros Organic, Sigma Aldrich) and used without further purification. DCM, THF, toluene and ether were dried by passing through two purification columns in a MBRAUN SPS-800 Solvents Purification System. Anhydrous DMF, DMSO, pyridine were obtained from commercial sources. All reactions were performed under positive pressure of dry N_2 or argon using oven-dried or flame-dried glassware unless otherwise stated.

Reactions were monitored by TLC carried out on pre-coated plastic-backed silica gel plates (Machery-Nagel Polygram® SIL G/UV254) and visualised under UV light (254 nm) or by staining ($KMnO_4$). Flash column chromatography was carried out using Silica Gel 60 (0.040-0.063 mm).

^{1}H and ^{13}C NMR spectra were recorded at 298 K on a Bruker Avance 300 MHz or 400 MHz spectrometer and calibrated to the solvent proton/carbon signals. Chemical shifts are reported in ppm (δ), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, br = broad signal), coupling constants in Hz and integration. Assignments were based on both one- and two-dimensional experiments (COSY, HSQC).

Electrospray Ionisation Mass Spectrometry analyses were performed on a Micromass LCT-TOF mass spectrometer. Infrared (IR) Spectra were recorded on a Perkin-Elmer Paragon series 1000FTIR spectrometer. HRMS were recorded by electrospray ionisation at the Mass Spectrometry Unit of the University of St Andrews.

Optical rotation determination was performed using a Perkin Elmer Model 341 polarimeter, at 589 nm and given in 10^{-1}deg.cm^2.g^{-1}. Melting points were uncorrected.

Single crystal X-ray diffraction analyses were carried out by Prof. A. M. Z. Slawin. Measurements were made on a Rigaku Saturn70 diffractometer with Cu-Kα radiation.

5.1.2 Compounds preparation and characterisation
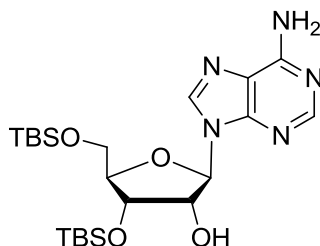
3',5'-Bis-*O,O*-(*tert*-butyldimethylsilyl)adenosine 68

TBDMSCl (3.72 g, 24.7 mmol) was added to a stirred solution of adenosine (3.0 g, 11.2 mmol) and imidazole (3.36 g, 49.4 mmol) in dry DMF (25 ml) at r.t., and the mixture was stirred at r.t. for 14 h. The mixture was then poured into brine (250 ml). The mixture was extracted into DCM (3 x 150 ml). The organic extracts were combined, dried (MgSO$_4$), filtered and the solvents were removed *in vacuo* at 70 °C. The mixture was separated by silica gel flash column chromatography, eluting with hexane : ethyl acetate : *i*-PrOH (5.6:4.0:0.4). The 2',3',5'-tris-*O*-silylated compound eluted first, followed by the 2',5'-bis-*O*-silylated isomer, then fractions containing both the 2',5'-bis-*O*-silylated isomer and the desired 3',5'-bis-*O*-silylated product **68** and followed by fractions containing pure **68**. The mixture containing the 2',5'-bis-*O*-silylated isomer and **68** was re-applied onto a second silica gel column. Fractions containing **68** as a single component were combined and evaporated *in vacuo* to give the title compound as a white solid (3.29 g, 25 %).

m.p. 193-194 °C, R$_f$ = 0.23 (hexane : EtOAc : *i*-PrOH, 5.6 : 4.0 : 0.4); $[\alpha_D]^{20}$ -35.5° (*c* = 0.5, CHCl$_3$); ^{1}H-NMR (400 MHz, CDCl$_3$) δ_H 8.34 (s, 1H, H-8), 8.10 (s, 1H, H-2), 6.04 (d, *J* = 4.3 Hz, 1H, H-1'), 5.92 (br, s, 2H, NH$_2$), 4.13 (br s, 2H, H-2' and 3'), 3.93

(dd, J = 3.5, 11.4 Hz, 1H, H-5'), 3.77 (dd, J = 2.9, 11.4 Hz, 1H, H-5'), 3.43 (br, 1H, OH), 0.95 (s, 9H, 5'-SiC(CH$_3$)$_3$), 0.90 (s, 9H, 3'- SiC(CH$_3$)$_3$), 0.17 (s, 6H, 2 x 5'-SiCH$_3$), 0.09 & 0.07 (s & s, 6H, 2 x 3'-SiCH$_3$); ^{13}C-NMR (75 MHz, CDCl$_3$) δ_C 160.0 (C-6), 153.4 (C-2), 150.1 (C-4), 139.5 (C-8), 120.2 (C-5), 89.2 (C-1'), 85.9 (C-4'), 75.6 (C-3'), 72.1 (C-2'), 62.8 (C-5'), 26.3, 26.2 (2 x (**CH**$_3$)$_3$C), 18.8, 18.5 (2 x (CH$_3$)$_3$**C**), -4.2, -4.4, -5.0, -5.1 (4 x CH$_3$Si); IR (nujol mull) v_{max}/cm^{-1} 3432, 3329, 3216, 2903, 1636, 1600, 1584, 1465; ES-MS m/z 496.19 [M+H]$^+$, 518.21 [M+Na]$^+$. These data are in accordance with the literature.[1]
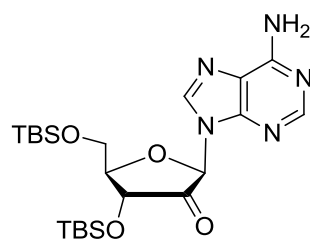
9-[3,5-Bis-*O*,*O*-(*tert*-butyldimethylsilyl)-β-D-*erythro*-pentofuran-2-ulosyl]adenine 69

Anhydrous pyridine (0.25 ml, 3 mmol) was added to a stirred suspension of finely divided chromium trioxide (151 mg, 1.51 mmol) in DCM (10 ml) at 0 °C, followed by acetic anhydride (0.143 ml, 1.51 mmol). The mixture was stirred at 0 °C for 20 mins, after which a solution of **68** (375 mg, 0.76 mmol) in DCM (30 ml) was added. The mixture was stirred at 0 °C for 20 mins and then at r.t. for 5 h.

The mixture was then diluted with ethyl acetate (200 ml) and filtered through a pad of celite (4 cm). The celite pad was washed with ethyl acetate (20 ml) and the filtrates were concentrated *in vacuo*. The residues were purified by flash chromatography through a short silica column, firstly washing with hexane:ethyl acetate 1 : 1 (100 ml) followed by elution with ethyl acetate. **69** was co-eluted with some unreacted **68**.

Therefore the mixture was treated again under the same oxidation procedures (CrO_3 (75 mg), pyridine (0.123 ml) and Ac_2O (0.07 ml)) and purification protocols to obtain complete conversion, giving **69** as a yellow oil (280 mg, 75 %).

^{1}H-NMR (400 MHz, $CDCl_3$) δ_H 8.02 (s, 1H, H-8), 7.65 (s, 1H, H-2), 6.03 (br s, 2H, NH_2), 5.62 (s, 1H, H-1'), 4.98 (d, 1H, J = 9.0 Hz, H-3'), 3.99 (ddd, 1H, J = 3.0 Hz, 3.1 Hz, 9.0 Hz, H-4'), 3.85 (dd, 1H, J = 3.1 Hz, 11.2 Hz, H-5'), 3.72 (dd, 1H, J = 3.0 Hz, 11.2 Hz, H-5'), 0.75 (s, 9H, 5'-SiC(CH$_3$)$_3$), 0.63 (s, 9H, 3'-SiC(CH$_3$)$_3$), 0.04 & 0.00

(s & s, 6H, 2 x 5'-SiCH$_3$), -0.17 & -0.27 (s & s, 6H, 2 x 3'-SiCH$_3$); ES-MS *m/z* 494.20 [M+H]$^+$, 516.13 [M+Na]$^+$.

These data are in accordance with the literature.[1]
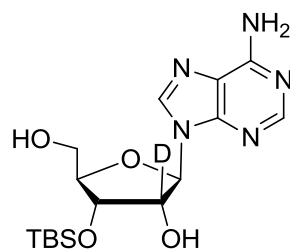
3'-*O*-*tert*-Butyldimethylsilyl-[2'-²H]adenosine 71

90 % Aqueous TFA (5 ml) was added to **69** (311 mg, 0.63 mmol) at 0 °C and the mixture was allowed to stir at 0 °C for 1 hour. The reaction was quenched by the slow addition of ice-cold saturated $NaHCO_3$ solution until tested neutral to blue litmus paper. The mixture was extracted with ethyl acetate (2 x 100 ml), the organic extracts were combined, washed with water (50 ml) and brine (50 ml), dried ($MgSO_4$), filtered and evaporated *in vacuo* to give crude **70** as a white solid (230 mg, 96 %).

The product containing **70** was immediately added a solution of sodium triacetoxyborodeuteride (4.78 mmol), prepared immediately prior to use by the addition of sodium borodeuteride (200 mg, 4.78 mmol) to a solution of glacial acetic acid (18.4 ml) in ethyl acetate (50 ml) at 0 °C, and stirring for 30 mins. The reaction mixture was stirred at r.t. for 1.5 hours, after which it was concentrated *in vacuo*. The residues were neutralised with saturated aqueous sodium bicarbonate solution (*ca.* 200 ml) and then extracted with ethyl acetate (100 ml x 2). The organic extracts were combined, dried ($MgSO_4$), filtered and evaporated *in vacuo* to afford **71** as off-white crystals (204 mg, 88 %).

¹H-NMR (400 MHz, CD_3OD) δ_H 8.25 (s, 1H, H-8), 8.13 (s, 1H, H-2), 5.91 (s, 1H, H-1'), 4.40 (d, 1H, *J* = 3 Hz, H-3'), 4.10 (q, 1H, *J* = 3 Hz, H-4'), 3.83 (d, 1H, *J* = 3 Hz,

H-5'), 3.80 (d, 1H, J = 3 Hz, H-5'), 0.92 (s, 9H, SiC(CH$_3$)$_3$), 0.91 & 0.85 (s & s, 6H, 2

x 3'- SiCH$_3$); ^{13}C-NMR (100 MHz, CD$_3$OD) δ_C 157.6 (C-6), 153.5 (C-2), 150.1 (C-4), 142.1 (C-8), 121.0 (C-5), 90.9 (C-1'), 89.1 (C-4'), 74.5 (t, C-2'), 74.4 (C-3'), 63.3 (C-5'), 26.4 (C(**CH**$_3$)$_3$), 19.3 (**C**(CH$_3$)$_3$), -4.4 & -4.8 (2 x CH$_3$Si); ES-MS *m/z* 383.08 [M+H]$^+$, 405.07 [M+Na]$^+$.

These data are in accordance with the literature.[2]
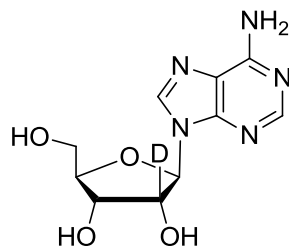
[2'-²H]Adenosine 66

A solution of **73** (158 mg, 0.413 mmol) and ammonium fluoride (80 mg, 2.16 mmol) in methanol (50 ml) was heated under reflux for 2 hours.

The solvents were then removed *in vacuo* and the residues were partitioned in deionised water (100 ml) and ethyl acetate (100 ml). The ethyl acetate layer was further extracted with deionised water (50 ml). The combined aqueous extracts were stirred with pre-washed Ambersep 900 OH resin (1 g) for 15 mins, then the solution was filtered and freeze-dried to give **66** as an off-white powder (110 mg, quant.).

m.p. 230-232 °C; $[\alpha_D]^{20}$ -54.8° (c = 0.2, H_2O); ^{1}H-NMR (400 MHz, DMSO-d_6) δ_H 8.38 (s, 1H, H-8), 8.17 (s, 1H, H-2), 7.38 (br s, 2H, NH_2), 5.91 (s, 1H, H-1'), 5.46 (br s, 2H, OH), 5.22 (br s, 1H, OH), 4.17 (d, 1H, J = 3 Hz, H-3'), 3.99 (q, 1H, J = 3 Hz, H-4'), 3.71 (dd, 1H, J = 3 Hz, 12 Hz, H-5"), 3.58 (d, 1H, J =12 Hz, H-5'); ^{13}C-NMR (100 MHz, DMSO-d_6) δ_C 156.1 (C-6), 152.3 (C-2), 149.0 (C-4), 139.9 (C-8), 119.3 (C-5), 87.8 (C-1'), 85.9 (C-4'), 73.6 (t, C-2'), 70.6 (C-3'), 61.6 (C-5'); ES-MS *m/z* 269.10 [M+H]⁺, 291.08 [M+Na]⁺; HRMS (calculated for [M+H]⁺ $C_{10}H_{13}DN_5O_4$ = 269.1103) found 269.1099. These data are in accordance with the literature.[2]
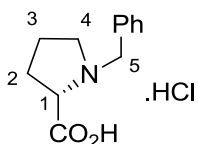
N-Benzyl L-proline hydrochloride 99

L-Proline (17.3 g, 0.15 mol) and potassium hydroxide pellets (25.4 g, 0.45 mol) were added to isopropanol (100 ml) and the mixture was stirred at 40 °C until the solid had dissolved. Then benzyl chloride (18.9 ml, 0.16 mol) was added slowly via a dropping funnel over 2 h while the temperature of the reaction was maintained at 40-50 °C. The mixture was stirred overnight at 40 °C.

The mixture was cooled in an ice-bath, and the pH was adjusted to 6 with conc. HCl. Chloroform (40 ml) was added and the mixture was stirred for 2 h at r.t., then the mixture was filtered to remove precipitated KCl and the residues were washed with $CHCl_3$ (20 ml). The combined filtrate was evaporated *in vacuo* to a viscous brown oil. To this oil was added acetone (500 ml) and the mixture was stirred for 10 mins, which produced a white precipitate. The precipitate was collected by filtration, washed with acetone (100 ml) and dried *in vacuo* over $CaCl_2$ to give **99** as an off-white solid (19.4 g, mixture of the free amino acid and HCl salt, 63 % based on the free amino acid).

m.p. 178-179 °C; $[\alpha_D]^{20}$ -27.4° (*c* = 1, EtOH); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 8.09 (br, 1H, COOH), 7.49-7.46 (m, 2H, ArH), 7.37-7.35 (m, 3H, ArH), 4.36 (q, 2H, *J* = 13.0 Hz, $PhCH_2$), 3.92-3.87 (m, 1H, NCH_2), 3.76-3.68 (m, 1H, NCH_2), 3.03-2.94 (m, 1H, NCH), 2.38-2.19 (m, 2H, CH_2), 2.07-1.88 (m, 2H, CH_2); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 171.0 (COOH), 130.7 (2 x CH, Ph), 130.4 (C, Ph), 129.5 (CH, Ph), 129.1 (2 x CH, Ph), 67.0 (CH, C-1), 57.4 (CH_2, C-5), 53.2 (CH_2, C-4), 28.7 (CH_2, C-2), 22.6 (CH_2, C-3). ES-MS *m/z* 206.23 $[M+H]^+$, 228.22 $[M+Na]^+$.
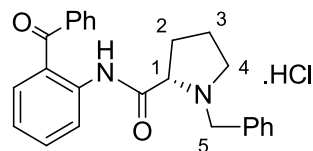
(*S*)-2-[*N*-(*N'*-Benzylprolyl)amino]-benzophenone hydrochloride 96

A solution of *N*-benzyl L-proline (2.05 g, 10 mmol) and *N*-methylimidazole (2.75 ml, 34 mmol) in CH_2Cl_2 (50 ml) was cooled to 0 °C. Methanesulfonyl chloride (1 ml, 10 mmol) was added dropwise over 10 mins and the mixture was stirred for 10 mins. Then *o*-aminobenzophenone (1.78 g, 9 mmol) was added in one portion and the mixture was heated to 45 °C and stirred for 16 h under reflux.

The mixture was diluted with CH_2Cl_2 (100 ml) and poured into sat. NH_4Cl solution (100 ml). The organic layer was collected, dried (Na_2SO_4) and the solvents were removed *in vacuo*. The residues were dissolved in acetone (25 ml) and conc. HCl (1.66 ml, 20 mmol) was added. The mixture was left to stand at 4 °C overnight and the crystalline product was collected by filtration, washed with ice-cold acetone (10 ml) and dried over $CaCl_2$ *in vacuo*. **96** was obtained as light-yellow crystals (3.03 g, 88 %).

m.p. 101-102 °C; $[\alpha_D]^{20}$ -131.2° (*c* = 0.5, $CHCl_3$); ^{1}H NMR (400 MHz, CD_3OD) δ_H 7.79-7.76 (m, 2H, ArH), 7.67-7.57 (m, 2H, ArH), 7.53-7.46 (m, 5H, ArH), 7.41-7.37 (m, 4H, ArH), 4.35-4.32 (m, 2H, PhCH$_2$), 3.61-3.58 (m, 1H, NCH), 3.34-3.31 (m, 2H, CH$_2$), 2.47-2.37 (m, 1H, CH$_2$), 2.21-2.08 (m, 1H, CH$_2$), 1.95-1.79 (m, 1H, CH$_2$), 1.67-1.55 (m, 1H, CH$_2$); ^{13}C NMR (125 MHz, CD_3OD) δ_C 197.7(C=O, ketone), 167.4 (C=O, amide), 138.7 (C, Ar), 136.1 (C, Ar), 134.5 (CH, Ar), 133.3 (CH, Ar), 132.2 (2 x CH, Ar), 131.7 (CH, Ar), 131.4 (2 x CH, Ar), 131.3 (2 x C, Ar), 130.4 (2 x CH, Ar), 129.7 (2 x CH, Ar), 127.1 (CH, Ar), 125.8 (CH, Ar), 68.2 (CH, C-1), 59.4 (CH$_2$, C-5), 56.0 (CH$_2$, C-4), 29.7(CH$_2$, C-2), 23.9 (CH$_2$, C-3); ES-MS *m/z* 385.39 [M+H]$^+$, 407.41 [M+Na]$^+$. These data are in accordance with the literature.[3]
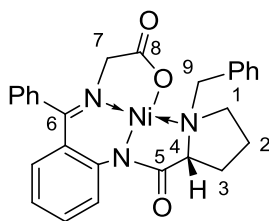
Ni(II)-complex of glycine Schiff base with (*S*)-2-[N-(N'-benzoylprolyl)amino]-benzophenone 97

Ni-(*S*)-BPB-Gly

To a stirred solution of **96** (2.53 g, 6 mmol), glycine (2.25 g, 30 mmol) and nickel(II) nitrate hexahydrate (3.49 g, 12 mmol) in methanol (30 ml) at 50 °C was added a solution of sodium hydroxide (1.92 g, 48 mmol) in methanol (15 ml). The mixture was stirred at 50 °C for 45 mins, then glacial acetic acid (2.7 ml) was added in one portion, and at once ice-cold water (50 ml) was added to the mixture. When the mixture began to separate, a few small crystals of **97** were added as seeds to initiate crystallisation†. The mixture was then left to stand to allow crystallisation to complete and the product was collected by filtration with a Bünchner funnel, washed with water and dried over $CaCl_2$ *in vacuo* to afford **97** as a bright-red crystalline powder (2.5 g, 84 %).

m.p. 207-209 °C; ^{1}H NMR (400 MHz, CD_3OD) δ_H 8.31-8.30 (m, 2H, ArH), 7.93-7.90 (m, 1H, ArH), 7.52-7.43 (m, 3H, ArH), 7.38-7.32 (m, 2H, ArH), 7.21-7.02 (m, 4H, ArH), 6.70-6.59 (m, 2H, ArH), 4.12 (d, *J* = 12.4 Hz, 1H, NCH), 3.70 (dd, *J* = 20.1, 66.4 Hz, 2H, CH_2), 3.51-3.44 (m, 2H, CH_2), 3.25 (s, 2H, CH_2), 2.53-2.31 (m, 2H, CH_2), 2.20-2.01 (m, 2H, CH_2); ^{13}C NMR (125 MHz, CD_3OD) δ_C 183.7 (C=O, C-8), 180.8 (C=O, C-5), 173.6 (C=N, C-6), 143.3 (C, Ar), 136.5 (C, Ar), 135.3 (C, Ar), 134.3 (CH, Ar), 133.0 (2 x CH, Ar), 131.0 (t, *J* = 29 Hz, 2 x CH, Ar), 130.2 (t, *J* =

29 Hz, 2 x CH, Ar), 130.2 (CH, Ar), 128.0 (CH, Ar), 127.3 (CH, Ar), 127.1 (C, Ar), 125.7 (2 x CH, Ar), 122.4 (2 x CH, Ar), 72.3 (CH, C-4), 65.2 (CH$_2$, C-9), 62.2 (CH$_2$, C-7), 59.7 (CH$_2$, C-1), 55.5 (C, C-3), 32.0 (CH$_2$, C-3), 24.8 (CH$_2$, C-2); ES-MS m/z 498.42 [M+H]$^+$, 520.40 [M+Na]$^+$.

These data are in accordance with the literature.[3]

†When crystals of **97** were not available this procedure was adopted. After the addition of water, the aqueous layer was decanted and the red gum was washed with water (20 ml x 3) by decantation, and then dissolved in methanol (20 ml). The methanol solution was poured into a crystallisation basin and the solvents were allowed to evaporate slowly overnight. This gave **97** as red plates.
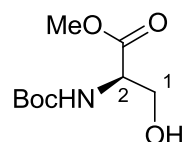
(2*R*)-2-*tert*-Butoxycarbonylamino-3-hydroxypropionic acid methyl ester 107

Boc$_2$O (14.0 g, 13.8 ml, 64.3 mmol, 1 eq.) was added to a solution of D-serine ester HCl (10.0 g, 64.3 mmol, 1 eq.) and triethylamine (26.8 ml, 193 mmol, 3 eq.) in DCM (80 ml). The mixture was stirred at r.t. overnight, then it was diluted with DCM (60 ml), and washed with NaHSO$_4$ (1 M, 50 ml x 2), sat. NaHCO$_3$ (100 ml) and brine (100 ml). The organic extracts were dried (MgSO$_4$), filtered and concentrated *in vacuo* to afford **107** as a colourless oil (13.0 g, 92 %).

$[\alpha_D]^{20}$ -12.0° (*c* = 1, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ_H 3.93 & 3.90 (d & d, *J* = 3.5 Hz, 1H, O=CCH, tautomers), 3.81 & 3.79 (d & d, *J* = 3.6 Hz, 2H, OCH$_2$, tautomers), 3.73 (s, 3H, OCH$_3$), 1.40 (s, 9H, 3 x CH$_3$); ^{13}C NMR (125 MHz, CDCl$_3$) δ_C 171.4 (C=O, ester), 155.7 (C=O, Boc), 80.0 (C, Boc), 62.9 (CH$_2$, C-1), 55.6 (CH, C-2), 52.4 (CH$_3$, MeO), 28.1 (3 x CH$_3$, Boc); ES-MS *m/z* 242.07 [M+Na]$^+$.

These data were in accordance with the literature.[4]
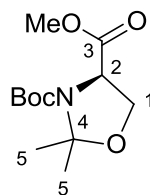
(3*R*)-3-*tert*-Butyl-4-methyl-2,2-dimethyloxazolidine-3,4-dicarboxylate 108

107 (2.00 g, 9.12 mmol) was dissolved in freshly distilled acetone (70 ml) and freshly distilled 2,2-dimethoxypropane (20 ml). The mixture was cooled to 0 °C and boron trifluoride etherate (40 µl) was added to the mixture. The mixture was stirred at 4 °C for 24 hours or until TLC indicated a complete conversion.

Then the solvents were removed *in vacuo*, the residues were dissolved in DCM (60 ml), washed with $NaHCO_3$ (50 ml x 2) and brine (50 ml), dried ($MgSO_4$), filtered and evaporated to give the crude product as a viscous oil. The oil was purified by silica gel column chromatography, eluting with hexane: ethyl acetate (8 : 2) to give **108** as a colourless viscous oil (2.28 g, 96 %).

R_f 0.36 (hexane : ethyl acetate 4:1); $[\alpha_D]^{20}$ +55.3° (*c* = 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 4.48-4.45 & 4.37-4.34 (m & m, 1H, O=CCH, rotamers), 4.16-4.08 & 4.04-3.98 (m & m, 2H, OCH_2, rotamers), 3.73 (s, 3H, OCH_3), 1.65 & 1.61 (s & s, 3H, CH_3 rotamers), 1.51 & 1.47 (s & s, 3H, CH_3 rotamers), 1.47 & 1.39 (s & s, 9H, 3 x CH_3 rotamers); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 171.6 & 171.2 (C=O, ester, rotamers), 152.0 & 151.1 (C=O, Boc, rotamers), 95.0 & 94.3 (C, C-4, rotamers), 80.8 & 80.2 (C, Boc, rotamers), 66.2 & 65.9 (CH, C-2, rotamers), 59.2 & 59.1 (CH_2, C-1, rotamers), 52.3 & 52.2 (CH_3, MeO, rotamers), 28.3 & 28.2 (3 x CH_3, Boc, rotamers) ; 25.9 & 25.1 (CH_3, C-5, rotamers), 24.9 & 24.3 (CH_3, C-5', rotamers) ES-MS *m/z* 282.12 $[M+Na]^+$. These data were in accordance with the literature.[4]
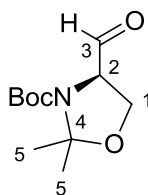
(4*R*)-4-Formyl-2,2-dimethyloxazolidine-3-carboxylic acid *tert*-butyl ester **87** (D-Garner's aldehyde)

Ester **108** (2.1 g, 8.1 mmol) was dissolved in dry toluene (40 ml) and the solution was cooled to -78 °C. Then DIBAL (25 % in toluene, 8.1 ml, 12.1 mmol) was added slowly via syringe to the stirred solution. The mixture was stirred at -78 °C for 2 h, then methanol (5 ml) was added slowly to quench the reaction while keeping the reaction temperature below -50 °C.

The mixture was allowed to warm to room temperature and potassium sodium tartrate 36 % solution (50 ml) was added and stirred vigorously for 30 mins until the cloudy mixture became clear.

The mixture was diluted with ether (50 ml) and the aqueous layer was separated and extracted with ether (50 ml). The combined organic extracts were dried ($MgSO_4$), filtered and evaporated.

The residues were purified by column chromatography, eluting with hexane: ethyl acetate (4:1), to give **87** as a colourless oil (1.2 g).

R_f 0.38 (hexane : ethyl acetate 4:1); $[\alpha_D]^{20}$ +79.1° (*c* = 1, $CHCl_3$); [^1]H NMR (400 MHz, $CDCl_3$) δ_H 9.56 & 9.50 (d & d, 1H, CHO, rotamers), 4.31-4.29 & 4.17-4.14 (m & m, 1H, O=CCH, rotamers), 4.06 & 4.04 (s & s, 2H, OCH_2, rotamers), 1.60 & 1.56

(s & s, 3H, CH_3 rotamers), 1.51 & 1.45 (s & s, 3H, CH_3 rotamers), 1.47 & 1.39 (s & s, 9H, 3 x CH_3 rotamers); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 199.3 & 199.2 (CHO, rotamers), 152.5 & 152.2 (C=O, Boc, rotamers), 94.9 & 94.2 (C, C-4, rotamers), 81.2 & 80.9 (C, Boc, rotamers), 64.7 & 64.6 (CH, C-2, rotamers), 63.8 & 63.3 (CH_2, C-1, rotamers), 28.1 (3 x CH_3, Boc); 26.6 & 25.7 (CH_3, C-5, rotamers), 24.6 & 23.7 (CH_3, C-5, rotamers); ES-MS m/z 230.13 $[M+H]^+$, 252.13 $[M+Na]^+$. These data were in accordance with the literature.[4]
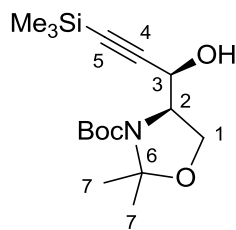
(4*R*)-4-((1*R*)-1-Hydroxy-3-trimethylsilanyl-prop-2-ynyl)-2,2-dimethyl-oxazolidine-3-carboxylic acid *tert*-butyl ester 88

Ethylmagnesium bromide (1M in THF, 30.2 ml, 30.2 mmol, 1.5 eq.) was added to a solution of ethynyltrimethylsilane (4.55 ml, 32.2 mmol, 1.6 eq.) in dry THF (50 ml) at 0 °C. The mixture was then heated under reflux for 2 h.

In a separate flask, copper(I) iodide (8.38 g, 44 mmol, 2.2 eq.) was dissolved in a mixture of dimethyl sulfide (25 ml) and dry THF (60 ml) at r.t., when the solid have all dissolved the solution was cooled to -78 °C. The Grignard reagent prepared above was slowly cannulated into the copper iodide solution at -78 °C, after the addition, the mixture was allowed to warm to -30 °C and stirred at that temperature for 30 mins.

The mixture was then cooled back to -78 °C, and a solution of D-Garner's aldehyde **87** (5.30 g, 20.1 mmol, 1 eq.) in dry THF (20 ml) was added dropwise. The mixture was stirred at -78 °C for a further 15 mins, then allowed to warm to r.t. and stirred overnight.

The mixture was diluted with ether (120 ml) and quenched with saturated NH_4Cl solution (200 ml). The aqueous layer was separated and extracted with ether (50 ml). The combined ether extracts were washed with sat. NH_4Cl (50 ml x2), 0.5 M HCl (50

ml), sat. NaHCO$_3$ (50 ml) and brine (50 ml), dried (MgSO$_4$), filtered and concentrated *in vacuo.*

A slurry was obtained which were applied to a silica gel chromatography column, and eluted with hexane: ethyl acetate (4 : 1), to give the addition product as a pale oil (4.61 g, 70 %).

R$_f$ 0.44 (hexane : ethyl acetate 4:1); $[\alpha_D]^{20}$ +42.0° (c = 0.5, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ_H 4.67 & 4.66 (s & s, 1H, C**H**OH, rotamers), 3.96-3.91 & 3.87-3.84 (m & m, 1H, NCH, rotamers), 3.66-3.62 & 3.60-3.56 (m & m, 2H, OCH$_2$, rotamers), 1.34 & 1.33 (s & s, 3H, CH$_3$ rotamers), 1.31 & 1.30 (s & s, 3H, CH$_3$ rotamers), 1.30 & 1.29 (s & s, 9H, 3 x CH$_3$ rotamers), 0.01 & 0.00 (s & s, 9H, 3 x SiCH$_3$ rotamers); ^{13}C NMR (125 MHz, CDCl$_3$) δ_C 153.5 (C=O, Boc), 100.7 (C, C-6), 99.7 ($\equiv$C, C-4), 91.3 ($\equiv$C, C-5), 79.4 (C, Boc), 64.3 (CH$_2$, C-1), 64.1 (CH, C-2), 47.5 (CH, C-3), 29.3 (2 x CH$_3$, C-7); 28.3 (3 x CH$_3$, Boc), -0.29 (3 x SiCH$_3$); ES-MS *m/z* 350.16 [M + Na]$^+$; HRMS (calculated for [M+Na]$^+$ C$_{15}$H$_{33}$NONaSi$_3$ = 350.1768) found 350.1767. These data are in accordance with the literature.[5]
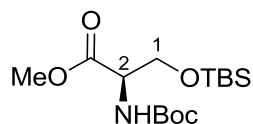
(2*R*)-2-((*tert*-Butoxycarbonyl)amino)-3-((*tert*-butyldimethylsilyl)-oxy)propionic acid methyl ester 113

A solution of *N*-Boc-D-serine methyl ester **107** (268 mg, 1.22 mmol, 1 eq.) in dry DCM (10 ml) and pyridine (108 µl, 1.34 mmol, 1.1 eq.), was cooled to 0 °C. *tert*-Butyldimethylsilyl trifluoromethanesulfonate (triflate) (355 mg, 309 µl, 1.34 mmol, 1.1 eq.) was added slowly *via* syringe and the mixture was stirred at r.t. overnight.

The mixture was diluted with DCM (40 ml), washed with 0.1 M HCl (50 ml), sat. NaHCO$_3$ (50 ml), brine (50 ml), dried (MgSO$_4$), filtered and concentrated to give a colourless oil. The crude product was purified by column chromatography, eluting with hexane: ethyl acetate (4:1) to give **113** as a colourless oil (312 mg, 77 %).

R$_f$ 0.69 (hexane: ethyl acetate 4:1); $[\alpha_D]^{20}$ -9.41° (c = 1, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ_H 5.36-5.31 (m, 1H, NCH), 4.05-3.73 (m, 2H, CH$_2$), 3.72 (s, 3H, OCH$_3$), 1.44 (s, 9H, 3 x CH$_3$), 0.85 (s, 9H, 3 x CH$_3$), 0.02 & 0.01 (s & s, 6H, 2 x SiCH$_3$); ^{13}C NMR (125 MHz, CDCl$_3$) δ_C 171.2 (C=O, ester), 155.4 (C=O, Boc), 79.8 (C, Boc), 63.7 (CH$_2$, C-1), 55.5 (CH, C-2), 52.2 (CH$_3$, MeO), 28.1 (3 x CH$_3$, Boc), 25.6 (3 x CH$_3$, *t*-BuSi), 18.1 (C, *t*-BuSi), -5.6 & -5.7 (2 x SiCH$_3$).

These data are in accordance with the literature.[6]
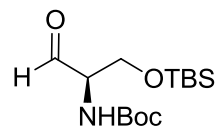
(2*R*)-*tert*-Butyl (1-((tert-butyldimethylsilyl)oxy)-3-oxopropan-2-yl)carbamate 112

A solution of ester **113** (240 mg, 0.72 mmol, 1 eq.) in dry toluene (10 ml) was cooled to -78 °C and DIBAL (25 % w/w in toluene, 0.72 ml, 1.08 mmol, 1.5 eq.) was added *via* syringe. Then the mixture was stirred at -78 °C for 2 h.

The reaction was quenched at -78 °C with the slow addition of methanol (2 ml), and then allowed to warm to r.t. Potassium sodium tartrate solution (36 %, 25 ml) was added and the mixture was stirred vigorously for 1 h until the cloudy solution became clear.

The mixture was extracted into ether (30 ml x 3), the organic layers were combined, dried (MgSO$_4$), filtered and evaporated *in vacuo* to give a colourless oil (219 mg, quant.) which was pure by TLC and used without further purification.

R$_f$ 0.51 (hexane: ethyl acetate 4:1); $[\alpha_D]^{20}$ -18.5° (c = 1, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ_H 9.65 (s, 1H, CHO), 5.38-5.37 (m, 1H, NCH), 4.26-3.84 (m, 2H, CH$_2$), 3.74 (s, 3H, OCH$_3$), 1.46 (s, 9H, 3 x CH$_3$), 0.86 (s, 9H, 3 x CH$_3$), 0.05 (s, 6H, 2 x SiCH$_3$).
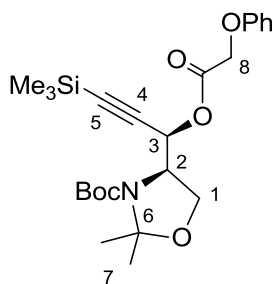
(*R*)-*tert*-Butyl 2,2-dimethyl-4-((*R*)-1-(2-phenoxyacetoxy)-3-(trimethylsilyl) prop-2-yn-1-yl)oxazolidine-3-carboxylate 116

A solution of alcohol **88** (250 mg, 0.76 mmol, 1 eq.) and triethylamine (0.16 ml, 1.15 mmol, 1.5 eq.) in DCM (10 ml) was cooled to 0 °C. Phenoxyacetyl chloride (156 mg, 130 µl, 0.917 mmol, 1.2 eq.) was added to the mixture and stirred overnight.

The mixture was diluted with DCM (50 ml) and quenched with sat. $NaHCO_3$ (50 ml). The organic layer was separated and washed with 0.5 M HCl (25 ml), sat. $NaHCO_3$ (50 ml) and brine (50 ml), dried ($MgSO_4$), filtered and the solvents were removed *in vacuo*.

The crude residues were applied to a chromatography column, eluting with hexane: ethyl acetate (9: 1), and collecting the fraction with R_f = 0.32 to give **116** as a colourless viscous oil (307 mg, 87 %). The oil crystallised slowly into white needles upon standing.

m.p. 69-70 °C; R_f 0.64 (hexane : ethyl acetate 4 : 1); $[\alpha_D]^{20}$ +21.9° (*c* = 0.1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.14-7.09 (m, 2H, ArH), 6.84-6.81 (m, 1H, ArH), 6.74-6.71 (m, 2H, ArH), 5.82-5.71 (m, 1H, C**H**OH), 4.48 (s, 2H, O=CC**H**$_2$), 4.00-3.90 (m, 1H, NC**H**), 3.78-3.75 (m, 2H, OC**H**$_2$), 1.46 (s, 3H, C**H**$_3$), 1.42 (s, 3H, C**H**$_3$),

1.32 & 1.31 (s, 9H, 3 x CH_3), 0.00 (s, 9H, 3 x $SiCH_3$); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 167.3 (C=O, ester), 157.6 (C, Ph), 152.6 (C=O, Boc), 129.5 (2 x CH, Ph), 121.8 (CH, Ph), 114.5 (2 x CH, Ph), 99.2 & 98.6 (C, C-6, rotamers), 95.0 & 94.5 ($\equiv$C, C-4 rotamers), 93.3 & 92.5 ($\equiv$C, C-5, rotamers), 80.8 (C, Boc, rotamers), 65.2 & 64.8 (CH, C-2, rotamers), 65.1 (CH_2, C-8), 59.3 & 58.3 (CH, C-3 rotamers), 64.5 (CH_2, C-1), 28.3 (3 x CH_3, Boc), 26.9 & 25.8 (CH_3, C-7, rotamers), 24.7 & 23.6 (CH_3, C-7', rotamers), -0.45 (3 x $SiCH_3$). HRMS (calculated for $[M+Na]^+$ $C_{24}H_{35}NO_4NaSi$ = 484.2131) found 484.2115.

†A single crystal X-ray structure is available for this compound (Appendix 1)
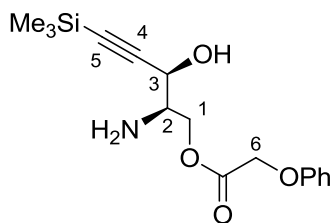
(2R)-2-Amino-(3R)-3-hydroxy-5-(trimethylsilyl)pent-4-yn-1-yl 2-phenoxyacetate 119

90 % Trifluoroacetic acid (2 ml) was added to ester **116** (257 mg, 0.56 mmol, 1 eq.) and the reaction was stirred for 10 mins at r.t.. The mixture was diluted with water (20 ml) and sat. $NaHCO_3$ was added slowly until the solution tested neutral to blue litmus paper and then extracted with ethyl acetate (100 ml). The organic layer was separated, washed with brine (50 ml), dried (Na_2SO_4), filtered and the solvents were removed to give the compound as a colourless oil (162 mg, 91 %) which was used without further purification.

R_f 0.48 (ethyl acetate : hexane 4: 1); $[\alpha_D]^{20}$ +4.50° (c = 0.5, EtOH); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.19-7.08 (m, 2H, ArH), 6.90-6.85 (m, 1H, ArH), 6.80-6.77 (m, 2H, ArH), 4.50 (d, J = 4.9 Hz, 1H, CHOH), 4.37 (s, 2H, O=CCH$_2$), 4.07-4.00 (m, 1H, NCH), 3.82-3.68 (m, 2H, OCH$_2$), 0.00 (s, 9H, 3 x SiCH$_3$); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 169.4 (C=O, ester), 157.1 (C, Ph), 129.8 (2 x CH, Ph), 122.3 (CH, Ph), 114.8 (2 x CH, Ph), 103.1 ($\equiv$C, C-4), 91.9 ($\equiv$C, C-5), 67.3 (CH$_2$, C-6), 63.1 (CH, C-3), 62.0 (CH$_2$, C-1) 54.5 (CH, C-2), -0.3 (3 x SiCH$_3$). HRMS (calculated for [M+Na]$^+$ C$_{16}$H$_{23}$NO$_4$NaSi = 344.1294) found 344.1279.
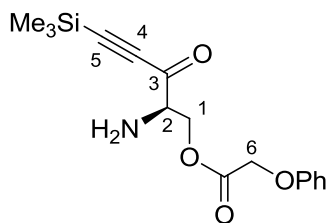
(2*R*)-2-Amino-3-oxo-5-(trimethylsilyl)pent-4-yn-1-yl 2-phenoxyacetate **120**

A solution of amino alcohol **119** (72 mg, 0.22 mmol, 1 eq.) in acetone (5 ml) was cooled to 0 °C and then Jones reagent (1 M, 0.56 ml, 2.5 eq.) was added slowly. The mixture was stirred at r.t. for 3 h.

After which the reaction was quenched by the addition of isopropanol (1 ml) and stirred for 10 mins. The mixture was evaporated until near-dryness, and ether (50 ml) was added and the contents were swirled for 5 mins. The insoluble chromium salts were removed by decanting and the organic layer was washed with water (25 ml) and brine (25 ml), dried (Na_2SO_4), filtered and concentrated *in vacuo* to give a yellow oil. The oil was applied to a silica gel column and eluted with hexane:ethyl acetate (4: 1). The compound was obtained as a colourless oil (60 mg, 84 %).

R_f 0.29 (hexane : ethyl acetate 4 : 1); $[\alpha_D]^{20}$ -14.3° (*c* = 0.03, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.24-7.20 (m, 2H, ArH), 6.95-6.91 (m, 1H, ArH), 6.87-6.84 (m, 2H, ArH), 4.73-4.30 (m, 1H, NCH), 4.47-4.45 (m, 2H, O=CCH$_2$), 4.12-3.99 (m, 1H, OCH$_2$), 0.15 (s, 9H, 3 x SiCH$_3$); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 182.6 (C=O, C-3), 168.9 (C=O, ester), 157.1 (C, Ph), 129.8 (2 x CH, Ph), 122.3 (CH, Ph), 114.8 (2 x CH, Ph), 103.8 (≡C, C-4), 99.5 (≡C, C-5), 67.3 (C, C-6), 62.7 (CH$_2$, C-1), 62.1 (CH, C-2), -1.0 (3 x SiCH$_3$).

$$Ph-Si \begin{array}{c} Ph \\ \\ Ph \end{array} \diagdown$$
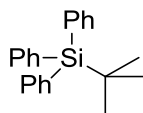
tert-Butyltriphenylsilane 124

tert-Butyldiphenylsilyl chloride (10 ml, 38.5 mmol, 1 eq.) was added dropwise to a solution of phenyllithium (2.0 M in di-*tert*-butyl ether, 23.1 ml, 46.2 mmol, 1.2 eq.) at 0 °C over 30 mins. The mixture was then stirred at room temperature for 16 h.

The reaction was quenched with the addition of ether (15 ml) and water (25 ml). A white precipitate was formed which was collected by filtration under suction, the solid was washed with water (50 ml) and cold ether (20 ml) and dried over $CaCl_2$ *in vacuo* to give **124** as a white powder(Yield 6.3 g, 52 %).

m.p. 158 °C; ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.67-7.63 (m, 6H, ArH), 7.47-7.39 (m, 9H, ArH), 1.25 (s, 9H, 3 x CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 136.5 (6 x CH), 134.8 (3 x C), 129.1 (3 x CH), 127.6 (6 x CH), 28.8 (3 x CH_3), 18.7 (C); ES-MS *m/z* 339.38 $[M+Na]^+$. These data were in accordance with the literature.[7]
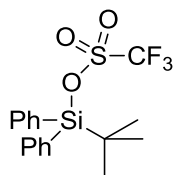
tert-Butyltriphenylsilyl trifluoromethanesulfonate 122

A solution of *tert*-butyldiphenylsilyl trifluoromethanesulfonate (TBDPSOTf) was prepared immediately before use as follows.[7]

A solution of *tert*-butyltriphenylsilane **124** (3.8 g, 12.0 mmol, 1.04 eq.) in dry DCM (25 ml) and cooled to 0 °C, and trifluoromethanesulfonic acid (triflic acid) (1.02 ml, 11.5 mmol, 1 eq.) was added slowly. The mixture was stirred at r.t. for 20 mins and then used immediately without further purification.
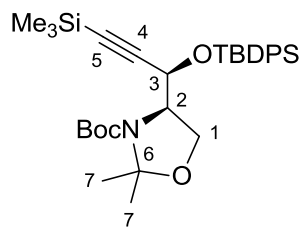
(*R*)-*tert*-Butyl 4-((*R*)-1-((*tert*-butyldiphenylsilyl)oxy)-3-(trimethylsilyl)prop-2-yn-1-yl)-2,2-dimethyloxazolidine-3-carboxylate **121**

A solution of alcohol **88** (2.43 g, 7.4 mmol, 1 eq.) and pyridine (2 ml) in DCM (20 ml) and cooled to 0 °C. A solution of TBDPSOTf **122** in DCM prepared as described above (25 ml, 11.5 mmol, 1.5 eq.) was added dropwise and the mixture was stirred at r.t. overnight.

The reaction was quenched with methanol (5 ml), and diluted with DCM (50 ml), washed with 0.5 M HCl (50 ml), sat. NaHCO$_3$ (50 ml) and brine (50 ml), dried (MgSO$_4$) and evaporated to give a yellow residue.

The residues were purified by column chromatography, eluting with hexane: ethyl acetate (9:1) to afford **121** as a colourless oil (3.27 g, 75 %)

R$_f$ 0.44 (hexane : ethyl acetate 4:1); $[\alpha_D]^{20}$ +34.0° (*c* = 1, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ_H 7.75-7.66 (m, 4H, ArH), 7.46-7.35 (m, 6H, ArH), 5.01 &4.85 (d & d, *J* = 4.9 Hz, 1H, OCH, rotamers), 4.43-4.38 & 4.36-4.33 (m & m, 1H, NCH, rotamers), 4.19-4.09 & 3.98-3.93 (m & m, 2H, OCH$_2$, rotamers), 1.64 &1.63 (s & s, 3H, CH$_3$, rotamers), 1.57 & 1.53 (s & s, 3H, CH$_3$ rotamers), 1.49 & 1.45 (s & s, 3H, CH$_3$, rotamers), 1.18 (s, 6H, 2 x CH$_3$), 1.08 (s, 9H, 3 x CH$_3$), 0.08 & 0.04 (s & s, 9H, 3 x SiCH$_3$, rotamers); ^{13}C NMR (125 MHz, CDCl$_3$) δ_C 151.6 (C=O, Boc), 136.0 (2 x

CH, Ph), 135.8 & 135.7 (2 x CH, Ph), 133.2 (C, Ph), 133.1 (C, Ph), 129.9 (CH, Ph), 129.7 & 129.5 (CH, Ph), 127.7 & 127.5 (2 x CH, Ph), 127.4 & 127.3 (2 x CH, Ph), 103.6 ($\equiv$C, C-4), 95.1 ($\equiv$C, C-5), 92.0 (C, C-6), 79.9 (C, Boc), 65.1 & 64.7 (CH$_2$, C-1), 64.9 & 64.2 (CH, C-2), 61.2 & 60.2 (CH, C-3), 28.4, 28.1, 26.9 (3 x CH$_3$, Boc), 26.8 (3 x CH$_3$, t-BuSi), 25.3 & 24.4 (2 x CH$_3$, C-7), 19.3 (C, t-BuSi), -0.48 (3 x SiCH$_3$); ES-MS m/z 588.22 [M+Na]$^+$; HRMS (calculated for C$_{32}$H$_{47}$NO$_4$NaSi$_2$ = 588.2941) found 588.2954.
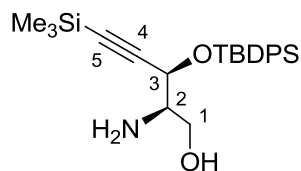
(2*R*, 3*R*)-2-Amino-3-((*tert*-butyldiphenylsilyl)oxy)-5-(trimethylsilyl)pent-4-yn-1-ol 125

90 % Trifluoroacetic acid (5 ml) was added to **121**(728 mg, 1.28 mmol) and the reaction was stirred at r.t. for 30 mins. The mixture was then neutralised with sat. $NaHCO_3$ until neutral to blue litmus paper, then the solution was extracted with ether (50 ml x 2). The ether layer was separated, dried (Na_2SO_4), filtered and concentrated to give **125** as a colourless oil (544 mg, 99 %).

R_f 0.29 (ethyl acetate : hexane 4:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.75-7.68 (m, 4H, ArH), 7.45-7.35 (m, 6H, ArH), 5.01 &4.85 (d & d, *J* = 4.9 Hz, 1H, OCH, rotamers), 4.43-4.38 & 4.36-4.33 (m & m, 1H, NCH, rotamers), 4.19-4.09 & 3.98-3.93 (m & m, 2H, OCH_2, rotamers), 1.64 &1.63 (s & s, 3H, CH_3,rotamers), 1.57 & 1.53 (s & s, 3H, CH_3, rotamers), 1.49 & 1.45 (s & s, 3H, CH_3, rotamers), 1.18 (s, 6H, 2 x CH_3), 1.08 (s, 9H, 3 x CH_3), 0.08 & 0.04 (s & s, 9H, 3 x $SiCH_3$, rotamers); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 136.1 (2 x CH, Ph), 135.9 (2 x CH, Ph), 133.2 (C, Ph), 132.9 (C, Ph), 130.0 (CH, Ph), 129.7 (CH, Ph), 127.7 (2 x CH, Ph), 127.4 (2 x CH, Ph), 103.9 ($\equiv$C, C-4), 92.1 ($\equiv$C, C-5), 63.3 (CH_2, C-1), 64.9 & 64.2 (CH, C-2, rotamers), 61.2 & 60.2 (CH, C-3, rotamers), 30.3, 28.3, 26.6 (3 x CH_3, *t*-BuSi, rotamers), 26.9 (3 x CH_3, *t*-BuSi, rotamers), 19.4 (C, *t*-BuSi), -0.44 (3 x $SiCH_3$); ES-MS *m/z* 426.18 $[M+H]^+$, 448.16 $[M+Na]^+$; HRMS (calculated for $[M+H]^+$ $C_{24}H_{36}NO_2Si_2$ = 426.2285) found 426.2275.
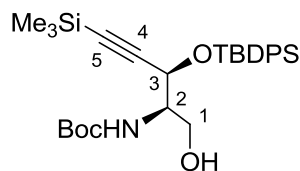
***tert*-Butyl ((2*R*,3*R*)-3-((*tert*-butyldiphenylsilyl)oxy)-1-hydroxy-5-(trimethylsilyl)pent-4-yn-2-yl)carbamate 126**

A solution of amino alcohol **125** (371 mg, 0.87 mmol), Boc$_2$O (225 µl, 1.04 mmol, 1.2 eq.) and triethylamine (225 µl) in DCM (40 ml) was stirred at r.t. for 16 h.

The mixture was diluted with DCM (20 ml), and washed with NaHSO$_4$ (1M, 50 ml x 2), sat. NaHCO$_3$ (50 ml) and brine (50 ml). The organic extracts were dried (MgSO$_4$), filtered and concentrated *in vacuo*, and purified by silica gel column chromatography, eluting with hexane: ethyl acetate (4 : 1) to give **126** as a colourless oil (429 mg, 94 %).

R$_f$ 0.36 (hexane: ethyl acetate 4:1); $[\alpha_D]^{20}$ -19.0° (*c* = 0.1, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$) δ_H 7.77-7.70 (m, 4H, ArH), 7.48-7.36 (m, 6H, ArH), 5.01 (m, 1H, OCH), 4.52 (m, 1H, NCH), 3.91-3.89 (m, 2H, OCH$_2$), 3.79 (br s, 1H, OH), 1.45 (s, 9H, 3 x CH$_3$), 1.10 (s, 9H, 3 x CH$_3$), 0.05 (s, 9H, 3 x SiCH$_3$); ^{13}C NMR (125 MHz, CDCl$_3$) δ_C 155.8 (C=O, Boc), 136.0 (2 x CH, Ph), 135.8 (2 x CH, Ph), 133.0 (C, Ph), 132.5 (C, Ph), 130.0 (CH, Ph), 129.7 (CH, Ph), 127.7 (2 x CH, Ph), 127.4 (2 x CH, Ph), 103.7(≡C, C-4), 92.1(≡C, C-5), 79.6 (C, Boc), 64.2 (CH, C-2), 62.5 (CH$_2$, C-1), 56.1 (CH, C-3), 28.3 (3 x CH$_3$, Boc), 26.9 (3 x CH$_3$, *t*-BuSi), 19.3 (Si-C), -0.56 (3 x SiCH$_3$); ES-MS *m/z* 548.04 [M+Na]$^+$; HRMS (calculated for [M+Na]$^+$ C$_{29}$H$_{43}$NO$_4$NaSi$_2$ = 548.2628) found 548.2621.
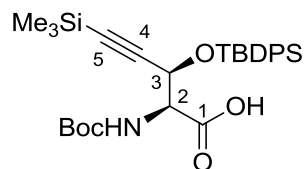
(2S,3R)-2-((*tert*-Butoxycarbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)-5-(trimethylsilyl)pent-4-ynoic acid 127

A solution of alcohol **126** (259 mg, 0.493 mmol, 1 eq.) in acetone (10 ml) was cooled to 0 °C, then Jones reagent† (1M, 1.23 ml, 2.5 eq.) was added dropwise to the stirred solution. The reaction was stirred at r.t. for 4 h.

The reaction was quenched by the addition of isopropanol (2 ml) and stirred until the orange colour disappeared (about 10 mins). The mixture was evaporated in vacuo until only a small amount of solvent remained. Then ether (100 ml) was added to the residue and the flask was swirled for 5 mins. The insoluble chromium salt was decanted from the ether solution, and the ether layer was washed with water (50 ml) and brine (50 ml), dried ($MgSO_4$) and filtered.

The solvents were removed to give **127** as a colourless sticky oil, which solidified upon standing (248 mg, 93 %). The solid may be further purified by recrystallisation in ether /hexane to give colourless rectangular plates.

m.p. 154-155 °C; $[\alpha_D]^{20}$ -25.7° (c = 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 7.73-7.72 (m, 4H, ArH), 7.47-7.36 (m, 6H, ArH), 5.42 (d, J = 9.4 Hz, 1H, SiOCH), 4.85 (d, J = 2.5 Hz, 1H, COOH), 4.57 (dd, J = 9.4 Hz, 2.4 Hz, 1H, O=CCH), 1.50 (s, 9H, 3 x CH_3), 1.06 (s, 9H, 3 x CH_3), 0.01 (s, 9H, 3 x $SiCH_3$); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 173.8 (COOH), 155.6 (C=O, Boc), 136.1 (2 x CH, Ph), 136.0 (2 x CH, Ph), 132.7

(C, Ph), 131.9 (C, Ph), 130.1 (CH, Ph), 129.8 (CH, Ph), 127.7 (2 x CH, Ph), 127.4

(2 x CH, Ph), 101.7($\equiv$C, C-4), 92.5($\equiv$C, C-5), 80.2 (C, Boc), 65.1 (CH, C-2), 58.5

(CH, C-3), 28.3 (3 x CH_3, Boc), 26.8 (3 x CH_3, t-BuSi), 19.3 (Si-C), -0.60 (3 x

$SiCH_3$); ES-MS m/z 561.97 $[M+Na]^+$, 533.95 $[M-H]^-$; HRMS (calculated for

$[M+Na]^+$ $C_{29}H_{41}NO_5NaSi_2$ = 562.2421) found 562.2427.

†Jones reagent (1M) was prepared by dissolving chromium trioxide (3.35 g) and conc. H_2SO_4 (3

ml) in distilled water and made up to 25 ml.

‡A single crystal X-ray structure of this compound is available (Appendix 2).
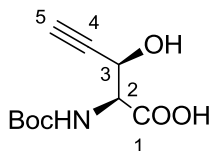
(2*S*,3*R*)-2-[(*tert*-butoxycarbonyl)amino]-3-hydroxypent-4-ynoic acid;
N-Boc-β-ethynyl-L-serine **128**

A solution of 10 % KOH in methanol (2 ml) was added to carboxylic acid **127**(170 mg, 0.315 mmol), the reaction was stirred at r.t. for 6 h. The solution was then diluted with methanol (40 ml) and neutralised with pre-washed Dowex 50Wx4-400 H⁺ ion-exchange resin until the solution was neutral to damp pH paper. The mixture was filtered and the resins were washed with methanol (10 ml). The combined filtrates were evaporated to give a product containing **128**.

The product was carried forward to the next step without further purification.

An analytical sample was prepared as follows. The product mixture was dissolved in water (20 ml) and the pH was adjusted to 9-10 with a few drops of NaOH(aq). The aqueous solution was washed with ether (50 ml x 2). The aqueous layer was collected and the pH was readjusted to 5 with HCl (aq). The solution was lyophilised and the residues were extracted with ethanol (10 ml). The mixture was filtered and the filtrate evaporated to give a sample of **128** as a white foam.

$[\alpha_D]^{20}$ +55.6° (c = 0.05, EtOH); ^{1}H NMR (400 MHz, D_2O) δ_H 4.94 (br s, 1H, OCH), 4.49 (d, J = 3.5 Hz, 1H, NCH), 2.93 (br s, 1H, ≡CH), 1.47 (s, 1H, 3 x CH_3); ^{13}C NMR (125 MHz, D_2O) δ_C 173.1 (COOH), 158.1 (C=O, Boc), 81.9 (C-O, Boc), 80.6 (≡C, C-4), 75.6 (HC≡, C-5), 61.7 (C-OH, C-3), 58.5 (C-N, C-2), 27.5 (CH_3, Boc); ES-MS 228.08 [M-H]⁻.
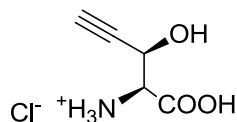
(2S,3R)-2-amino-3-hydroxypent-4-ynoic acid hydrochloride;
β-Ethynyl-L-serine hydrochloride 4

Conc. HCl (2 ml) and water (2 ml) were added to the product mixture from the previous reaction containing **128** and the mixture was stirred for 2 h at r.t., and then diluted with distilled water (20 ml) and washed with ether (20 ml x 3).

The aqueous layer was collected and lyophilised to give β-ethynyl-L-serine hydrochloride as a pale yellow solid (49 mg, 93 % from two steps).

m.p. 208 °C (decompose); $[\alpha_D]^{20}$ -69.6° (c = 0.05, EtOH) [lit.[8] -71.5°, c = 1, H_2O]; ^{1}H NMR (400 MHz, D_2O) δ_H 4.99 (dd, J = 2.2 & 3.0 Hz, 1H, OCH), 4.10 (d, J = 3.0 Hz, 1H, NCH), 2.99 (d, J = 2.2 Hz, 1H, ≡CH); ^{13}C NMR (125 MHz, D_2O) δ_C 168.7 (C=O), 78.6 (HC≡), 77.7 (C≡), 59.4 (C-OH), 57.5 (C-N); HRMS (calculated for $[M+H]^+$ $C_5H_8NO_3$ = 130.0499) found 130.0499.
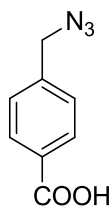
4-(Azidomethyl)benzoic acid;

α-Azido-*p*-toluic acid 129

4-(Bromomethyl)benzoic acid (α-bromo-*p*-toluic acid) (1.00 g, 4.65 mmol, 1 eq.) and sodium azide (317 mg, 4.88 mmol, 1.05 eq.) were dissolved in methanol (30 ml), the solution was heated under reflux for 2 h.

The solvents were removed *in vacuo* and the white residues were partitioned between 0.2 M HCl (50 ml) and ether (50 ml). The ether layer was separated and washed with water (20 ml) and brine (20 ml), dried ($MgSO_4$) and filtered.

Evaporation of the solvents gave **129** as a white powder (794 mg, 96 %).

m.p. 129-130 °C [lit.[9] 129-131 °C]; ^{1}H NMR (DMSO-d_6, 400 MHz) δ_H 7.96 (d, J = 8.5 Hz, 2H, ArH), 7.48 (d, J = 8.5 Hz, 2H, ArH), 4.55 (s, 2H, CH_2); ^{13}C NMR (DMSO-d_6, 125 MHz) δ_C 167.5 (COOH), 141.1 (C, Ar), 130.9 (C, Ar), 130.2 (CH, Ar), 128.8 (CH, Ar), 55.3 (CH_2); HRMS (calculated for [M-H]$^-$ $C_8H_6N_3O_2$ = 176.0466) found 176.0459.
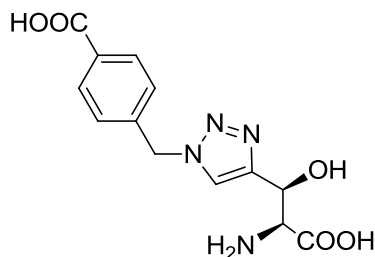
4-({4-[(1S,2S)-2-amino-2-carboxy-1-hydroxyethyl]-1H-1,2,3-triazol-1-yl}methyl)benzoic acid 132

To a solution of β-ethynyl-L-serine hydrochloride (10 mg, 0.06 mmol) in 50 % aqueous DMSO (1 ml) was added a solution of α-azido-p-toluic acid (30 mg, 0.18 mmol, 3 eq.) in DMSO (0.2 ml), a solution of sodium ascorbate (12 mg, 0.06 mmol, 1 eq.) in water (0.2 ml) and a solution of Cu-TBTA complex (10 mM, 0.6 ml, 0.006 mmol, 0.1 eq.) in 50 % DMSO. The mixture was stirred at r.t. for 12 h.

The mixture was purified through a short C-18 RP silica gel column, eluting with methanol : water (1 : 1). Fractions containing 132 were combined and concentrated *in vacuo*. The residues were purified by semi-preparative HPLC, using this HPLC condition: Mobile phase A: 0.05 % TFA/water; B: 0.05 % TFA/CH$_3$CN; Gradient: initial 0 % B to 15 % B in 15 min, 80 % B in 20 min, returning to the initial conditions at 25 min. Under these conditions, the compound eluted between 17.0 to 17.5 mins. Evaporation of the solvents from this fraction, followed by freeze-drying afforded an analytical sample of **132** as a white solid (4 mg, 22 %).

^{1}H NMR (DMSO-d_6, 400 MHz) δ_H 8.20 (s, 1H, triazole CH), 7.95 (d, J = 8.2 Hz, 2H, ArH), 7.40 (d, J = 8.2 Hz, 2H, ArH), 5.72 (s, 2H, CH$_2$), 5.27 (d, J = 4.0 Hz, 1H, OCH), 4.24 (br, s, 1H, NCH); ES-MS 307.12 [M+H]$^+$, 329.11 [M+Na]$^+$, 345.08 [M+K]$^+$; HRMS (calculated for [M+H]$^+$ C$_{13}$H$_{15}$N$_4$O$_5$ = 307.1037) found 307.1033.

5.2 Biological Methods

5.2.1 General Methods

Reagents were obtained from commercial sources (Acros Organic, Sigma, Fisher) and used without further purification. Enzymes and enzymatic reaction buffers were obtained from commercial sources (Novagen, Promega) and diluted with sterile deionised water as required. pEHISTEV vector was kindly provided by Dr H. Liu (University of St Andrews). DNA primers were ordered from Eurogentec.

Streptomyces cattleya DSM 46488 was obtained from DSMZ (Germany) as dried culture. Competent cells: *Escherichia coli* BL21 Gold cells were kindly provided by Dr L. Major (University of St Andrews) and *E. coli* DH10B were kindly provided by Dr C. Zhao (University of St Andrews).

All microbiological works were carried out using standard sterile techniques under a Gallenkamp laminar flowhood. Experiments involving cell-free extracts or protein solutions were carried out at 4 °C or on ice unless otherwise stated. Glasswares, equipments and consumables for biological works were sterilised by autoclaving, flaming or spraying with 70 % ethanol as appropriate prior to use. Sterilised consumables were used as supplied. Media were sterilised by autoclaving.

Cell cultures were incubated in a temperature controlled Gallenkamp orbit incubator or an Innova 2000 platform shaker. Centrifugation were carried out on a Beckman Avanti centrifuge. Micro-centrifugation were carried out on a Hettich Mikro 200 bench-top centrifuge. Cell lysis were performed either by sonication with a Vibra Cell apparatus from Sonics & Materials Inc., or by the use of a Constant Systems cell disruptor running at 30 kPsi. Protein purifications were carried out manually or by the

use of an Akta-Express protein purification systems connected to pre-packed columns from GE Healthcare (Ni-Sepharose 6 FF, Desalt 16/10, Superdex 75, Superdex 200).

PCR were carried out on a Techne TC-512 machine. Electroporation of competent cells were carried out on a BioRad Gene Pluser X-cell apparatus. The polymerase reactions were carried out using the recommended procedures from the suppliers. DMSO (5 % v/v final concentration) was supplemented to all PCR reactions mixtures.

Plasmids from *E. coli* were prepared using the QIAprep Spin Miniprep kit (Qiagen) from an overnight culture grown in LB (4.5 ml) with the appropriate antibiotic.

Optical density was measured at 600 nm using a Jenway 6300 Spectrophotometer. Protein and DNA concentrations were determined by a Nanodrop ND1000 Spectrophotometer.

^{19}F NMR spectra were recorded with proton-decoupling in 10 % D_2O in H_2O at 298 K on a Bruker Avance 400 MHz spectrometer and calibrated to an external reference ($CFCl_3$).

DNA Sequencing was carried out by the Sequencing Service, Univeristy of Dundee.

5.2.2 Electrophoresis Methods

Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) were performed on an Invitrogen XCell SureLock^{TM} mini-cell apparatus connected to an Amersham Pharmacia biotech EPS 301 power supply operating at a constant current of 125 mA for 35 mins. NuPAGE^{TM} Bis-Tris 12-well pre-case gel plates containing 4-12 % acrylamide were used. The gels were submerged in 1x NuPAGE^{TM} buffer during electrophoresis.

Protein samples for SDS-PAGE were prepared by adding 5 μl of SDS-PAGE loading buffer (4x) to 20 μl of protein sample, vortexed and heated at 95 °C for 10 mins to denature the protein. 10 μl samples were loaded onto the wells of the gel. Prestained PagerulerTM protein ladder was used as a guide for size determination.

The gels were stained with by soaking in Instant Blue^{TM} reagent for 60 minutes with gentle agitation. Destaining was not necessary for this reagent. The gels were then washed with water and the images were recorded with a digital colour scanner (Canon).

DNA agarose gel electrophoresis were performed on a HOEFER^{TM} HE33 mini horizontal submarine unit (Amersham Biosciences) connected to a power-pac 300 (BioRad) operating at a constant current of 70 V for 110 mins. DNA gels were cast from 0.8 % agarose solution in 1x TAE buffer at 60 °C and allowed to set for 30 mins at r.t. prior to use. The gels were submerged in 1x TAE buffer during electrophoresis.

DNA samples were prepared by adding 5 μl of DNA loading buffer (4x) to 20 μl of DNA sample. 10 μl samples were loaded onto the wells of the gel. Generuler^{TM} 1kb ladder was used as a guide for size determination.

The gels were stained by soaking in a 0.01 % ethidium bromide aqueous solution for 20 mins and visualised by a UV light transilluminator. UV photographs of the gels were taken as records.

5.2.3 Media and buffer

5.2.3.1 Culture media

Luria broth media (LB) for *E. coli*

Tryptone (10 g), yeast extract (5 g) and NaCl (10 g) were dissolved in distilled water (1 L) and the pH was adjusted to 7.25 with NaOH. The solution was sterilised by autoclaving before use.

Tryptic soy broth-yeast extract media (TSBY media) for *Streptomyces* sp.

Tryptic soy broth (30 g), yeast extract (5 g) and sucrose (103 g) were dissolved in distilled water (1000 ml) and the pH was adjusted to 7.20 with NaOH solution. The solution was sterilised by autoclaving before use.

Auto-induction media for *E. coli*

ZY media

Tryptone (10 g) and yeast extract (5 g) were dissolved in distilled water (925 ml). The solution was sterilised by autoclaving before use.

NPS (20 X)

$(NH_4)_2SO_4$ (66 g, 0.5 mol), KH_2PO_4 (136 g, 1 mol) and Na_2HPO_4 (142 g, 1 mol) were dissolved in distilled water (1000 ml, final volume). The solution was sterilised by autoclaving before use.

5052 solution (50 X)

Glycerol (250 g) was dissolved in distilled water (730 ml), followed by glucose (25 g) and α-lactose (100 g). The mixture was heated gently if necessary. The solution was sterilised by autoclaving before use.

5.2.3.2 Buffers for gel electrophoresis

TAE buffer (50 X stock solution)

Tris-base (242 g), glacial acetic acid (52 ml) and EDTA disodium salt (14.6 g) were dissolved in distilled water and made up to a final volume of 1000 ml.

TAE buffer (1 X)

20 ml of TAE (50 X) solution was made up to 1000 ml with distilled water.

SDS-PAGE loading buffer (4 X)

Dithiothreitol (DTT) (0.77 g), sodium dodecylsulfate (0.4 g) and bromophenol blue (0.02 g) were dissolved in 1 M Tris HCl pH 7 (1 ml) and distilled water (2 ml), gentle heating of the mixture may be required. When the solids are dissolved, glycerol (2 ml) is added. Store at 4 °C.

DNA agarose gel loading buffer (6 X)

Sucrose (30 g) and bromophenol blue (0.2 g) were dissolved in distilled water (final volume 100 ml).

5.2.3.3 Buffers for protein purification

Lysis buffer

Lysis buffer contains 20 mM Tris HCl pH 7.5, 20 mM imidazole HCl pH 7.5, NaCl 500 mM and β-mercaptoethanol 3 mM.

Wash buffer

Wash buffer contains 20 mM Tris HCl pH 7.5, 40 mM imidazole HCl pH 7.5, NaCl 500 mM and β-mercaptoethanol 3 mM.

Elution buffer

Elution buffer contains 20 mM Tris HCl pH 7.5, 400 mM imidazole HCl pH 7.5, NaCl 500 mM and β-mercaptoethanol 3 mM.

Desalt buffer

Desalt buffer contains 20 mM Tris HCl pH 7.5, NaCl 150 mM and β-mercaptoethanol 3 mM.

Gel filtration buffer

Gel filtration buffer contains 20 mM Tris HCl pH 7.5 and DTT 5 mM.

5-FDRPi for crystallisation is prepared in this buffer solution.

5.2.4 Cultures of *Streptomyces cattleya*

5.2.4.1 Growth of *S. cattleya* on solid media

Streptomyces cattleya NRRL 8057 or DSM 46488 can be grown on solid agar plate composed of soybean flour (2 % w/v), mannitol (2 % w/v), agar (1.5 % w/v) and tap water. The plates were maintained at 30 °C where the bacteria mature and develop purple spores after a period of 28 days. The purple spores can be collected by means of sterilised cotton swabs and stored as a suspension in 40 % glycerol at -80 °C.[10]

5.2.4.2 Fermentation culture

The fermentation of *S. cattleya* was carried out by inoculating a sterilised, defined media (90 ml) containing deionised water (450 ml), trace elements solution (150 ml), filtered carbon source solution (75 ml), phosphate buffer (75 ml, 150 mM, pH 7.0) and potassium fluoride (3 ml, 0.5 M) with the purple spores glycerol stock of *S. cattleya* and allowed to grow at 28 °C for 6 to 8 days in a 500 ml conical flask with shaking at 180 rpm.[10]

Trace elements solution contains deionised water (900 ml), NH_4Cl (6.75 g), NaCl (2.25 g), $MgSO_4 \cdot 7H_2O$ (2.25 g), $CaCO_3$ (1.13 g), $FeSO_4 \cdot 7H_2O$ (113 mg), $CoCl_2 \cdot 6H_2O$ (45 mg) and $ZnSO_4 \cdot 7H_2O$ (45 mg). The solution was sterilised by autoclaving before use.

Carbon source solution contains deionised water (900 ml), glycerol (45 g), monosodium L-glutamate (22.5 g), *myo*-inositol (1.8 g), *p*-aminobenzoic acid (0.45 mg). The solution was sterilised by autoclaving before use.

5.2.4.3 Isolation of genomic DNA from *S. cattleya*

Glycerol stock of purple *S. cattleya* spores (20 μl) was inoculated to sterilised TSBY media (200 ml) and the media was shaken at 30 °C for 36 h. The mycelia were harvested by centrifugation and the supernatant was discarded.

The mycelia was resuspended in TE25S buffer (5 ml) and lysozyme was added to a final concentration of 2 mg/ml. This mixture was incubated at 37 °C for 40 mins. Then proteinase K (50 μl) and 10 % SDS (300 μl) were added and the mixture incubated at 55 °C for 1 h. 5M Sodium chloride solution (1 ml) was then added, followed by phenol/chloroform (1:1 v/v; 5 ml) and the mixture was mixed by inversion for a period of 30 mins, then the mixture was centrifuged (5000 rpm, 5180 *g*, 5 mins). The aqueous layer was transferred to a clean tube and DNA was precipitated by the addition of 3M sodium acetate (1.5 ml) and isopropanol (6 ml). The mixture was centrifuged (13000 rpm, 16060 *g*, 10 mins) and the supernatant was discarded. 80 % Ethanol was carefully added without disturbing the precipitated DNA at the bottom of the tube, the tube was centrifuged (13000 rpm, 16060 *g*, 5 mins) and the supernatant was again discarded. The DNA was allowed to dry and then re-dissolved in DNase-free water (1.5 ml).

5.2.5 Purification of the 5-FDRPi

5.2.5.1 Over-expression of 5-FDRPi

The recombinant *E. coli* BL21 Gold cells containing the 5-FDRPi-pEHISTEV plasmid (restriction sites : *BamH* I and *Xho* I) were grown and induced by the auto-induction method described by Studier.[11]

An inoculum culture (20 ml) was grown in LB media containing 1 % glucose (w/v) and kanamycin (100 μg/ml) at 37 °C, 200 rpm for 16 h.

To each 2 L conical flask was added ZY media (372 ml), 1 M $MgSO_4$ (0.4 ml), 5052 solution (50X, 10 ml), NPS (20X, 20 ml) and kanamycin (100 μg/ml). Each flask was inoculated with the inoculum culture (5 ml), and then allowed to grow at 20 °C at 250 rpm for 36 h. The cells were harvested by centrifugation (4000 rpm, 3315 *g*, 20 mins).

5.2.5.2 Purification of the 5-FDRPi

The cell pellet was resuspended in lysis buffer (250 ml) and DNase (10 mg) and protease inhibitor cocktail tablets (Roche) were added to the suspension. The cells were lysed by passing through a cell disruptor (30 kPsi). The cell lysate was cleared by centrifugation (20000 rpm, 48384 *g*, 20 mins) and the supernatant was filtered through a 0.45 μm membrane.

The filtrates were loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated with lysis buffer. The purification was carried out using the Akta-Express Protein purification system. The column was washed with wash buffer and then eluted with elution buffer. The eluted fractions were passed through a Desalt 16/10 column (GE Healthcare) connected to the Akta-Express purification system and eluted with

desalt buffer. The fractions were concentrated by the use of a protein concentrator (Vivaspin concentrator, 30 kDa MWCO).

5.2.5.3 Removal of His$_6$-tag on the over-expressed 5-FDRPi

TEV protease (over-expressed from *E. coli*) was added to the solution of 5-FDRPi in desalt buffer (ratio 1:50, TEV:5-FDRPi). The mixture was placed on a tube rotating platform and the solution was gently mixed by slow rotation for 2 h at r. t.

The solution was then loaded to a Ni-Sepharose 6 FF column equilibrated with desalt buffer and the protein was eluted with desalt buffer. The fractions containing 5-FDRPi were pooled.

5.2.5.4 Size exclusion chromatography

5-FDRPi solution was loaded to a gel filtration column (Superdex 200) equilibrated with gel filtration buffer. The size exclusion chromatography was carried using the Akta-Express protein purification system at a flow rate of 1 ml/min.

The fractions containing 5-FDRPi were pooled and concentrated. The purity of the protein was assessed by SDS-PAGE.

The protein samples were then flash frozen in liquid nitrogen and stored at -80 °C.

5.2.6 Assays of the activities of native and site-directed mutants of 5-FDRPi

Mutagenesis PCR was carried out using the mismatched DNA primers specified in *Chapter 3* (Table 3.1). The 5-FDRPi-T vector recombinant plasmid was used as the template for the reaction. The *pFu* polymerase was used for this reaction and the extension time was adjusted for the length of the plasmid (*ca.* 6 kb). The mixture was allowed to complete 16 cycles of denaturation at 95 °C (1 min), annealing at 55 °C (45 s) and extension at 68 °C (4 mins). The PCR products were purified from a gel with DNA gel extraction kit (Roche). The PCR products were purified into buffer Tango and *Dpn* I was added and the mixture was incubated for 2 h at 37 °C to remove parental DNA. The PCR products were then transformed into *E. coli* DH10B cells by electroporation and the cells were plated on an agar plate containing ampicillin. Plasmids were extracted from the cell colonies obtained.

The plasmids containing correct mutagenesis as confirmed by DNA sequencing were cut by restriction enzymes (*Nco* I, *Xho* I) and cloned into the pEHISTEV vector. The pEHISTEV constructs were then transformed into *E. coli* BL21 Gold cells.

The cells were cultured in LB media and protein production was induced at OD_{600} = 0.5 with IPTG (final concentration 0.5 mM) at 25 °C for 24 hours. After the cells were harvested by centrifugation (5000 rpm, 5180 g, 20 mins), they were resuspended in cold phosphate buffer (50 ml, 20 mM, pH 7.8), and the cells were lysed by sonication at 0 °C (10 x 1 min brusts, with 1 min pauses in between). Specifically, protease inhibitors and DNase were not added to the buffer. The lysate were cleared by centrifugation (20,000 rpm, 48,384 g, 20 mins) and the soluble fractions were passed through a Ni-NTA column (3 ml), and the column was washed with wash buffer (50 ml) and then eluted with elution buffer (10 ml). The eluted protein solutions were

dialysed against cold phosphate buffer (5 L, 20 mM, pH 7.8) and then concentrated to approx. 20 mg/ml by protein concentrators (10,000 MWCO). The protein identities were confirmed by mass spectrometry analysis.

The individual 5-FDRPi mutant assays were carried out as follows. A mixture (700 μl) containing synthetic 5'-FDA solution (2 mM final concentration), PNP-MBP (over-expressed from E. coli, 3 mg/ml) and the mutant isomerases/native isomerase [control] (3 mg/ml) in phosphate buffer (20 mM, pH 7.8) was incubated at 37 °C for 6 hours.

Then D_2O (100 μl) was added to each tube, and the mixtures were centrifuged to remove any precipitated protein before being loaded into NMR tubes. $^{19}F\{^1H\}$ NMR analyses were carried out to establish whether the product 5-FDRP and 5-FDRulP were produced by the enzymes.

5.3 Chapter 5 References

1. F. Hansske, D. Madej, M. J. Robins, *Tetrahedron*, 1984, **40**, 125-135.

2. M. J. Robins, S. Sarker, V. Samano, S. F. Wnuk, *Tetrahedron*, 1997, **53**, 447-456.

3. H. Ueki, T. K. Ellis, C. H. Martin, T. U. Boettiger, S. B. Bolene, V. A. Soloshonok, *J. Org. Chem.*, 2003, **68**, 7104-7107.

4. L. Williams, Z. Zhang, F. Shao, P. J. Carroll, M. M. Joullié, *Tetrahedron*, 1996, **52**, 11673-11694.

5. X. Zhang, W. A. van der Donk, *J. Am. Chem. Soc.*, 2007, **129**, 2212-2213.

6. P. H. H. Hermkens, J. H. V. Maarseveen, H. C. J. Ottenheijm, C. G. Kruse, H. W. Scheeren, *J. Org. Chem.*, 1990, **55**, 3998-4006.

7. W. J. Vloon, J. C. van Den Bos, N. P. Willard, G.-J. Koomen, U. K. Pandit, *Recl. Trav. Chim. Pays-Bas*, 2010, **110**, 414-419.

8. M. Sanada, T. Miyano, S. Iwadare, *J. Antibiotics* 1986, **39**, 304-305.

9. Z. Zhou, C. J. Fahrni, *J. Am. Chem. Soc.*, 2004, **126**, 8862-8863.

10. K. K. J. Chan, D. O'Hagan, *Meth. Enzymol.*, 2012, **516**, 291-235.

11. F. W. Studier, *Protein Expression Purif.*, 2005, **41**, 207-234.

Appendix 1

Crystallographic Data for (*R*)-*tert*-butyl 2,2-dimethyl-4-((*R*)-1-(2-phenoxyacetoxy)-3-(trimethylsilyl) prop-2-yn-1-yl)oxazolidine-3-carboxylate **116**
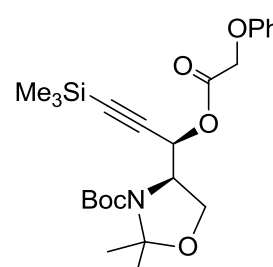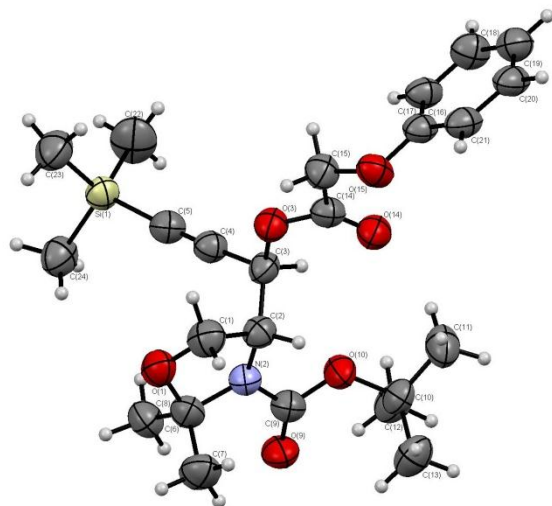
Empirical Formula	$C_{24}H_{35}NO_6Si$
Formula Weight	461.63
Crystal Colour, Habit	colourless, needles
Crystal Dimensions	0.300 X 0.020 X 0.020 mm
Crystal System	monoclinic
Lattice Type	Primitive
No. of Reflections Used for Unit	
Cell Determination (2q range)	1132 (82.2 - 139.0°)
Omega Scan Peak Width	
at Half-height	0.00°
Lattice Parameters	a = 11.92(2) Å
	b = 6.307(7) Å
	c = 17.82(2) Å
	β = 96.53(4) °
	V = 1331(3) Å^3
Space Group	P2$_1$ (#4)
Z value	2
D$_{calc}$	1.151 g/cm^3
F$_{000}$	496.00
μ(CuKα)	10.753 cm^{-1}

Appendix 2

Crystallographic Data for (2*S*,3*R*)-2-((*tert*-Butoxycarbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)-5-(trimethylsilyl)pent-4-ynoic acid 127
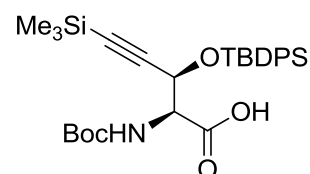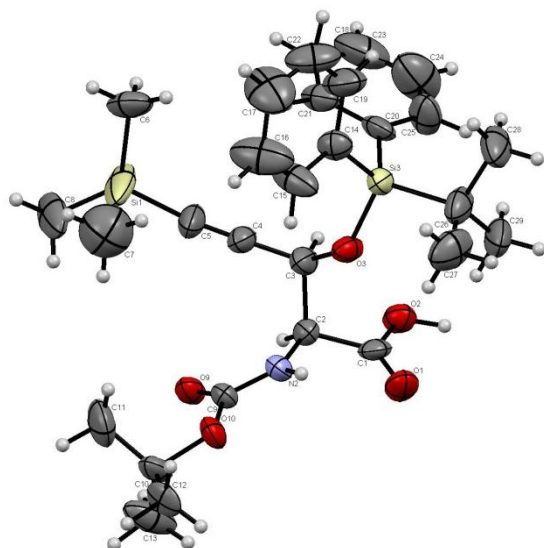
Empirical Formula	$C_{29}H_{41}NO_5Si_2$
Formula Weight	539.82
Crystal Colour, Habit	colourless, prism
Crystal Dimensions	0.200 X 0.200 X 0.100 mm
Crystal System	monoclinic
Lattice Type	Primitive
Lattice Parameters	a = 9.557(2) Å
	b = 12.285(2) Å
	c = 14.320(2) Å
	β = 106.775(6) °
	V = 1609.9(4) Å^3
Space Group	$P2_1$ (#4)
Z value	2
D_{calc}	1.114 g/cm^3
F_{000}	580.00
μ(CuKα)	12.767 cm^{-1}

Appendix 3

5-FDRPi nucleotide sequence

SCATT20080

Streptomyces cattleya gene for 5-deoxy-5-fluoro-α-D-ribose 1-phosphate isomerase

1161 bases

```
ATGGGTGATCAGTCCGTACAGCCTTTGGCCAAGGGCACGGGGTCCGGGACCCCGGAGCCGAAACCCGCTCTCCGC
TGGGAAGAGCCTCCCGAAGGGCCCGTGCTGGTCCTCCTCGACCAGACCCGGCTCCCCGTCGAGGAGGTCGAACTG
TTCTGTACGGACGTGCCCGCGCTCGTCCAGGCCATCCGTACCCTCGCCGTCCGCGGCGCGCCGCTGCTCGGGCTC
GCCGGAGCGTACGGCGTCGCCCTGGCCGCCGCCCGTGGCTACGACGTCGGGCAGGCCGCCGACGAACTCGCCGGC
GCCCGGCCCACCGCCGTCAACCTCTCCTACGGGGTGCGCCGCGCGCTGGCCGCGTACCGTACCGCGGTCACCGGC
GGCGCCGACGACACGGGCGCGGCGGCGGCCACCCTCGCCGAGGCCCGCGCGCTGCACGCCGAGGACGCCAGGGCC
AGCGAACGCATGGCCCGCAACGGCCTGGCGCTGCTGGACGAACTCGTCCCCGGCGGCGGCTACGGGTGCTGACC
CACTGCAACACCGGCGCCCTGGTCTCCGGCGGCGAGGGCACCGCCCTGGCCGTCGTCCTCGCCGCCCACCGCGGC
GGACTCCTGCGCCGGCTGTGGGTGGACGAGACCCGGCCGCTGCTCCAGGGCGCCCGGCTGACCGCCTACGAGGCC
GCCCGGGCCGGCGTCGCCCACACGTTGCTGCCGGACGGCGCGGCCGGGTCGCTCTTCGCGGCCGGCGAGGTCGAC
GCGGTGCTGATCGGCGCCGACCGGATCGCCGCGGACGGCTCGACCGCCAACAAGGTCGGCAGCTACCCGCTGGCC
GTCCTCGCCCGGTACCACAACGTCCCCTTCGTCGTGGTCGCCCCCACCACCACGATCGACCTGGCCACCCCCGAC
GGCACCGCGATCGAGGTGGAGCAGCGCCCCGCGCAGGAGGTGACCGAGCTGACCGGACCGCGCCCCGGCCCGGAC
CGCGAGGGCGCCACCGGCATCCCCGTCGCGCCCCTGGGCACGCCGGCGTACAACCCGGCGTTCGACGTCACCCCG
CCCGAACTGATCACCGCCGTGGTCACCGAGACCGGCGTGGCCTCCCCGGTCACCGGCTCCTCCATAGCCGCCCTG
GCCGCCCGCCCCGGCCCCGTCCGCGCCCAGCCGTGA
```

5-FDRPi peptide sequence

emb CAR92348.1

Streptomyces cattleya peptide for 5-deoxy-5-fluoro-α-D-ribose 1-phosphate isomerase

386 residues; MW = 39.23 kDa

```
MGDQSVQPLAKGTGSGTPEPKPALRWEEPPEGPVLVLLDQTRLPVEEVELFCTDVPALVQAIRTLAVRGAPLLGL
AGAYGVALAAARGYDVGQAADELAGARPTAVNLSYGVRRALAAYRTAVTGGADDTGAAAATLAEARALHAEDARA
SERMARNGLALLDELVPGGGYRVLTHCNTGALVSGGEGTALAVVLAAHRGGLLRRLWVDETRPLLQGARLTAYEA
ARAGVAHTLLPDGAAGSLFAAGEVDAVLIGADRIAADGSTANKVGSYPLAVLARYHNVPFVVVAPTTTIDLATPD
GTAIEVEQRPAQEVTELTGPRPGPDREGATGIPVAPLGTPAYNPAFDVTPPELITAVVTETGVASPVTGSSIAAL
AARPGPVRAQP
```