UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

School of Chemistry

Studies on Thiamine Biosynthesis

by

Filipa Teixeira Martins

Thesis presented for the degree of Doctor of Philosophy

September 2009

To my parents and to my Sensei

"A university's mission is not to produce people of authority but people of true ability, intelligence and conviction who will dedicate themselves to the good of the people. Can universities raise and send individuals into society capable of shouldering the next age? The answer to this vital question will determine not only the victory or defeat of the academic world but the destiny of the age as well."

Daisaku Ikeda in

"Soka Education – A Buddhist vision for teachers, students and parents"

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF CHEMISTRY

Doctor of Philosophy

STUDIES ON THIAMINE BIOSYNTHESIS

by Filipa Teixeira Martins

In *Escherichia coli*, and other prokaryotes, thiamine (vitamin B1) is assembled by coupling 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (Hmp-PP) and 4-methyl-5-(β-hydroxyethyl)thiazole phosphate (Thz-P). The thiazole moiety is biosynthesised from 1-deoxyxylulose-5-phosphate, tyrosine, and cysteine, and at least six genes are required including *thiH*, *thiG*, *thiS*, and *thiF*. Whilst in aerobes, the C2-N3 fragment of Thz-P derives from glycine in a reaction catalysed by the flavoenzyme ThiO, in anaerobes such as *E. coli*, dehydroglycine is formed from tyrosine in a ThiH dependent reaction. This biosynthetic step requires the cleavage of the Cα-Cβ bond of tyrosine and a release of an aromatic side chain. ThiH shows sequence similarity with the 'radical S-adenosylmethionine' (AdoMet) family of proteins, including conserved cysteine ligands to the essential [4Fe-4S] cluster and has been shown to form a complex with ThiG. With the purpose of studying the mechanistic enzymology by which Thz-P is assembled it was crucial to isolate ThiH in the *holo*-form. Several expression systems and purification methodologies were investigated. The optimisation of the purification method, together with *in vitro* chemical reconstitution with exogenous iron and sulfide allowed the successful isolation of *holo*-ThiH.

To facilitate the mechanistic investigation of Thz-P biosynthesis, an *in vitro* assay was developed, and the reaction products formed *in vitro* were elucidated and quantified. The aromatic by-product derived from the side chain of tyrosine is *p*-cresol and the remaining fragment yields glyoxylate, a product of hydrolysis of dehydroglycine.

TABLE OF CONTENTS

List of figures

List of schemes

List of tables

Declaration of authorship

I, Filipa Teixeira Martins, declare that the thesis entitled 'Studies on thiamine biosynthesis' and the work presented in it are my own.
I also confirm that:

- this work was carried out while in the candidature for a research degree at this University;
- I have clearly stated where any part of this thesis has been previously submitted for a degree or any other qualification at this University or any other institution;
- I have clearly attributed and quoted where I have consulted published work of others, and the source is always given;
- I have acknowledge all main sources of help
- Where my thesis is based on work carried out by myself jointly with others, I have made it clear what was done by others and what I contributed myself;
- Part of this work has been published as:

  - Kriek, M., Martins, F., Leonardi, R., Fairhurst, S. A., Lowe, D. J., and Roach, P. L. (2007) Thiazole synthase from *Escherichia coli*: an investigation of the substrates and purified proteins required for activity *in vitro*. *J Biol Chem 282*, 17413-23
  - Kriek, M., Martins, F., Challand, M. R., Croft, A., and Roach, P. L. (2007) Thiamine biosynthesis in *Escherichia coli*: identification of the intermediate and by-product derived from tyrosine. *Angew Chem Int Ed Engl 46*, 9223-6

Signed:

Date: September 2009

Acknowledgments

I would like to thank Dr. Peter Roach for his supervision during this project; for his enthusiasm, brilliant ideas, help and advice. But specially for giving me the opportunity to work in this project, where I feel that I have developed has a scientist.

I would like to thank Dr. Marco Kriek for all the work done during this project. In particular for the development of the *in vitro* assay for ThiH, the development of chemical reconstitution and concentration of ThiH methods, and the purification of flavodoxin and flavoprotein. Also, for all the preliminary studies that provided very useful information and allowed for this project to be a success. But above all this, I am very grateful for all his help, support and hard work throughout this project, because without him it would have just taken so much longer to get there!

I would like to thank Dr. Rob Wood, for all his help throughout my project and for proofreading chapter 2 and 6 of this thesis.

I would like to thank Dr. Gabriel Cavalli for his support and for proofreading the introduction chapter of this thesis.

I would like to thank Martin Challand for proofreading chapters 1 and 4 of this thesis but mainly for his great enthusiasm in the last stages of my stay at the 'Roach Group'. "You really lifted my spirit and kept my interest for science going during the writing up of my thesis…thank you!".

And because a PhD is not only about science…

During my PhD years many people outside the lab have supported me in many ways…I would like to show here my deepest appreciation to all of them.

First of all, I would like to warmly thank my parents, Maria Helena and Armindo, who have always supported me and made so many efforts to help me get through with my PhD. I would also like to thank my brother Zé and his wife Maria João for their warmth, for being good listeners and for all the cheering up in moments of doubt.

I would like to thank my bebé Lefteris, for being so patient and supportive in moments where even I was feed up with myself…for all the daimoku that I could so deeply feel. For making me laugh, for making me chant…for the pushing and the cuddling…"tough love is necessary at times and you gave it in the right amount. Thank you!".

I would also like to thank my good friend Naruo Yoshikawa for bringing into my life the greatest tool of the universe, because without it this work wouldn't have been possible.

I would especially like to thank my Sensei for his constant encouragement, for always believing in my potential; because in all the moments when I didn't believed in myself and life was hard he was the only one that was always there for me! "Thank you so much Sensei!".

I would like to thank all the Soka Gakkai members, from UK and Europe, for the incredible support and training during the past three years of my life…for all the joyful activities…for all the smiles…for all the words of encouragement. Especially to my good friends: Gabriel, Adeline, Ashley, Shoko, Gary and Scott thank you so much!

I would like to thank all the people that one way or another made may PhD years more enjoyable. In particular, I would like to thank my good friend Hugo for all the laughs and especially for his eternal friendship; to Ricardo, Carla and Violeta for all the nonsense conversations outside the lab and for their unconditional support.

Abbreviations

Abs or A_x	absorbance (at x nm)
Ado•	5′-deoxyadenosyl radical
AdoH	5′-deoxyadenosyl
AdoMet	S-adenosyl-L-methionine
AIR	5-aminoimidazole ribotide
amp	ampicillin resistance gene
arabinose	β-(+)-L-arabinose
araC	Gene encoding AraC, the arabinose operon regulatory protein
ATP	Adenosine triphosphate
B. subtilis	*Bacillus subtilis*
BioB	Biotin synthase
C. jejuni	*Campylobacter jejuni*
CoA	Coenzyme A
C. tepidum	*Chlorobaculum tepidum*
C. thermocellum	*Clostridium thermocellum*
Cys	Cysteine
DTB	Dethiobiotin
DTT	Dithiothreitol
D. vulgaris	*Desulfovibrio vulgaris*
Dxp	1-deoxy-D-xylulose-5-phosphate
E. coli	*Escherichia coli*
EPR	Electron paramagnetic resonance
ESI-MS	Electrospray ionisation mass spectrometry
FldA	Flavodoxin
FM	Filipa Martins (author)
FPLC	Fast protein liquid chromatography
Fpr	Flavoprotein
FT	Flow through
g	Gram

GC-MS	Gas chromatography – mass spectrometry
h	Hour
HBA	4- hydroxybenzyl alcohol
HE-TPP	hydroxyethylidene-thiamine pyrophosphate
His-tag	Hexa-histidine tag
Hmp	4-amino-5-hydroxymethyl-2-methylpyrimidine
I.D.	Internal diameter
kb	Kilo basepairs
KDa	Kilo Daltons
L	Litre
LipA	Lipoyl synthase
MCS	Multiple cloning site
MOPS	3-(N-Morpholino)propanesulfonic acid
M. thermoacetica	*Moorella thermoacetica*
MK	Marco Kriek
MWCO	Molecular weight cut off
NAD^+	Nicotinamide adenine dinucleotide, reduced
NADH	Nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate, reduced
NADPH	Nicotinamide adenine dinucleotide phosphate
nm	Nanometer
min	Minute
M	Molecular weight marker
NMR	Nuclear magnetic resonance
OD_x	Optical density at x nm
OPD	*o*-phenylene diamine
ORF	Open reading frame
p	Plasmid
P	Pellet
PEI	Polyethyleneimine
PfAE	Pyruvate formate lyase active enzyme

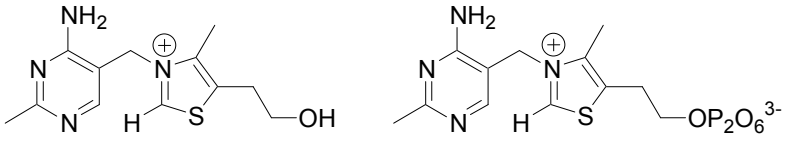
PI	Isoelectric point
PMSF	Phenylmethylsulfonyl fluoride
R^2	Square of the correlation coefficient
RL	Roberta Leonardi
RP-HPLC	Reverse phase high performance liquid chromatography
s	Second
S	Supernatant
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
S. cerevisiae	*Saccharomyces cerevisiae*
S. typhimurium	*Serovar typhimurium*
T	Thiamine
Thz-P	4-methyl-5-(β-hydroxylethyl)thiazole phosphate
T. maritima	*Themotoga maritime*
T. tengcongensis	*Thermotoga tengcongensis*
thiH-His	Gene encoding ThiH-His
ThiH-His	ThiH with a C-terminal hexahistidine tag
ThiS-COSH	ThiS thiocarboxylate
TP	Thiamine monophosphate
TPP	Thiamine pyrophosphate
Tris	tris(hydroxymethyl)aminomethane
Tyr	Tyrosine
U	Unit
UV	Ultraviolet
V. vulnificus	*Vibrio vulnificus*
W	Watts

CHAPTER 1

Introduction

1.1 Overview of Thiamine

Thiamine **1**, also known as vitamin B1, is a water soluble vitamin and a essential component of human diet, with a recommended daily allowance of 1.4 mg[1]. The main biologically active form of thiamine is thiamine pyrophosphate **2** (TPP). TPP is an essential cofactor of several enzymes involved in important metabolic pathways. The biosynthesis of this cofactor occurs in most microorganisms and higher plants. Humans are not able to produce this vitamin and its deficiency seriously affects cardiovascular (wet beriberi) and nervous systems. The most devastating neurological disorder known to be caused by thiamine deficiency is the Wernicke-Korsakoff syndrome, an alcohol related illness, that is the third most common cause of dementia in the United States[1-3]. The chemical structures of thiamine **1** and thiamine pyrophosphate **2** are shown in figure 1.1.



1 **2**

*Figure 1.1 Structure of thiamine **1** and thiamine pyrophosphate **2**.*

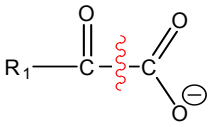
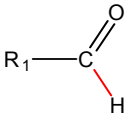
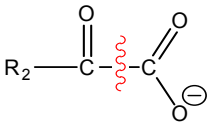
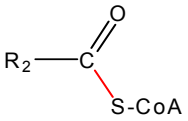
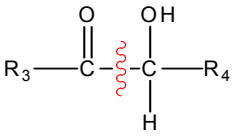
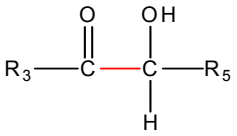
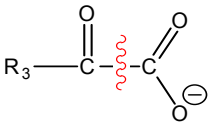
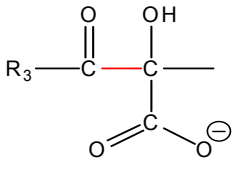
Thiamine was first discovered in 1910 by Umetaro Suzuki in Japan while he was investigating a cure for beriberi. At the time he did not determine the chemical constitution of thiamine and named it aberic acid. It was only later, in 1926, that vitamin B1 was crystallised by Jansen and Donath[4], two Dutch chemists, that at the

time published the wrong formula. Ten years later, Williams [5] finally elucidated the correct formula and soon after that the first chemical synthesis for thiamine was reported. Thiamine is commercially available and widely used as a nutritional supplement[6].

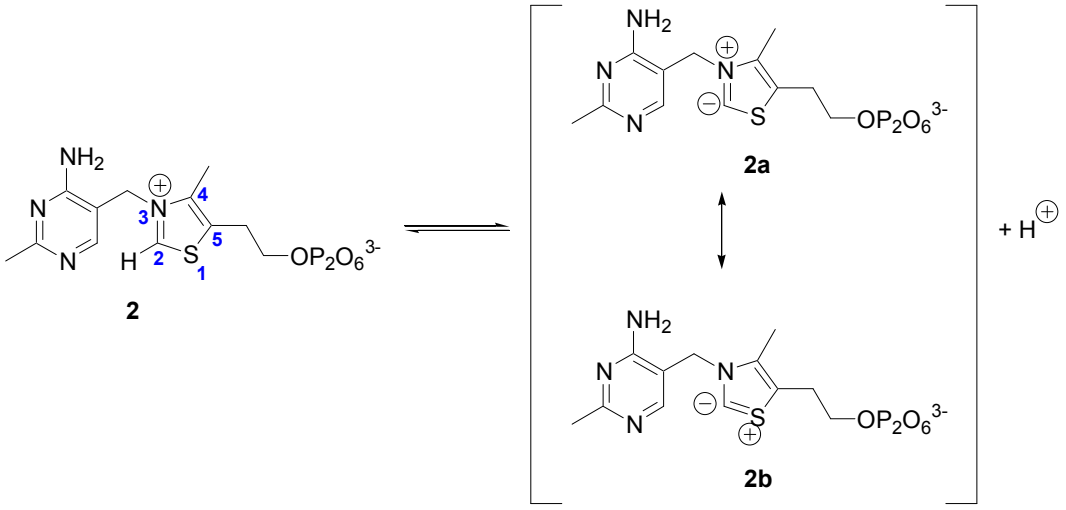
1.2 Thiamine pyrophosphate: importance and biochemical functions

Thiamine pyrophosphate plays an important role in the cleavage of bonds adjacent to a carbonyl group, such as the decarboxylation of α-keto acids, and in chemical rearrangements in which an activated aldehyde group is transferred from one carbon to another[7] (table 1.1).

Table 1.1 Examples of TPP-dependent reactions. Adapted from[7].

Enzyme	*Pathway*	*Bond cleaved*	*Bond formed*
Pyruvate decarboxylases	Alcohol fermentation		
Pyruvate dehydrogenase	Synthesis of acetyl-CoA		
α-Ketoglutarate dehydrogenase	Citric acid cycle		
Transketolase	Carbon-fixation reactions of photosynthesis		
Acetolactate synthase	Valine, leucine biosynthesis		

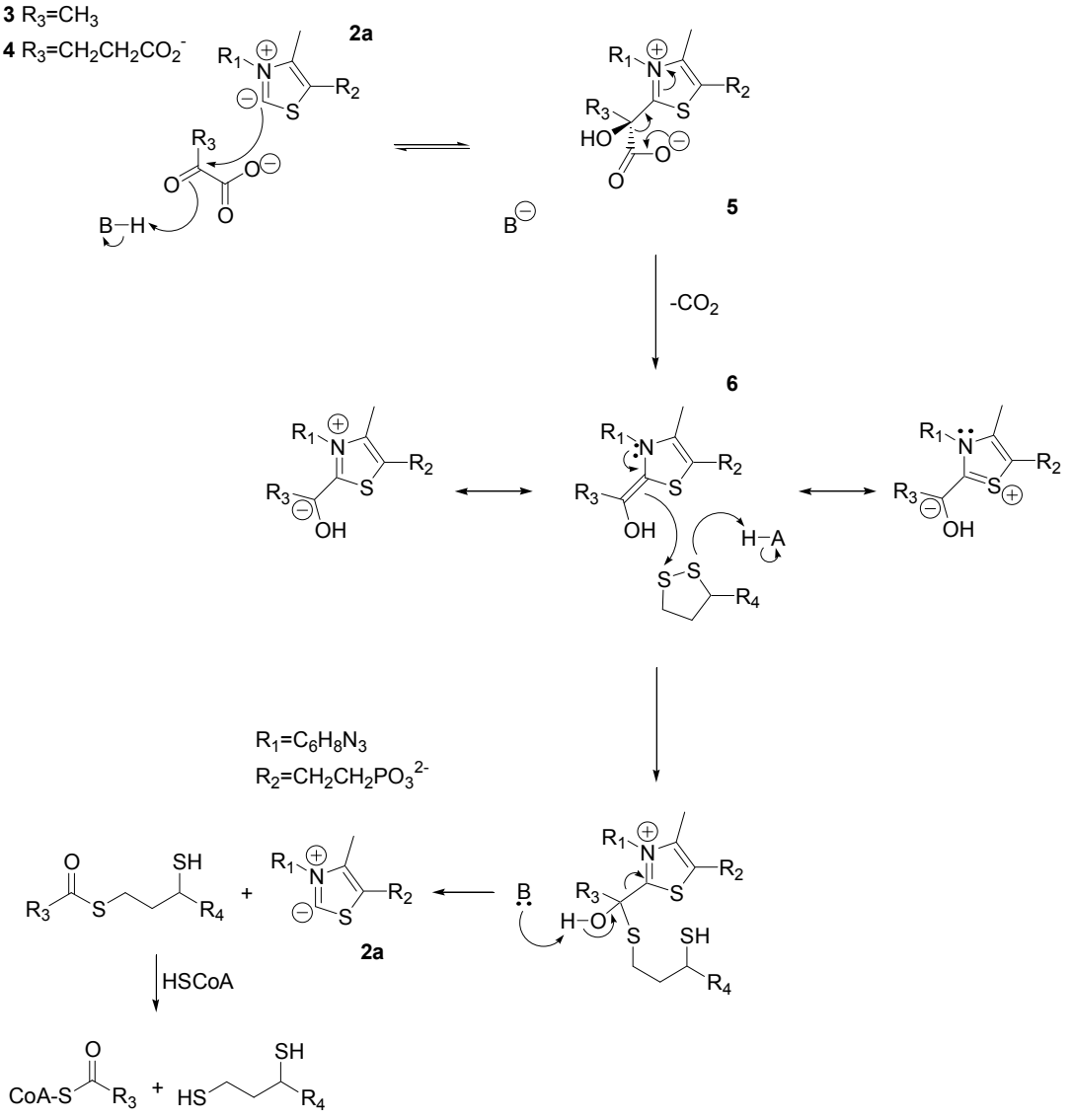
In most TPP-dependent enzymes the loss of the proton at C-2 of the thiazole ring is the first step in the catalysis. This proton is relatively acidic [pKa of 17-19[8,9]], due to the stabilisation of the corresponding carbanion **2a** by the adjacent sulfur and by the positive charge on the nitrogen[10] (scheme 1.1). This carbanion is both a potent nucleophile and a good leaving group. The subsequent catalytic steps are characterised by the nucleophilic attack of **2a** on the substrate, followed by cleavage of a C-C bond. This cleavage releases the first product, leaving a second, resonance stabilised carbanion (enamine).



Scheme 1.1 Formation of the stabilised carbanion of TPP.

Pyruvate dehydrogenase and α-ketoglutarate dehydrogenase are two important enzymes in the Krebs cycle. The E1 components of these complexes, catalyse the decarboxylation of pyruvate **3** and α-ketoglutarate **4**, respectively[11], through the mechanism shown in scheme 1.2. The nucleophilic addition of the resonance stabilised thiazolium carbanion to the carbonyl group of the α-keto acid, followed by protonation, leads to the formation of a α-hydroxyacid **5**. This adduct readily undergoes decarboxylation producing hydroxyethylidene-TPP **6** (HE-TPP), stabilised by resonance. This hydroxyethylidene group is then oxidised and transferred as an acetyl into to the lipoamide prosthetic group of the pyruvate dehydrogenase and α-ketoglutarate dehydrogenase E2 components, releasing **2a**. The E2 components of the respective complexes then transfer the acetyl or succinyl groups to coenzyme A (CoA),

with the oxidised form of the lipoamide being regenerated by the E3 component, which is identical in the two dehydrogenases. Decreased activities of pyruvate dehydrogenase and α-ketoglutarate dehydrogenase reduces adenosine triphosphate (ATP) synthesis, and lower ATP levels have been found in brain regions damaged by thiamine deficiency[1].



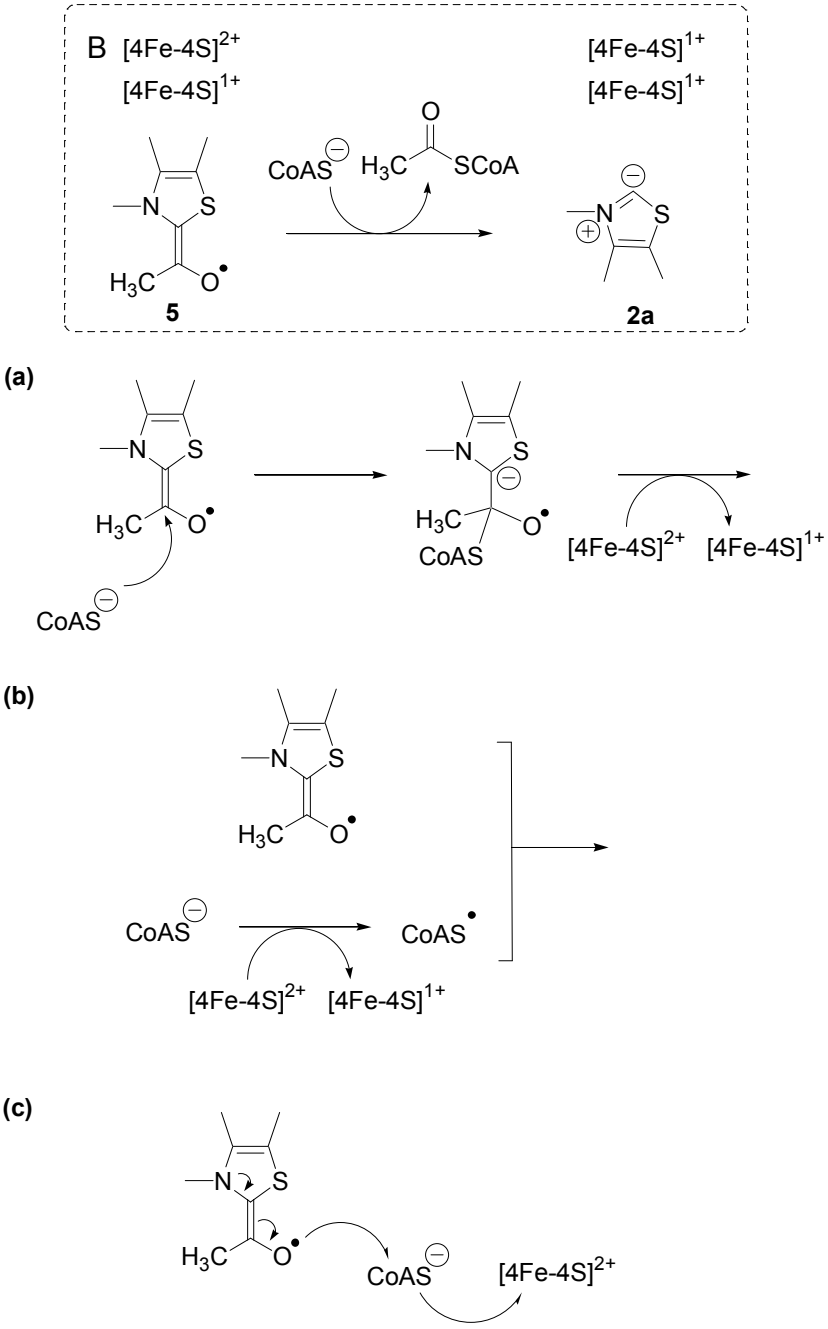
Scheme 1.2 Decarboxylation of α-keto acids catalysed by the TPP-dependent E1 components of pyruvate and α-ketoglutarate dehydrogenases.

Whilst most TPP dependent enzymes follow the Breslow mechanism[12] for TPP activation, the fate of the intermediate formed upon decarboxylation, HE-TPP **6**, differs amongst TPP dependent enzymes. An example of this is pyruvate ferredoxin oxidoreductase (PFOR), a thiamine dependent enzyme that catalyses the anaerobic oxidation of pyruvate (with CoA) to acetyl-CoA. In PFOR, the oxidation of this intermediate is associated with the reduction of an external electron acceptor, for example, the clostridial ferredoxin, which contains three [4Fe-4S] clusters (cluster A, B and C)[13].

$$CoASH + pyruvate + ferredoxin_{ox} \rightleftharpoons H_3C - CO - SCoA + CO_2 + ferredoxin_{red}$$

After the formation of HE-TPP, one electron is transferred to one of the three $[4Fe\text{-}4S]^{2+}$ clusters, which generates a paramagnetic $[4Fe\text{-}4S]^{1+}$ cluster and a HE-TPP radical[14]. A subsequent electron transfer from HE-TPP radical to cluster B ($[4Fe\text{-}4S]^{2+}$) generates a second $[4Fe\text{-}4S]^{1+}$ cluster. The stability of the HE-TPP radical is dependent on the presence of CoA[13].

The subsequent steps upon formation of the HE-TPP radical intermediate were studied in detail by Furdui and Ragsdale[14] and three mechanisms have been proposed (scheme 1.3); (a) kinetic coupling mechanism: the nucleophilic attack by CoA thiolate on the HE-TPP radical generates an anion radical that transfers one electron to cluster B; (b) biradical mechanism: an electron transfers from CoA to cluster B generating a thiyl radical that combines with the HE-TPP radical; (c) wire mechanism: CoA induces an electron transfer pathways between the HE-TPP radical and cluster B[14].

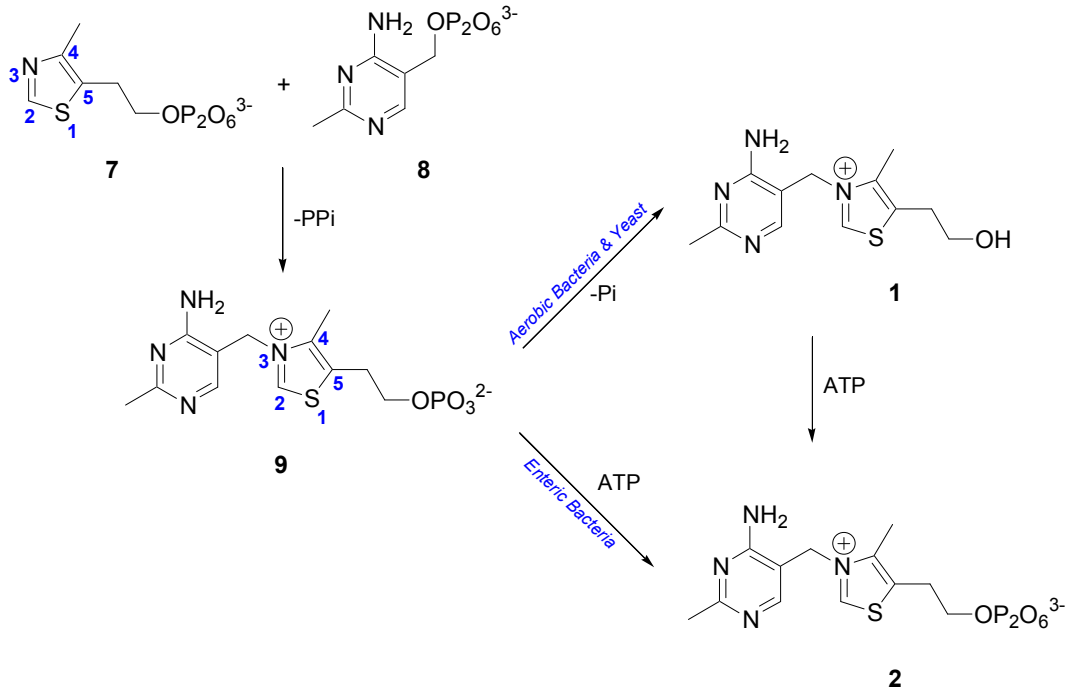


Scheme 1.3 Proposed mechanisms for the formation of the thiazole carbanion 2a via an HE-TPP radical intermediate; (a) kinetic coupling mechanism; (b) biradical mechanism; and (c) wire mechanism. Adapted from[14].

1.3 Thiamine biosynthesis: prokaryotes and eukaryotes

The biosynthesis of thiamine involves the separate biosynthesis and subsequent coupling of 4-methyl-5-(β-hydroxyethyl) thiazole phosphate **7** (Thz-P) and 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate **8** (Hmp-PP) to form thiamine monophosphate **9** (TP). These two precursors are biosynthesised through independent pathways that are considerably different in prokaryotes and eukaryotes.

All organisms able to produce thiamine first assemble thiamine monophosphate (TP), which is then converted to thiamine pyrophosphate (TPP), either by direct phosphorylation (in enteric bacteria) or by de-phosphorylation to thiamine followed by pyrophosphorylation (in aerobic bacteria and yeast) (scheme 1.4)[6].



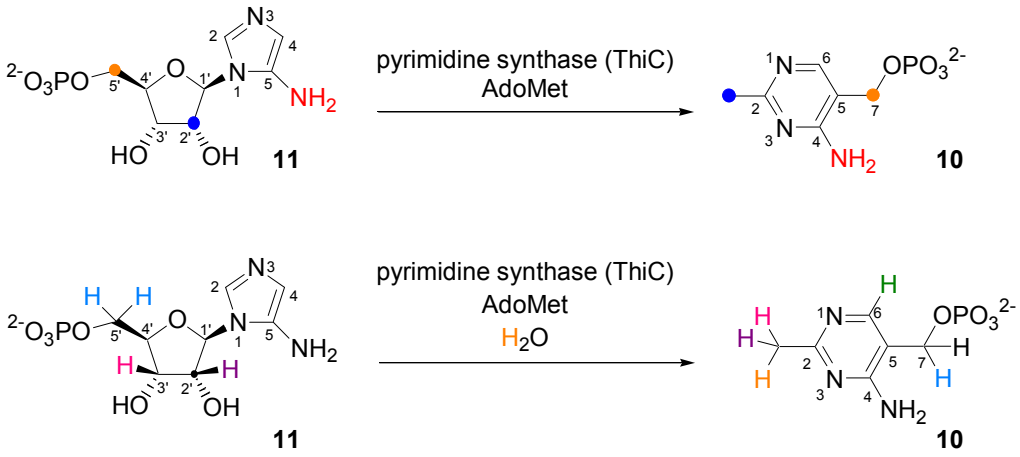
*Scheme 1.4 Assembly of thiamine pyrophosphate from Thz-P **7** and Hmp-PP **8** in enteric bacteria and aerobic bacteria and yeast.*

In prokaryotes significant progress has been made in the identification of the precursors and genes involved, comprehensive reviews of this topic are those by White and Spenser[6,15]. In eukaryotes, however, the details have only recently begun to emerge. The following paragraphs summarise the biosynthesis of these two precursors in prokaryotes and eukaryotes, with particular emphasis on the biosynthesis of Thz-P.

Biosynthesis of 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (Hmp-PP)

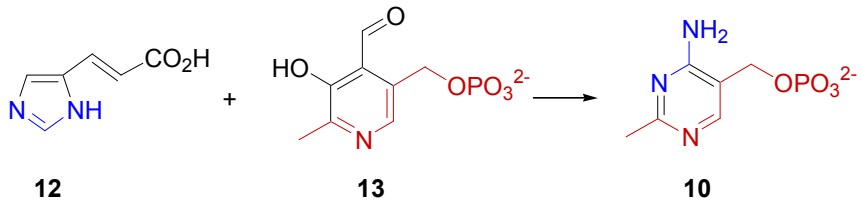
Studies on the prokaryotes *E. coli and S. typhimurium*, established that 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate **10** (Hmp-P) derives from 5-aminoimidazole ribotide **11** (AIR)[16-18]. It is still unclear if the direct product of the pyrimidine-forming reaction is Hmp or Hmp-P, since AIR is phosphorylated. However, because the enzyme ThiD can catalyse the phosphorylation of 4-amino-5-hydroxymethyl-2-methylpyrimidine (Hmp) to Hmp-P and then on to Hmp-PP, Hmp is generally accepted to be the first product of the pyrimidine biosynthesis.

This remarkable rearrangement of aminoimidazole ribonucleotide is catalysed by the ThiC gene product. Isotopic labelling studies (scheme 1.5) indicate that all the carbon and nitrogen atoms of Hmp derive from AIR, an intermediate in both *de novo* purine and histidine biosynthesis[16,17,19,20]. These studies suggest that the C5 of the pyrimidine derives from C4′ of AIR, the C2 methyl substituent originates from C2′ of AIR, and also that the nitrogens of the aminoimidazole are retained in the pyrimidine. Moreover, studies show that pyrimidine synthase activity is dependent on an iron sulfur cluster and S-adenosyl methionine (AdoMet) due to the production of 5-deoxyadenosine radical (Ado•), which suggests that this enzyme belongs to the ‘radical AdoMet’ family (discussed in detail in section 1.4). The mechanism of this rearrangement has not yet been elucidated[21].



*Scheme 1.5 Biosynthesis of Hmp-P **10** from AIR **11**.*

The biosynthesis of the thiamine pyrimidine in *S. cerevisiae*, derives from histidine **12** and pyridoxal-5′-phosphate **13** (PLP) in a reaction where the THI5 gene is involved. The mechanism for this rearrangement is not yet elucidated[22]. Scheme 1.6 summarises this conversion.



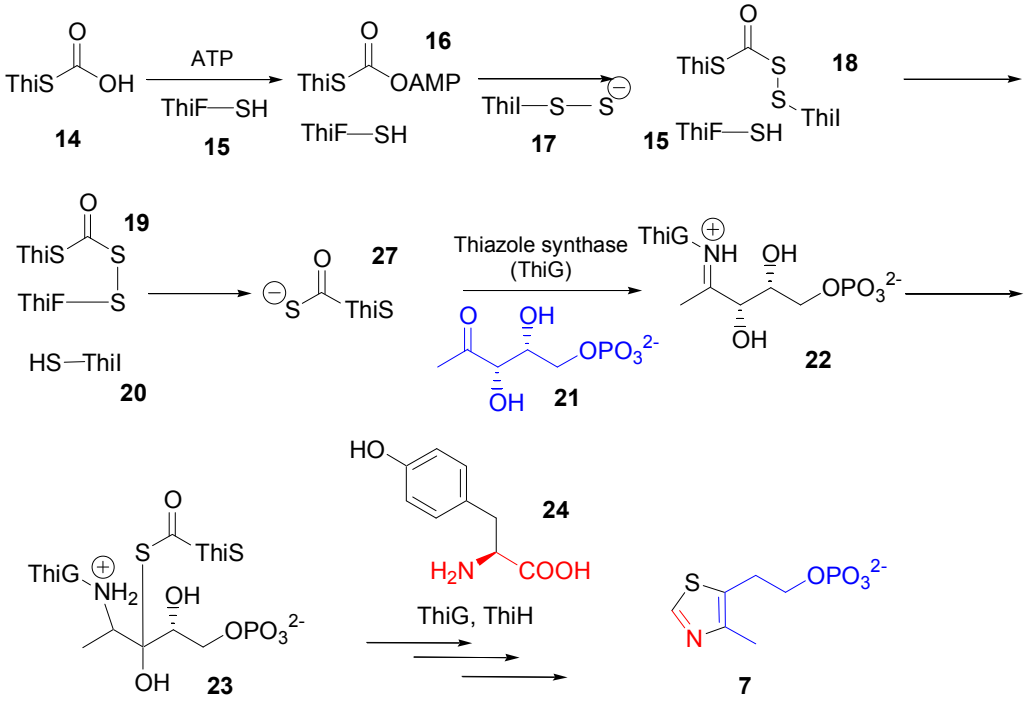
*Scheme 1.6 Biosynthesis of Hmp-P **10** from histidine **12** and PLP **13**.*

Zeidler and co-workers used incorporation studies with ^{13}C, ^{15}N and ^{2}H labelled substrates, with subsequent NMR analysis to show that Hmp originates from a NCHNH fragment that derives from histidine **12** (blue), and a $CH_3CNCHCCH_2OPO_3^{2-}$ fragment that derives from PLP **13** (red)[22].

Biosynthesis of 4-methyl-5-(β-hydroxyethyl) thiazole phosphate (Thz-P)

Considerable progress has been made recently on the elucidation of the biosynthetic pathway by which 4-methyl-5-(β-hydroxyethyl) thiazole phosphate (Thz-P) is assembled. This thiazole moiety, required for thiamine biosynthesis, has two different biosynthetic pathways in prokaryotes and eukaryotes.

In prokaryotes, the precursors and proteins involved differ slightly in aerobic and anaerobic bacteria. Common to both are the precursors 1-deoxy-D-xylulose-5-phosphate (Dxp), cysteine (Cys) and the proteins ThiF, ThiS, ThiI, ThiG and IscS. *E. coli* and other enteric bacteria depend on ThiH, an oxygen sensitive protein, which utilises tyrosine (Tyr) to provide the C2-N3 fragment of the thiazole[23,24]. *B. subtilis*, and other aerobes, substitute ThiH with ThiO, an oxygen dependent flavoenzyme, that uses glycine (Gly) to provide the C2-N3 unit[25,26].



Scheme 1.7 Thz-P biosynthetic pathway in E. coli. Adapted from[27].

IscS, an enzyme also involved in iron-sulfur cluster biosynthesis, catalyses the sulfur transfer from cysteine to ThiI to give ThiI-persulfide **17** (scheme 1.7)[28,29]. ThiF catalyses the adenylation of ThiS, at its carboxy terminus to give **16** and AMP is subsequently displaced by the ThiI-persulfide to give **18**. Disulfide interchange, involving Cys184 on ThiF, generates the mixed acyl disulfide **19**, which acts as a sulfur donor for thiazole formation[29,30]. The intermediates **22** and **23** have been identified in studies using *B. subtilis* but are thought to be common intermediates in Thz-P in *E. coli*[31]. The investigation of the mechanistic enzymology by which thiazole **7** is assembled from tyrosine **24** in *E. coli* is the main focus of this project.

ThiS, a 7.2 kDa protein, has similar structure to ubiquitin and MoaD[32]. All three proteins have a low degree of sequence similarity aside from the carboxy-terminal Gly-Gly pair. ThiF, a 27 kDa protein, has similar structure to ubiquitin-like (Ubl) activating enzymes and also to the MoeB, a molybdopterin biosynthetic protein[33-36]. MoeB and MoaD, molybdopterin biosynthetic proteins, have similar function to those of the ThiF-ThiS complex. MoeB uses ATP and catalyses the formation of MoaD-thiocarboxylate which acts as a sulfur source in molybdenum cofactor biosynthesis. The relationship between ThiF-ThiS complex with ubiquitin is less direct, as instead of formation of an acyl disulfide, ubiquitin forms a thioester. In this rearrangement, the ubiquitin-activating enzyme E1 catalyses the formation of an acyl adenylate of ubiquitin, which reacts with the cysteine on E1 to form a thioester. This thioester functions as an activated ubiquitin donor[37,38]. The functional similarity between these enzymes, ubiquitin-E1, ThiF-ThiS complex and MoaD-MoaB suggests that they share a common ancestor.

Studies by Settembre[31], elucidated the structure of the ThiG-ThiS complex and gave further information about the thiazole synthase active site and also about the similarities between the ThiG-ThiS complex and the ThiF-ThiS complex. ThiG-ThiS crystal structure indicates that the surface involved in the complex formation is the same as the surface suggested being involved in ThiS-ThiF complex formation[36]. Therefore, regardless of the structural difference existing between thiazole synthase and the proteins MoaE and E2-like proteins, a common binding motif is used for the

formation of the complex of ThiS, MoaD, and ubiquitin with their respective partners. For clarity, table 1.2 summarises the protein systems related to thiazole biosynthesis.

Table 1.2 Protein systems related to thiazole biosynthesis.

pathway	*Ubiquitin homolog*	*Activating enzyme (E1 homolog)*	*Conjugating enzyme/synthase (E2 homolog)*
Thiamine biosynthesis	ThiS	ThiF	ThiG
Molybdopterin biosynthesis	MoaD	MoeB	MoaE
Protein degradation	Ubiquitin	UBA1 (E1)	UBC (E2)

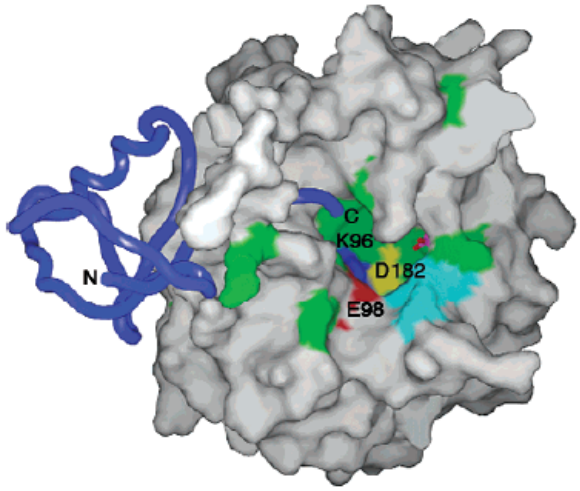
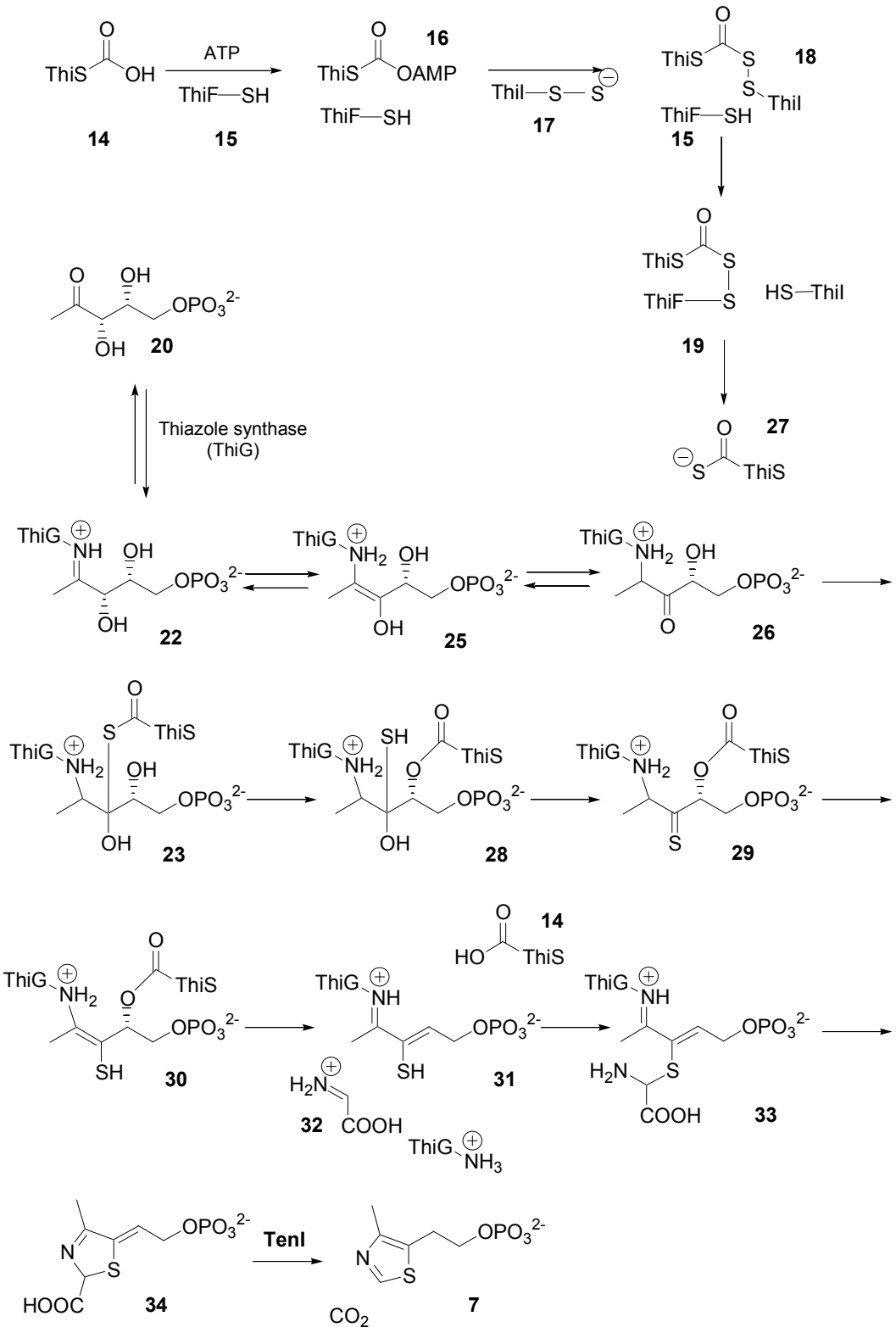


Figure 1.2 Putative active site of thiazole synthase. Thiazole synthase is shown as a surface representation, and ThiS is shown as a blue coil. Conserved residues with predicted functions are shown: Lys96 (blue), Glu98 (red), and Asp182 (yellow). Conserved residues without an assigned function are coloured in green, and an exposed hydrophobic surface is coloured in cyan. Phosphate is shown in red ball-and-stick representation. Adapted from[31].

The location of the ThiG active site was deduced bearing in mind several factors (figure 1.2): (a) thiazole synthase binding region was found to contain most of the conserved residues, which shows similarities to the thiamine phosphate synthase[39]; (b) a phosphate ion was found to bind to the region that contains a cluster of highly conserved residues; (c) Lys96, which forms an imine with Dxp, is closely positioned to the phosphate binding site; (d) the C-terminal appendage of ThiS extends to the region of thiazole synthase where Lys96 and the phosphate binding site are found. This indicated that the phosphate of Dxp most likely binds in identical way to the bound phosphate in this structure. This binding site was found to contain a P-loop identical to that of thiamine phosphate synthase for the binding of the phosphate to its substrate[31].

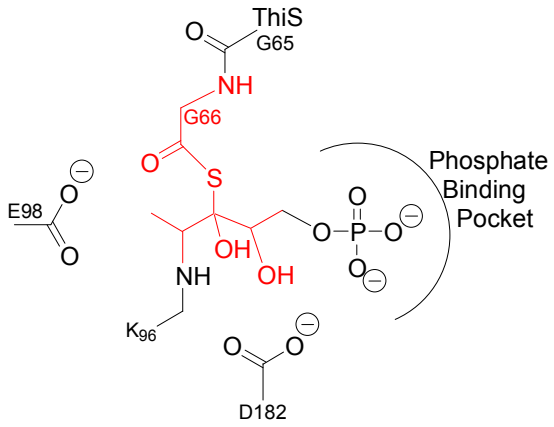
Studies on the aerobe *B. subtilis* by Begley and co-workers have identified five enzymes involved in the formation of Thz-P, ThiS, ThiF, ThiG, ThiO and IscS. These studies have further elucidated the biosynthetic pathway by which Thz-P is assembled in prokaryotes and a mechanism has been proposed (scheme 1.8)[40].

Dxp **21** forms an imine with Lys96 of the thiazole synthase (ThiG) to give **22**. Tautomerisation followed by addition of thiocarboxylate to the C3 ketone of **26** gives **23**, this intermediate then undergoes an S/O acyl shift followed by loss of water to yield thioketone **29**. Tautomerisation followed by the elimination of the ThiS **14**, the sulfur carrier protein, gives **31**. Addition of the glycine imine **32**, formed in a separate reaction, followed by transamination produces the thiazole tautomer **34**. The enzyme TenI is responsible for the aromatisation of **34** to Thz-P **7**. This mechanistic proposal has considerable experimental support[40-43].



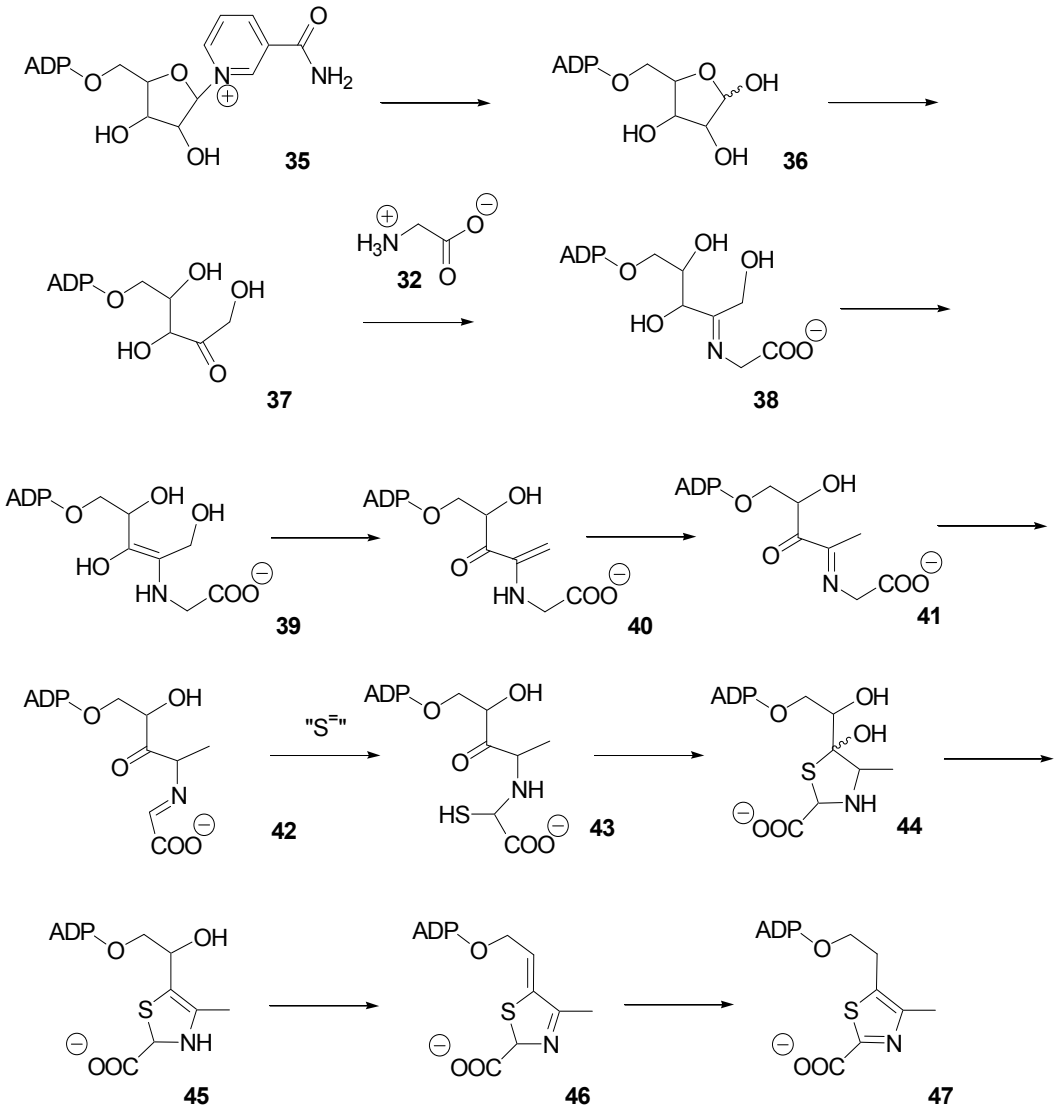
Scheme 1.8 Mechanistic proposal for the thiazole synthase (ThiG) catalysed formation of the thiazole moiety of Thz-P. Adapted from[44].

Based on this mechanistic proposal combined with the structure of thiazole synthase complexed to ThiS, Settembre and co-workers[31] further elucidated the catalytic mechanism of this complex by modelling the thiazole synthase active site (scheme 1.9). The residues likely to be involved in the catalysis are: Glu98 and Asp182. Glu98 has close proximity to Dxp, and Asp182 is close to the C3 and C4 of Dxp.



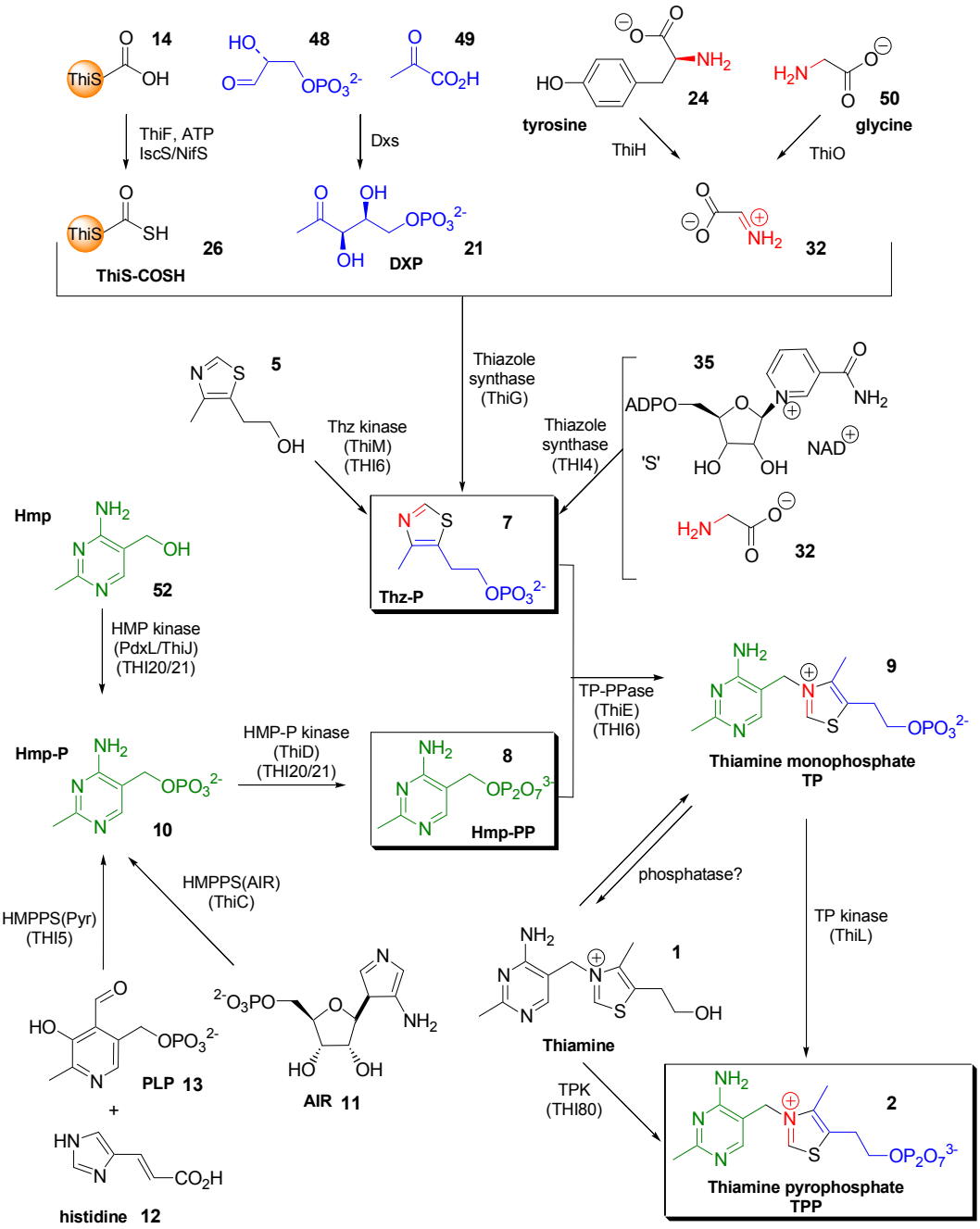
Scheme 1.9 Diagram representing Dxp bound to the thiazole synthase active site. The scheme shows a representation of Dxp and key amino acid side chains. Atoms shown in black were observed experimentally, and atoms shown in red were modelled using structural and mechanistic information. Adapted from[44].

The biosynthesis of Thz-P biosynthesis in eukaryotes has been shown to differ significantly. The proposed mechanism is outlined in scheme 1.10. The reaction is catalysed by the thiazole synthase (THI4), and whilst this reaction has not been yet fully reconstituted *in vitro*, THI4 was successfully isolated and revealed the presence of three tightly bound metabolites (**42**, **46**, **47**) and the structure of the complex between the enzyme and 46 has been determined[45-48]. These studies suggest that NAD is a precursor to the thiazole. The reaction is initiated by the cleavage of the *N*-glycosyl bond of NAD **35** followed by the opening of the ring and tautomerisation and loss of water to give **40**, this is followed by two tautomerisations and sulfide addition to yield **43**. This intermediate then cyclises to give **44**. The loss of two water molecules gives **46** followed by the thiazole aromatisation completes the reaction with the formation of **47**.



Scheme 1.10 Proposed mechanism for thiazole biosynthesis in S. cerevisiae. The sulfur source indicated by "S=" has not yet been identified. Adapted from[49].

The pathways of thiamine biosynthesis in prokaryotes and eukaryotes are summarised in scheme 1.11 and table 1.3 lists the enzymes of thiamine biosynthesis in *E. coli, B. subtilis* and *S. cerevisiae*. Comprehensive reviews of thiamine biosynthesis are those by White and Spenser[6,15], and more recently by Webb and co-workers[50].



Scheme 1.11 Thiamine biosynthesis in prokaryotes and eukaryotes. Adapted from[50].

Table 1.3 Gene names corresponding to enzyme activities of thiamine biosynthesis in E. coli, B. subtilis and S. cerevisiae. Adapted from[50].

Enzyme activity	*E. coli*	*B. subtilis*	*S. cerevisiae*
Thz-P biosynthesis			
S donor	IscS	NifS	
S activation	ThiS, ThiF	ThiS, ThiF	
Thz-P synthase	ThiG & ThiH	ThiG & ThiO	
Thz-P synthase			Thi4
HMP-PP biosynthesis			
HMPP kinase	ThiD	ThiD	Thi20/21
HMP synthase	ThiC	ThiA	
HMP synthase			Thi5/11/12/13
TPP biosynthesis			
TP synthase	ThiE	ThiC	Thi6
TP kinase	ThiL	ThiL	
TPK		ThiN	Thi80
Salvage enzymes			
HMP kinase	PdxL/ThiJ	ThiD	Thi20/21
Thiazole kinase	ThiM	ThiK	Thi6
Thiamine kinase	ThiK		

1.4 'Radical AdoMet' superfamily and ThiH

In prokaryotes, at least seven genes, *thiFSGH*, *thiI*, *iscS*, and *dxs* are required to convert Dxp, cysteine and tyrosine to Thz-P [24,29,30,51-54]. *In vitro* studies, in particular, using *B. subtilis* and *E. coli* proteins have provided significant amount of information about the mechanism by which Thz-P is biosynthesised and were discussed in detail in section 1.3[27,40,41,43,51,55].

In *E. coli*, the imine intermediate (**32**, dehydroglycine) is proposed to be a product of tyrosine cleavage reaction catalysed by the enzyme ThiH. Genetic and biochemical studies have provided evidence of the presence of an oxygen sensitive iron sulfur cluster in ThiH[24,56,57], this cluster is proposed to reduce S-adenosyl methionine (AdoMet) to form a highly reactive 5'-deoxyadenosyl radical (Ado$^{\bullet}$) that will generate a tyrosyl radical that will further produce the glycine imine[51]. However, the role of the iron sulfur cluster in Thz-P biosynthesis is still unclear as attempts to reconstitute *in vitro* enzymatic activity were unsuccessful[58].

'Radical AdoMet' superfamily

ThiH is a 45 kDa protein that has been classified as a member of the 'radical AdoMet' superfamily of proteins[52,59]. The common feature of radical AdoMet family is the presence of a highly conserved CXXXCXXC motif (table 1.4). Over 600 proteins have been identified as potential 'radical AdoMet' members.

Lysine 2,3-aminomutase (KAM)[60,61], pyruvate formate lyase activase (PFL-AE)[62], anaerobic ribonucleotide reductase (ARR-AE)[63], biotin synthase (BioB)[64,65] and lipoyl synthase (LipA)[65,66] were the first enzymes to be classified as members of the 'radical AdoMet' superfamily (figure 1.3).

Table 1.4 Alignment E. coli ThiH with sequences from biotin synthase (BioB), lipoic acid synthase (LipA), anaerobic ribonucleotide reductase activating enzyme (ARR-AE), pyruvate formate lyase activating enzyme (PFL-AE) from E coli and lysine-2,3-aminomutase (LAM) from Bacillus halodurans.

Protein	Residues	Sequence
ThiH	83-94	N L **C** A N D **C** T Y **C** G F
BioB	51-62	G A **C** P E D **C** K Y **C** P Q
LipA	92-103	A I **C** T R R **C** P F **C** D V
ARR-AE	24-35	S G **C** V H E **C** P G **C** Y N
PFL-AE	28-39	Q G **C** L M R **C** L Y **C** H N
KAM	131-71	N Q **C** S M Y **C** R Y **C** T R

'Radical AdoMet' enzymes catalyse a single electron reduction of AdoMet and subsequent homolysis to produce a highly reactive 5'-deoxyadenosyl radical (Ado•) and methionine. The reduction of AdoMet occurs via a reduced $[4Fe\text{-}4S]^{1+}$ cluster, present in all AdoMet family enzymes, generating Ado• which then abstracts an hydrogen from a substrate (scheme 1.12). This iron sulfur cluster is therefore essential for enzyme activity[66-69].

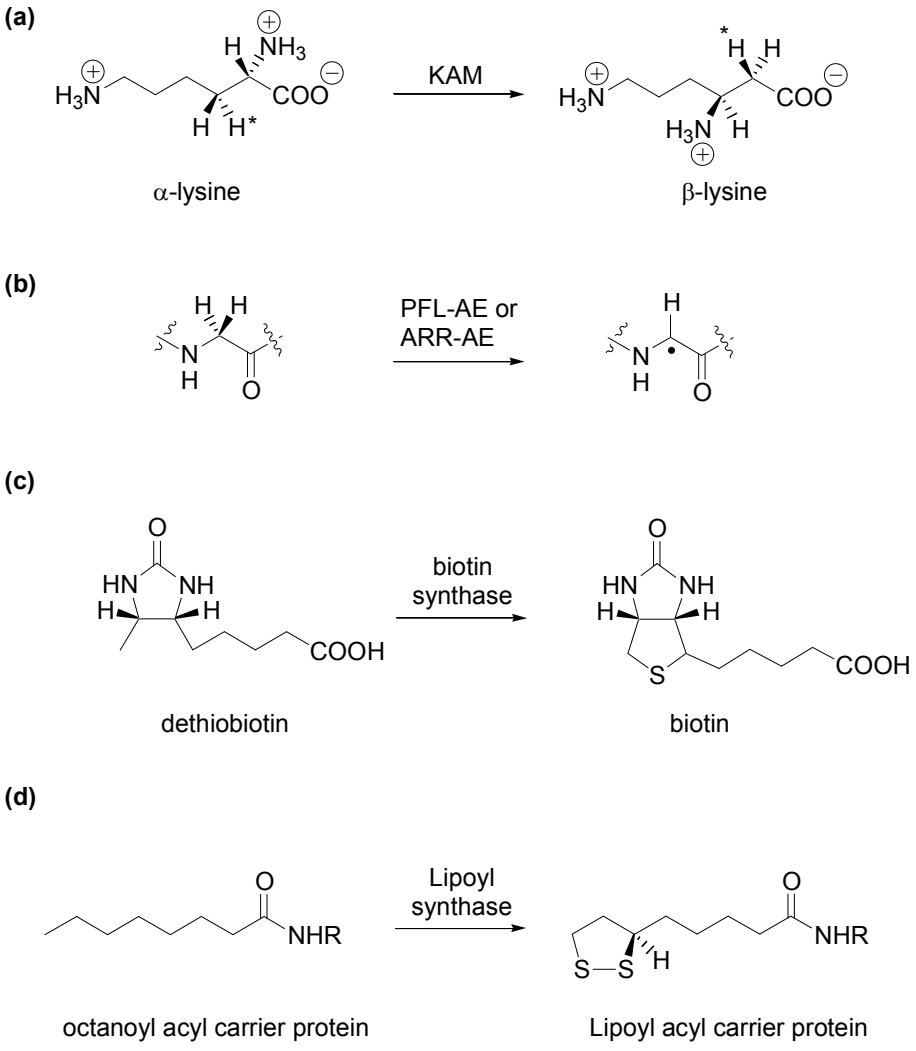
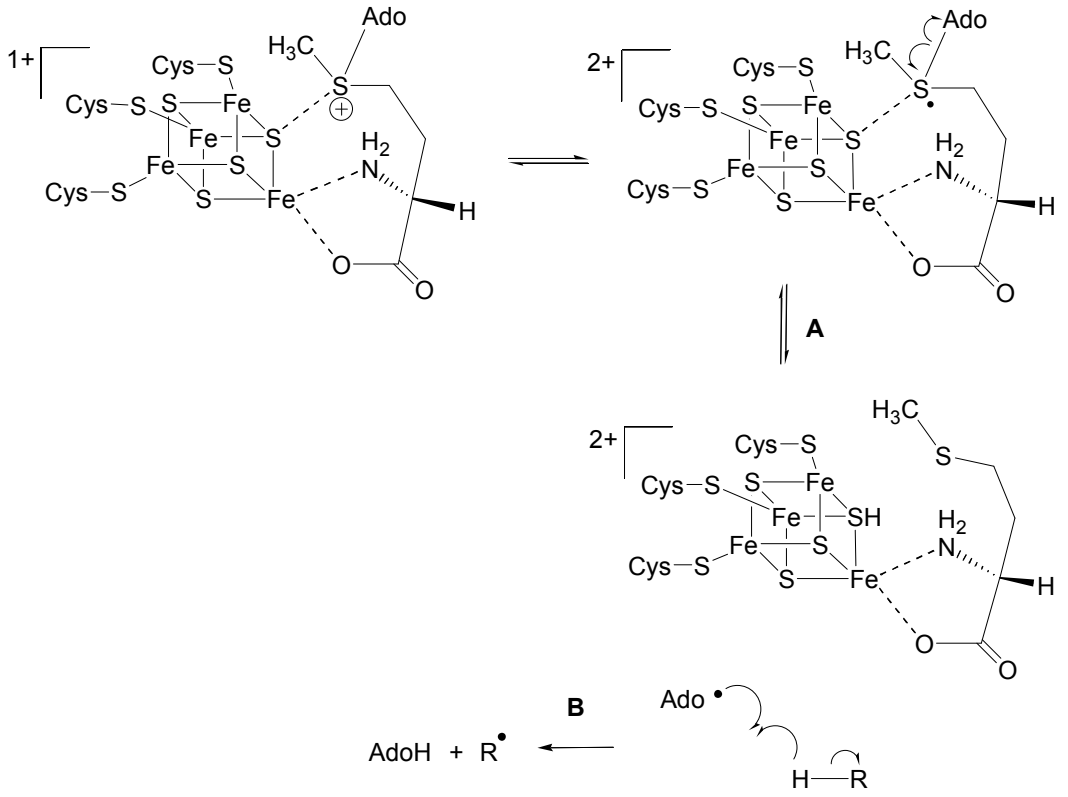


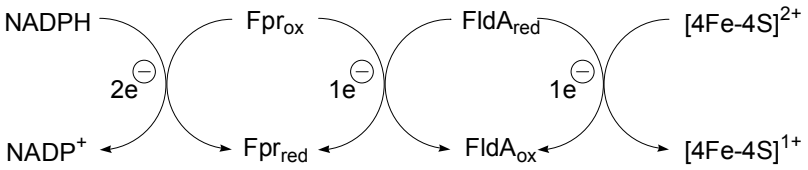
Figure 1.3 Examples of reactions catalysed by 'radical AdoMet' enzymes. (a) lysine 2,3-aminomutase (KAM); (b) general representation of the generation of glycyl radicals by pyruvate formate lyase activating enzyme (PFL-AE) and anaerobic ribonucleotide reductase activating enzyme (ARR-AE); (d) biotin synthase (BioB); lipoyl synthase (LipA). (*) marks the conserved hydrogen. Adapted from[70].



Scheme 1.12 Mechanism of radical AdoMet enzymes. The reaction is initiated by the reduction of [4Fe-4S] cluster that transfers an electron to AdoMet, forming the 5'-Ado radical and methionine. Steps A and B are probably concerted, and it is unlikely that significant amount of Ado• accumulates in the active site. Ado is a highly reactive primary radical that is assumed to undergo further reactions with substrates, usually through hydrogen abstraction to form a substrate radical.

Step 1: Reduction of the [4Fe-4S]$^{2+}$ cluster

The first step of AdoMet mediated reactions is the reduction of the [4Fe-4S]$^{2+}$ cluster to the +1 state. *In vitro*, dithionite[66] and 5-deazaflavin[71] can be used as reducing agents. The physiological reducing system [NADPH, flavodoxin (FldA), flavodoxin reductase (Fpr)] (scheme 1.13)[72-74] is commonly used and has proved to be more efficient for reducing PLF-AE, ARR-AE, BioB, LipA and MiaB. In other cases the reductant system is still unknown.



Scheme 1.13 Reduction of the iron sulfur cluster using NADPH, flavodoxin, flavodoxin reductase system. The number of electrons transferred is indicated at each stage.

Step 2: Formation of Ado•

Due to the strength of the S-adenosyl bond (~60 kcal mol^{-1}), AdoMet cannot spontaneously cleave homolytically to produce Ado• and methionine. Therefore, one of the important features of radical AdoMet enzymes is the presence of an iron sulfur cluster, allowing them to initiate radical chemistry, in which this unusual reaction takes place. The radical Ado radical has never been experimentally observed. However, electron paramagnetic resonance (EPR) spectroscopic studies using KAM and anhydroAdoMet, an AdoMet analogue, generate a stable, and therefore detectable, anhydroadenosyl radical[75]. Substrate carbon radicals have also been characterised by EPR with some enzymes: in KAM the substrate radical at C-3 has been detected using unsaturated analogues of lysine (trans-4, 5-dehydro-L-lysine) and 4-thia-L-lysine[76]; the glycyl radical of PLF[77], ARR[78], and benzylsuccinate synthase[79] as well as the substrate benzylic radical in HemN[80], have also been detected by EPR.

The [4Fe-4S]$^{+}$ clusters of KAM and PFL-AE have been shown to interact directly with AdoMet[81,82], and in the case of PFL-AE, the interaction occurs through the binding of this cofactor at the non-cysteinyl-ligated unique Fe site[81] (in these enzymes the cluster is chelated to three cysteines of the CX_3CX_2C triad) (scheme 1.12). This direct interaction may provide an easier route to the reduction and homolytic scission of AdoMet.

Step 3: Hydrogen abstraction by Ado•

The Ado• formed upon AdoMet cleavage will abstract hydrogen atoms, either from substrate molecules to form radical intermediates, or from glycyl residues of enzymes to activate them for further radical chemistry[83]. The 'radical AdoMet' enzymes can be sub categorised into two main subfamilies.

Subfamily 1

Reactions via a glycyl radical intermediate

These enzymes require an activase to bind to AdoMet. Enzymes that have been classified as members of this subfamily include: PLF-AE, ARR-AE, Benzylsuccinate synthase activase, glycerol dehydratase activase and 4-hydroxyphenylacetate decarboxylase activase. In this reaction, AdoMet acts as a co-substrate for the production of a glycyl radical intermediate generated on a highly conserved glycine residue. The consumption of AdoMet is stoichiometric with respect to the enzyme. The glycyl radical is transferred to the cysteine and the thiyl radical reacts with the substrate. This step is catalytic, as the reaction leads to a weak S-H bond that will donate a H• back to the product radical[70].

Subfamily 2

Reactions via a direct hydrogen transfer from the substrate to Ado•

Enzymes that have been classified as members of this subfamily include KAM and spore photoproduct lyase and the sulfur transfer enzymes BioB and LipA.

Members of this subfamily do not require a protein derived radical intermediate. The hydrogen is transferred directly between Ado• and the substrate. This reaction is reversible and the deoxyadenosine produced in the C-H bond cleavage returns a hydrogen to the rearranged product radical. AdoMet behaves as a cofactor in this reaction. However, in most cases studied AdoMet is a co-substrate, with one molecule being consumed per cleavage of C-H bond. For recently discovered enzymes such as

MoaA, ThiH and formyl glycine synthase, there is not enough experimental data available to classify them as members of one of the two subfamilies[70].

Within this subfamily there is a subclass of enzymes that catalyses a specific reaction, the formation of C-S bonds at non-activated carbons. So far only BioB, LipA and MiaB have been identified to be part of this family.

Biotin Synthase (BioB)

Biotin synthase is responsible for the conversion of dethiobiotin **48** (DTB) into biotin **52** via the insertion of a sulfur atom into an unreactive C-H bond of DTB. A mechanism has been proposed, although the detailed mechanistic enzymology by which this reaction occurs is still controversial (scheme 1.14)[84,85].

Jarrett and co-workers proposed that the active site of BioB contains two iron sulfur clusters. A $[4Fe\text{-}4S]^{2+/1+}$, characteristic of all 'radical AdoMet' enzymes, responsible for the reduction of AdoMet, and the second cluster, a $[2Fe\text{-}2S]^{2+}$, located at a different site[86-89]. EPR, Mossbauer and Raman spectroscopic studies and the recently solved crystal structure of BioB (figure 1.4), support this hypothesis[90].

The iron sulfur cluster present in BioB is oxygen sensitive, therefore isolated protein samples contain mainly [2Fe-2S], and addition of exogenous Fe^{2+} and S^{2-} is required to reconstitute biotin synthase activity *in vitro*. When chemically reconstituted, protein samples contain a 1:1 mixture of the two clusters ($[4Fe\text{-}4S]^{2+/1+}$, $[2Fe\text{-}2S]^{2+}$), as determined by Mossbauer spectroscopy[87]. It was also verified that the maximum turnover for this reaction was one. Figure 1.4 shows the 3D structure of the protein (a) and the section where AdoMet binds the iron sulfur cluster[90].

(a)

(b)

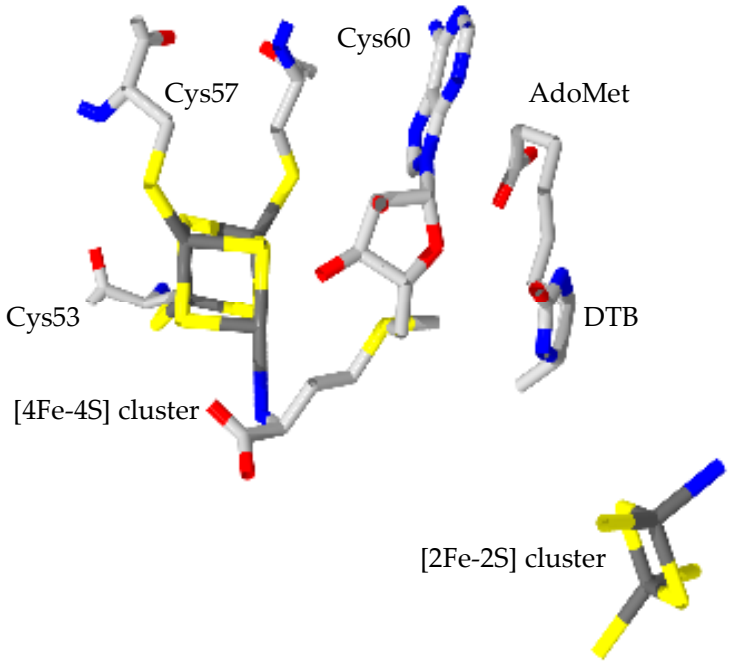
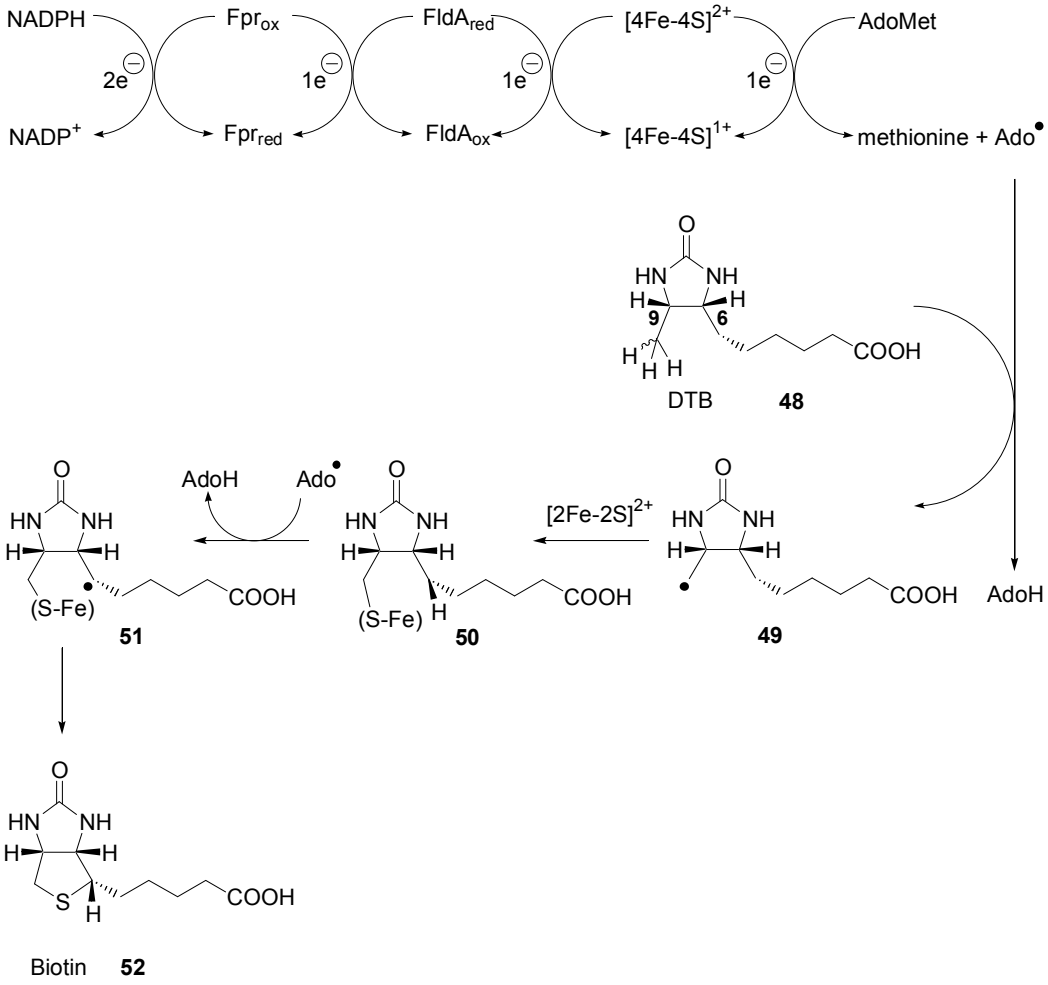


Figure 1.4 (a) Crystal structure of biotin synthase monomer (BioB) from E. coli; (b) amplification showing the AdoMet binding to the [4Fe-4S]$^{+1}$ cluster to the unique Fe site of BioB. Dethiobition (DTB) and the [2Fe-2S]$^{2+}$ cluster are also shown. Image created using POV-Ray version 3.6 and Swiss-Pdb Viewer Deep View version 4.0[90].



Scheme 1.14 Biotin synthase mechanism. The number of electrons transferred is indicated at each stage. Adapted from[42].

Studies have suggested that the sulfur source is the [2Fe-2S] cluster present in BioB[87-89]. Mossbauer studies have shown that during the formation of biotin, the amount of [2Fe-2S]$^{2+}$ cluster decreases, whereas the amount of [4Fe-4S] cluster remained constant[87,91]. However, studies suggest that the source of sulfur in this reaction derives from cysteine of PLP, not from the [2Fe-2S] cluster[92].

The formation of radicals at C-6 and C-9 of DTB has been further clarified by Lotierzo and co-workers[93]. Isotopic labelling studies demonstrated that biotin synthase catalyses the direct transfer of H• between Ado• and the substrate. However, the reaction of Ado

with the substrate remains to be clarified as isotopic studies proved to be unsuccessful[93].

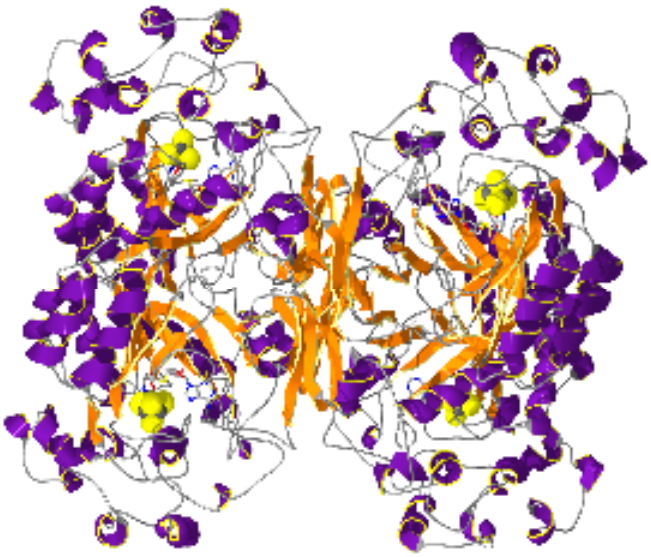
Lysine 2, 3-aminomutase (KAM)

Lysine 2, 3-aminomutase (KAM) is the most thoroughly studied radical AdoMet enzyme. KAM catalyses the interconversion between L-lysine and β-lysine, by an unusual reaction in which a hydrogen of C-3 exchanges position with the aminogroup on C-2[42]. Figure 1.5 shows the 3D structure of the protein (a) and the section where AdoMet binds the iron sulfur cluster[94].

There are many mechanistic similarities between this rearrangement and of the reactions catalysed by adesosylcobalamin(AdoCbl)-dependent enzymes[95], which is unsurprising as both AdoCbl and AdoMet act as sources of Ado radicals.

In the interconversion of lysine and β-lysine, the 3-*proR* hydrogen migrates to the 2-*proR* position of β-lysine with inversion of configuration, AdoMet is used as a cofactor given that the redox state of the substrate remains unchanged. Besides AdoMet this enzyme also requires pyridoxal phosphate (PLP) as a cofactor[96,97]. Studies using tritium labelled AdoMet showed that AdoMet is the intermediate carrier of hydrogen in KAM[98,99], and EPR experiments using isotopically labelled substrates, and substrate analogs of lysine have allowed various radical species that are proposed to be intermediates in the mechanism to be identified[100,101].

(a)



(b)

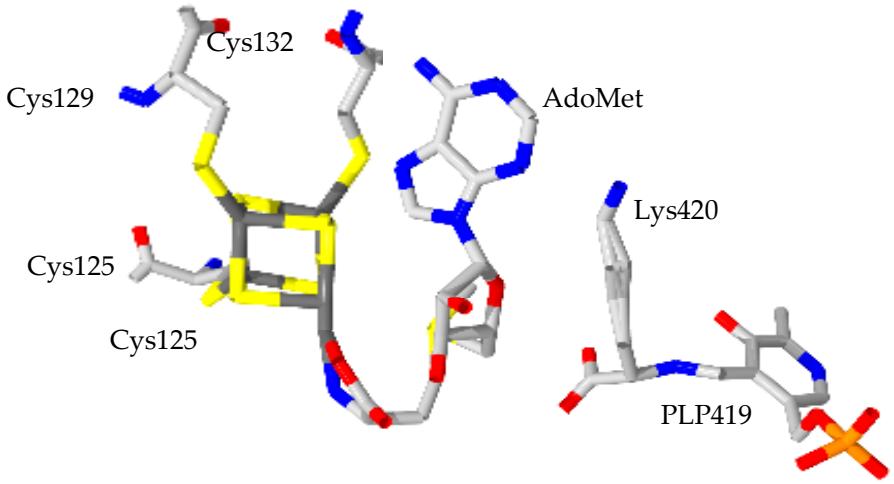
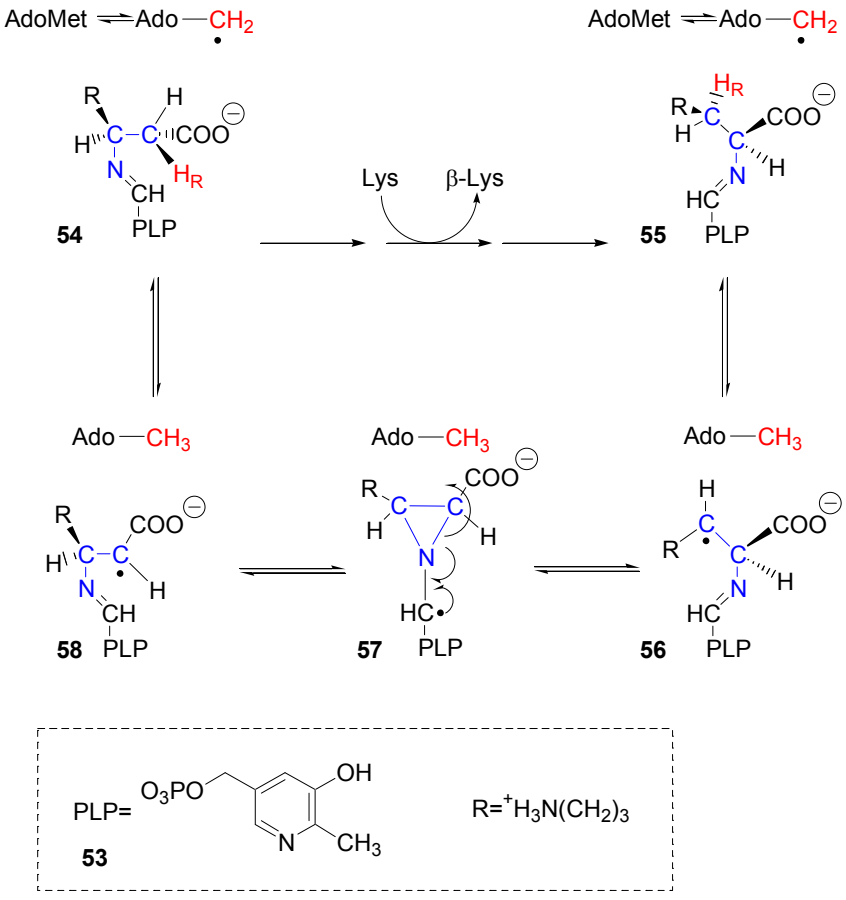


Figure 1.5 (a) Crystal structure of lysine 2,3-aminomutase (KAM) from C. subterminale; (b) amplification showing the AdoMet binding to the [4Fe-4S]$^{+1}$ cluster to the unique Fe site of KAM and lysine residue binding to pyridoxal-5′-phosphate *(PLP). Image created using POV-Ray version 3.6 and Swiss-Pdb Viewer Deep View version 4.0*[94].

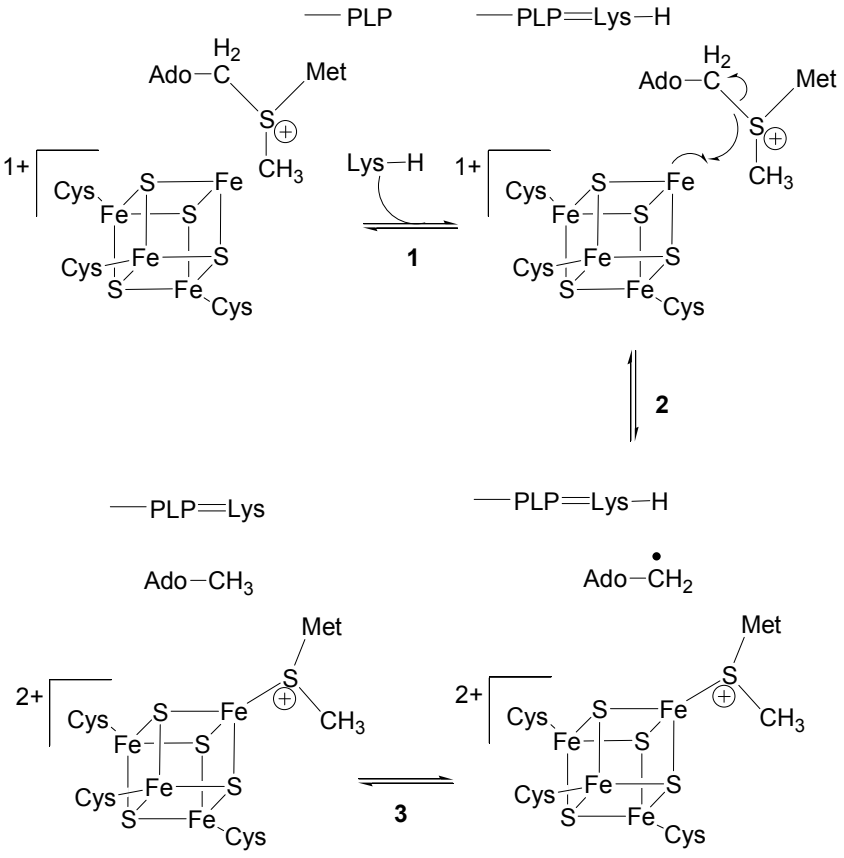
The mechanism of this reaction is outlined in scheme 1.16. Lysine binds to KAM as its external aldimine with PLP. The reversibly generated Ado•, from AdoMet in a reaction with the [4Fe-4S] cluster, abstracts a hydrogen atom from lysyl at C-3 to form AdoH and the lysyl radical **56**. Cyclisation by radical addition to the π-bond of the imine group produces the azacyclocarbinyl radical **57**, a quasisymmetric intermediate. Forward ring opening of the azacyclopropylcarbinyl ring produces the β-lysyl radical **58**, which abstracts a hydrogen atom from methyl group of AdoH to regenerate Ado• and form the external aldimine between β-lysine and PLP **53**. Release of β-lysine and binding of lysine in several steps recharges the enzyme for another cycle of catalysis[102].



Scheme 1.16 Radical rearrangement catalysed by lysine 2, 3-aminomutase (KAM). Adapted from[103].

In KAM, the mechanism by which AdoMet is cleaved by the iron sulfur cluster to generate Ado•, has been the subject of great interest. EPR spectroscopy studies have shown that KAM contains four forms of iron sulfur cluster: [3Fe-4S], $[4Fe\text{-}4S]^{3+}$, $[4Fe\text{-}4S]^{2+}$ and $[4Fe\text{-}4S]^{1+}$ [69,104]. Two of these oxidation states are important for the reversible cleavage of AdoMet, $[4Fe\text{-}4S]^{2+}$ and $[4Fe\text{-}4S]^{1+}$.

Several hypothetical mechanisms have been proposed for the generation of the radical that occurs upon the reaction of AdoMet with the iron sulfur centre [83]. One of these mechanisms involves the direct interaction between the unligated iron of the iron sulfur cluster and the sulfonium sulfur of AdoMet. This is associated with an inner sphere electron transfer that occurs simultaneously with the homolytic cleavage of C-S bond of AdoMet. This process leads to direct coordination of the sulfur of the methionine with the iron, after generation of Ado• (scheme 1.17)[83].



Scheme 1.17 Proposed mechanism for the reversible cleavage of AdoMet by the [4Fe-4S] cluster to generate Ado•. Adapted from[83].

An electron is transferred from [4Fe-4S]+ to AdoMet and the C-S bond cleaves homolytically, leaving methionine and generating the 5'deoxyadenosyl radical (Ado•). This reactive radical will then abstract an hydrogen atom (3-pro-*R*) from the external aldimine of lysine with PLP. This mechanism is supported by X-ray absorption spectroscopic studies (XAS) using Se-adenosyl-L-methionine (SeAdoMet)[105].

Comparison between the binding of AdoMet to the [4Fe-4S] cluster in KAM and BioB

Electro paramagnetic resonance (EPR) experiments with KAM showed that the rates of formation of the organic substrate radical and the rate of decomposition of the [4Fe-4S]$^{1+}$ cluster signal were the same, as were the rates of formation of 5'-deoxyadenosine and methionine [76]. The unligated iron is crucial for the interaction between the cluster and AdoMet, and EPR experiments demonstrated that addition of AdoMet (in absence of a substrate) to the protein changes the spectral properties of the cluster[97,106-108]. Recent spectroscopic studies resulted in two detailed proposals for the interaction of AdoMet and the Fe-S cluster, resulting in two different proposed mechanisms for AdoMet cleavage (scheme 1.18).

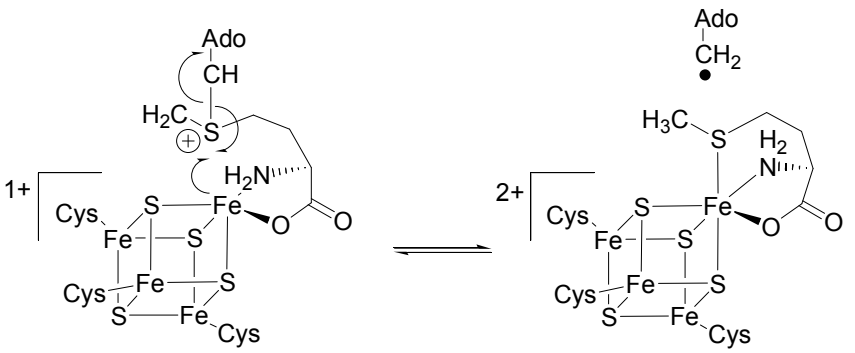
Studies of KAM with selenium X-ray absorption spectroscopy indicate the direct ligation of the sulfonium sulfur of AdoMet to the unligated iron [105]. This study was complemented by experiments using electron double nuclear resonance that verified that the α-amino and α-carboxyl groups of AdoMet coordinate to the unligated iron of the cluster. Therefore, the proposed mechanism, for KAM, entails an electron transfer from the iron sulfur cluster to the sulfonium of AdoMet, consequent ligation of the iron to the sulfur of AdoMet and simultaneous cleavage of the S-C bond to form 5'-adenosyl radical (Ado•) resulting in a methionine coordinating the iron atom of the cluster[97,109].

Studies with BioB and PFL, using EPR and Mossbauer spectroscopy, suggest that the sulfonium of AdoMet is more closely associated with one of the sulfides rather then the unligated iron of the [4Fe-4S]$^{1+}$ [81,110-112]. The proposed mechanism starts with an electron being transferred from the [4Fe-4S]$^{+1}$ cluster to AdoMet via a sulfide-

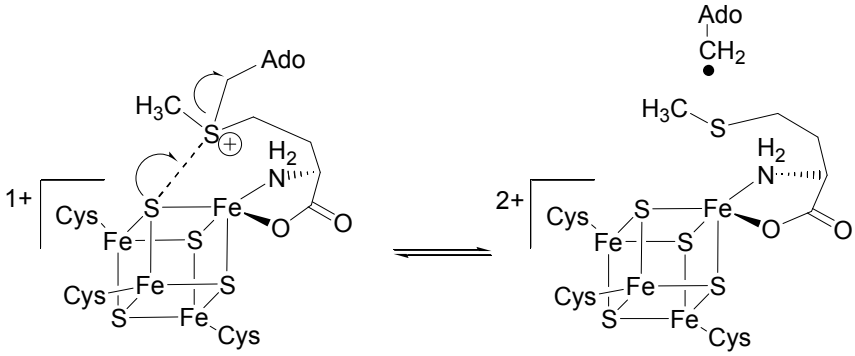
sulfonium interaction. The sulfonium-adenosyl bond from AdoMet cleaves homolytically to generate Ado•. In this proposed mechanism, methionine can only coordinate to the iron sulfur cluster through the amino and carboxyl groups[112].

These two different proposed mechanisms for AdoMet binding may suggest slightly different functions for AdoMet in each reaction. Whilst in KAM, AdoMet is regenerated after each turnover, in BioB and PFL-AE AdoMet acts as a co-substrate and is consumed after each turnover. Coordination between the iron ion of the cluster and the sulfur of the methionine may have significant influence on the reactivity of the sulfur and therefore can alter reformation of AdoMet in the reaction[109].

(a)



(b)

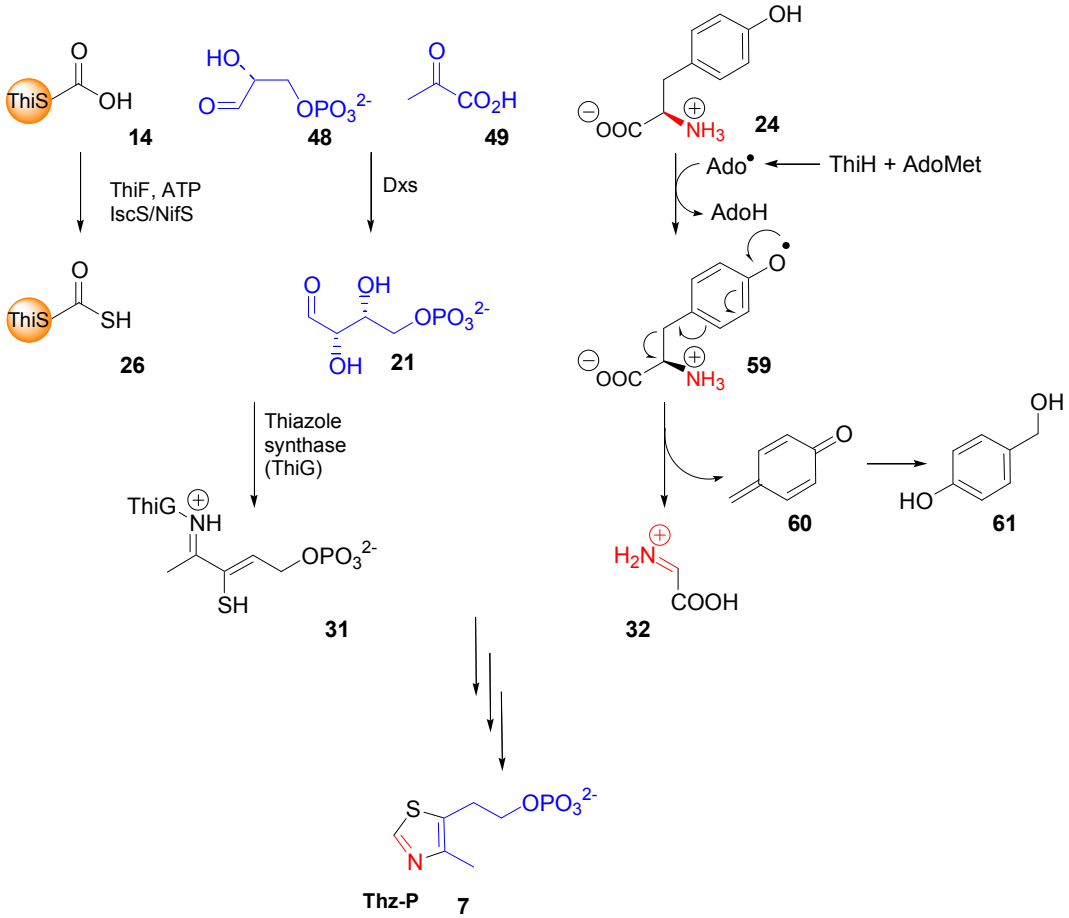


Scheme 1.18 Binding of AdoMet to the [4Fe-4S]$^{+1}$ cluster in: (a) KAM and (b) BioB. Adapted from[95].

Thiamine biosynthetic protein ThiH

As mentioned previously, ThiH is a member of the 'radical AdoMet' superfamily. As far as the classification of this protein within the radical AdoMet, ThiH is still part of an unknown subfamily because the mechanism by which the iron sulfur cluster (of ThiH) interacts with AdoMet to generate Ado and cleave tyrosine is still unknown. Thiazole synthase catalyses an AdoMet and NADPH dependent radical reaction, being therefore a novel example of the diverse chemistry catalysed by 'radical AdoMet' superfamily of enzymes.

The proposed mechanism for Thz-P biosynthesis by Leonardi[51], together with the new discoveries achieved by Begley and co-workers on *B. subtilis*, leads to an overall proposal for the biosynthesis of the thiazole moiety in *E. coli* (scheme 1.19).



Scheme 1.19 Proposed mechanism for Thz-P biosynthesis in E. coli. Adapted from[43,58].

The [4Fe-4S]$^{1+}$ cluster of ThiH catalyses the reductive cleavage of AdoMet to generate Ado•. This highly reactive radical will then abstract an hydrogen atom from tyrosine **24** and forms a tyrosyl radical **59** that will react cleaving the Cα-Cβ bond. Glycine **32** is released together with the quinone methide **60**. Hydration of **60** leads to *p*-hydroxybenzyl alcohol **61**, a known thiamine by-product[113].

1.6 Aims of this project

The main aim of this project is to understand the mechanistic enzymology by which the thiazole moiety of thiamine is assembled by *E. coli*. This initially requires the expression, isolation and characterisation of the enzymes involved in these reactions and subsequent reconstitution of the *in vitro* activity. The properties of ThiH and ThiG have been reported; in particular of ThiGH complex. ThiH has been classified as a member of the 'radical AdoMet' superfamily. All the members of this family have been shown to be Fe-S cluster enzymes requiring *S*-adenosyl methionine (AdoMet) for activity, and recent studies showed that the [4Fe-4S] cluster of ThiH is essential for the catalytic activity. However, the complex isolated has a deficiency in iron content; therefore novel expression and purification techniques need to be explored with the purpose of producing *holo*-ThiH, and consequently, increase the activity that will facilitate the development of an *in vitro* assay to investigate thiazole biosynthesis and in particular, the mechanism of tyrosine cleavage.

CHAPTER 2

Isolation and Reconstitution of *Escherichia coli* ThiH

2.1 Introduction

The investigation of the mechanism by which the thiamine precursor 4-methyl-5-(β-hydroxyethyl) thiazole phosphate (Thz-P) is assembled required the isolation of enzymes involved in an active, stable form. Significant progress has been made in elucidating the function of several of the enzymes involved in bacterial thiazole biosynthesis as discussed in Chapter 1. However, limited progress had been made in understanding the function and mechanism of ThiH, a protein that shows considerable sequence similarity to the 'radical AdoMet' family. Initial studies by R. Leonardi[58] established methods for the expression, isolation and characterisation of ThiH. This enzyme was isolated in a complex with ThiG in an approximate 1:1 ratio. However the iron content of the purified protein complex was much less then expected containing around 1Fe (0.96 ± 0.03 Fe/ThiH) and 2S (1.87 ± 0.09 S^{2-}/ThiH) atoms per ThiGH[58], as opposed to 4 moles of Fe/protein expected in a [4Fe-4S] cluster protein. The isolation of iron sulfur cluster proteins containing less iron then expected is not uncommon[114-116] as iron may be lost, mainly by oxidative degradation, either during the expression or purification procedure [114-116]. For proteins of the 'radical-AdoMet' family, in which the cluster rests in the $[4Fe\text{-}4S]^{2+}$ state, oxidation can convert it to a metastable form. Further exposure to oxidising conditions leads to additional degradation of the cluster which may result in a [2Fe-2S] or even an *apo*-form[117]. The [4Fe-4S] cluster of ThiH is thought to be essential for its activity and therefore oxidative degradation results in inactivation of the protein[24]. To avoid this problem, R. Leonardi[58] developed an anaerobic purification protocol (method 12). ThiGH complex, purified using this method, was tested for activity in an assay containing cell free lysates of de-repressed *E. coli* 83-1, tyrosine, 4-amino-5-hydroxymethyl-2-methylpyrimidine (Hmp) and

adenosine triphosphate (ATP). The protein was shown to be active[51]. A hypothesis for this observation is that the cellular machinery required for cluster (re-) assembly may have been present in the cell-free lysates and facilitates reactivation of the ThiH in the *holo-* form. It was further observed that the addition of AdoMet, ferric chloride, sodium sulfide and NADPH doubled the relative thiamine forming activity in these assays. The effect of AdoMet on the activity is not perhaps very surprising, and is consistent with enzymes from 'radical AdoMet' family, but the effect of ferric chloride, sodium sulfide and NADPH may suggest that chemical reconstitution of ThiH to an active form is occurring in the assay.

With the purpose of expressing ThiGH, several plasmids were assembled by R. Leonardi (figure 2.1). These plasmids were derived from either a pBAD-HisA (araBAD promoter) or pET-24d(+) (T7 promoter) parental plasmid. Expression and purification studies showed that the nature of the expression system had significant influence not only on the production of soluble ThiGH complex but also on the iron and sulfide content of the samples (table 2.1)[58]. It was observed that pET-derived plasmids produced poor yield of cell paste compared to the plasmids derived from pBAD-HisA, possibly due to the fact that production of protein was too fast for the iron sulfur cluster biosynthesis machinery to support; ultimately the protein becomes toxic to the cells, resulting in the formation of inclusion bodies[58]. Plasmids expressing ThiH with ThiG, ThiF and ThiS resulted in a higher level of soluble ThiGH and neither ThiF nor ThiS were detected in the purified samples. From the assembled plasmids, pRL1000 and pRL1020 produced ThiGH samples containing higher content of iron and sulfide comparing to other expression plasmids (table 2.1)[58].

As previously mentioned the partial degradation or incomplete assembly of the iron sulfur cluster might explain the lower content in iron of the samples. With the purpose of investigating ways to assist *in vivo* formation of the iron sulfur cluster of ThiH, R. Leonardi assembled the pRL1021 plasmid (figure 2.1), encoding for *iscSUA-hscAB-fdx* downstream from the *thiFSGH* operon. Proteins encoded by the *iscSUA-hscAB-fdx* operon assist Fe-S cluster assembly *in vivo*, a key step in the post-translational maturation of Fe-S cluster proteins[118,119], therefore improving expression of enzymes in

the *holo-* form[120-122]. Table 2.2 and figure 2.2 summarise the functions of each of these proteins.

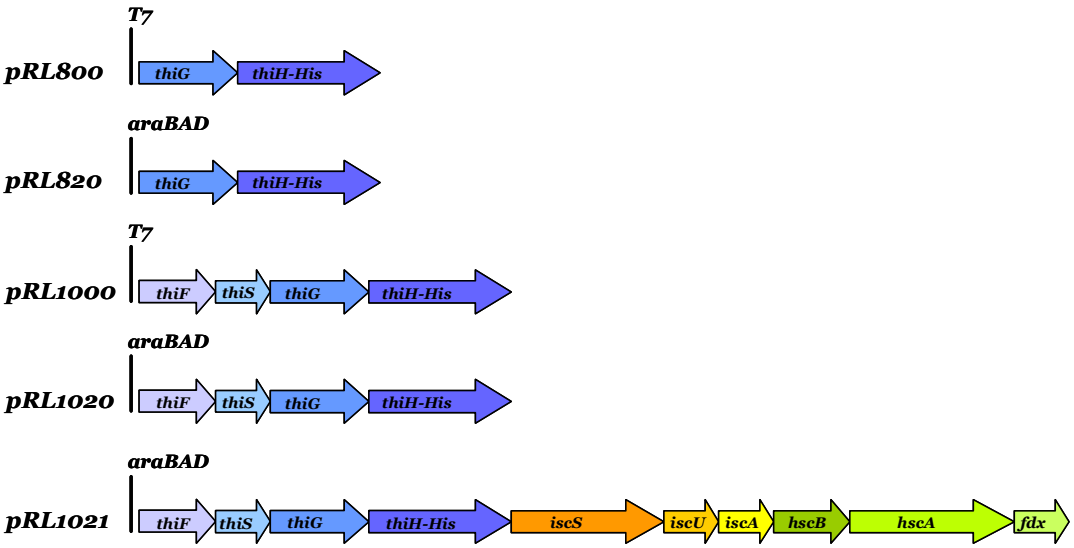


Figure 2.1 Artificial operons assembled in plasmids derived from either pBAD-HisA (araBAD promoter) or pET-24d(+) (T7 promoter) used to study ThiH expression.

Table 2.1 Sulfide and metal content of ThiGH-His samples isolated from different expression systems. The sulfide and iron values represent the mean of three independent measurements ± SD. Adapted from[58].

ThiGH-His from [vector/BL21(DE3)]	Sulfide	Iron
	mol/mol ThiGH-His	
pRL800	0.69 ± 0.02	0.45 ± 0.05
pRL820	2.31 ± 0.08	0.87 ± 0.02
pRL1000	2.32 ± 0.06	0.91 ± 0.03
pRL1020	1.87 ± 0.09	0.96 ± 0.03
pRL1021	2.4 ± 0.1	1.36 ± 0.03

Table 2.2 Proteins of the Isc cluster and their functions.

Gene	*Orthologs*	*Potential Functions*	*References*
iscS	*nifS, sufS*	*Cysteine desulfurase* Transfers the sulfur atom to IscU	123-127
iscU	*nifU, sufU*	*Scaffold protein* Assembly of iron sulfur clusters	128,129
iscA	*sufA*	*Scaffold protein* Assembly of iron sulfur clusters	130
iscR		*Iron sulfur cluster regulator* Repressor for the Isc operon	131
hscA/hscB		*Molecular chaperones* Facilitate [2Fe-2S] cluster transfer from IscU to apoferredoxin in ATP-dependent reactions	132-134
fdx		*Adrenodoxin-type ferredoxin* Electron donor in cluster assembly	135

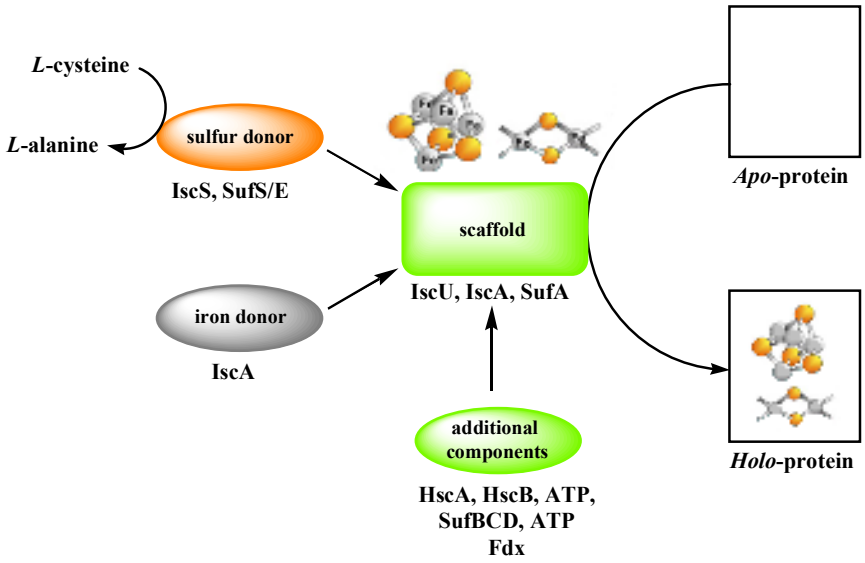


Figure 2.2 Biosynthesis of iron sulfur clusters in bacteria. IscA, IscU and SufA are the central components. They function as molecular scaffolds mobilising the sulfur atoms from IscS or SufS and iron atoms from an iron donor to synthesise [4Fe-4S] or [2Fe-2S] clusters. Adapted from[136].

Initial *in vitro* experiments to test the activity of purified ThiH, carried out by Dr. M. Kriek in our laboratory, showed limited or no activity [as estimated by the turnover of ^{14}C-tyrosine (results not shown)], although the reasons for the lack of turnover were unclear. To solve this problem, a wide range of parameters that might have affected the activity were investigated:

1. Purification methodology - several modifications were applied including: additions to the chromatography buffers to include protective reagents; modification of the chromatography media and methods used, including ion exchange methods; and modification of the affinity tag used during the purification.
2. *In vivo* assembly of *holo*-ThiH - modifications to the expression plasmid, effects of different *E. coli* strains and varying the growth conditions.
3. *In vitro* chemical reconstitution of purified ThiH into active *holo*-ThiH.

The studies described in this chapter were carried out in parallel to the development of an *in vitro* assay (discussed in Chapter 3).

2.2 Results and discussion

2.2.1 Improving cell lysis

The successful method for isolation of ThiGH, developed by Leonardi[24], involved an aerobic cell lysis step using sonication. However, the release of ThiH protein from the cells into an aerobic buffer might result in a rapid degradation of the oxygen sensitive iron sulfur cluster. Therefore, before attempting any other approach to solve the deficiency in iron content of ThiH, the effect of anaerobic cell lysis was investigated. This might produce purified protein samples with higher iron content and in turn result in a higher specific activity. Table 2.3 summarises the results obtained from both purification methods.

Table 2.3 Comparison of two cell lysis methods during ThiGH-His purification.

Cell Lysis	Yield of protein (mg/g cell)	Iron content mol/mol ThiGH
Aerobic	2.2	0.97 ± 0.03
Anaerobic	5.2	1 ± 0.05

ThiGH-His was expressed using pRL1020, a plasmid containing the *thiFSGH* operon under the control of the arabinose promoter, with a C-terminal His-tag attached to *thiH*. Purification of ThiGH was carried out using nickel affinity chromatography followed by a desalting gel filtration step, using aerobic cell lysis (method 12) [24]. A second experiment was carried out using the same protocol described above but cell lysis was carried out anaerobically by using a sonicator placed inside the anaerobic glove box (method 12a). The iron content of the purified protein samples was measured by the method of Fish (method 14)[137]. The new modified method (method 12a), produced more efficient cell lysis (judged by the increase of the yield of protein) but also, on average, no increase of the iron content of ThiGH samples was observed. The elution of the ThiGH complex was identical in both methods (figure 2.3).

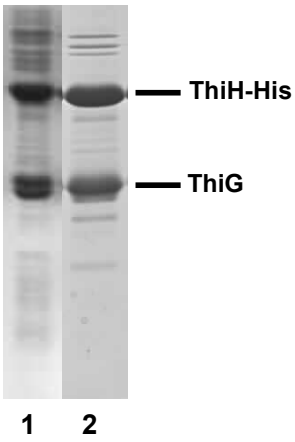


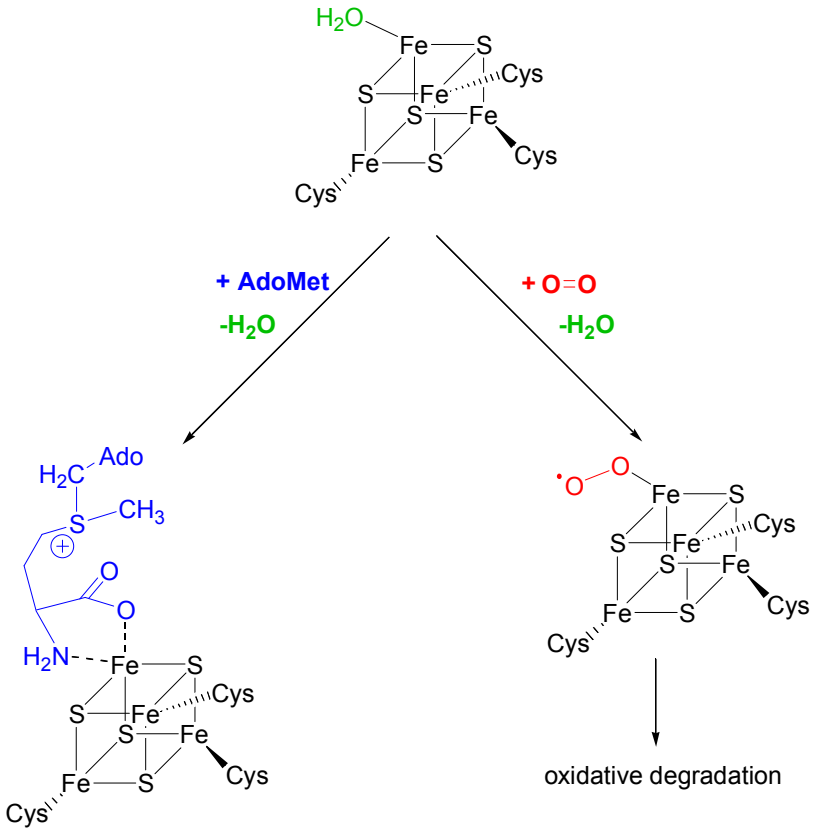
Figure 2.3 Coomassie Blue stained 15% SDS-PAGE gel of ThiGH-His purification expressed from pRL1020/BL21(DE3) cells in 2YT medium. Lane 1: aerobic sonication; lane 2: anaerobic sonication.

2.2.2 Improving *in vivo* assembly of *holo*-ThiH

Two aspects were considered for the improvement of *in vivo* assembly of *holo*-ThiH:

1. *E. coli* is a facultative anaerobe and is able to grow under aerobic or anaerobic conditions. In the absence of oxygen *E. coli* grows by anaerobic respiration with nitrate, nitrite, fumarate, dimethyl sulfoxide and trimethylamine N-oxide as electron acceptors or by fermentation[138,139].

ThiH production under anaerobic conditions is of interest as it may help to protect the Fe-S cluster from oxidative degradation. The [4Fe-4S] cluster of 'radical-AdoMet' proteins is known to be particularly oxygen sensitive because, in the resting state, it is bound only by three cysteine ligands (CXXXCXXC motif): the fourth coordinating site is thought to be occupied by a labile water molecule which may provide a particularly effective route for attack by dioxygen (scheme 2.1). However, anaerobic metabolism is rather inefficient; therefore growth and protein production under these conditions can be inefficient. The use of nitrate as the terminal electron acceptor, a process called denitrification, enhances growth whilst permitting the exclusion of dioxygen.



Scheme 2.1 Structure of the [4Fe-4S] cluster in the resting state. This cluster can either bind AdoMet or undergo oxidative degradation by dioxygen.

2. The addition of supplements to the growth medium may favour efficient assembly and/or protection of the essential iron sulfur cluster of ThiH. This included supplementing iron and other substrates to ensure efficient Fe-S cluster assembly.

Studies using purified *E. coli* dihydroxy-acid dehydratase, fumarase A, fumarase B and mammalian aconitase, all belonging to the hydrolyase class that contain a catalytically active [4Fe-4S] cluster, demonstrated that these enzymes are inactivated by the presence of dioxygen. In this work, Flint hypothesised that the presence of O_2, inactivates these enzymes by oxidizing their clusters to an unstable oxidation state, resulting in cluster degradation (scheme 2.1)[140]. This hypothesis was supported by the following findings: 1) a substrate protects against inactivation of the iron sulfur cluster enzyme, 2) iron is lost on inactivation as would be expected if the clusters were

oxidized, 3) enzymatic activity is recovered under conditions which lead to cluster re-synthesis[140]. Thus, the presence of a substrate in the growth medium or chromatography buffers might prevent the oxidation of ThiH [4Fe-4S] cluster. Potentially, the addition of tyrosine, AdoMet or iron, which might all bind to the active site of ThiGH, could protect the cluster during cell growth, or during purification, leading to an increase in the yield of *holo*-ThiGH.

With the aim of improving *in vivo* assembly of *holo*-ThiH, and two alternative expression systems available (pRL1020 encoding for *thiFSGH* and pRL1021 containing the additional *isc* operon); a number of experiments were carried out using modified growth conditions. The purification methods used were all derived from the initial purification procedure developed by Leonardi (method 12)[24]. All anaerobic growths were carried out at low temperatures in an attempt to prevent protein misfolding (overnight at 4 ºC).

Expression and purification experiments with pRL1020/BL21(DE3)

When modifying the cell culture and purification methods, the results are compared at each stage to the yield and iron content of ThiGH anaerobically isolated from aerobically cultured pRL1020/BL21(DE3).

Expression experiments using pRL1020/BL21(DE3) were carried out aerobic and anaerobically (table 2.4). Several parameters were considered when optimising the expression of ThiH, including the yield of cell paste from the growth, the yield of protein upon cell lysis and, after purification of ThiH, the iron content of the resultant protein. Anaerobic cell growths were carried out using glycerol minimal medium (GMM) and 2YT medium (with or without nitrate addition). A significant decrease of the yield of cells and protein content was observed (table 2.4, entries 1-4). In addition, no improvement was observed in the iron content of purified ThiGH. These results suggest that either anaerobic growth is not facilitating the assembly of *holo*-ThiH, or the *holo*-ThiH is being degraded during a subsequent step.

SDS-PAGE gel analysis of ThiGH-His purification using pRL1020/BL21(DE)3 expression system under aerobic conditions is shown as an example (figure 2.4). The numbers on the left refer to the molecular weight marker in kDa; ThiH-His (44 kDa), ThiG (27 kDa). This notation will be used throughout this thesis.

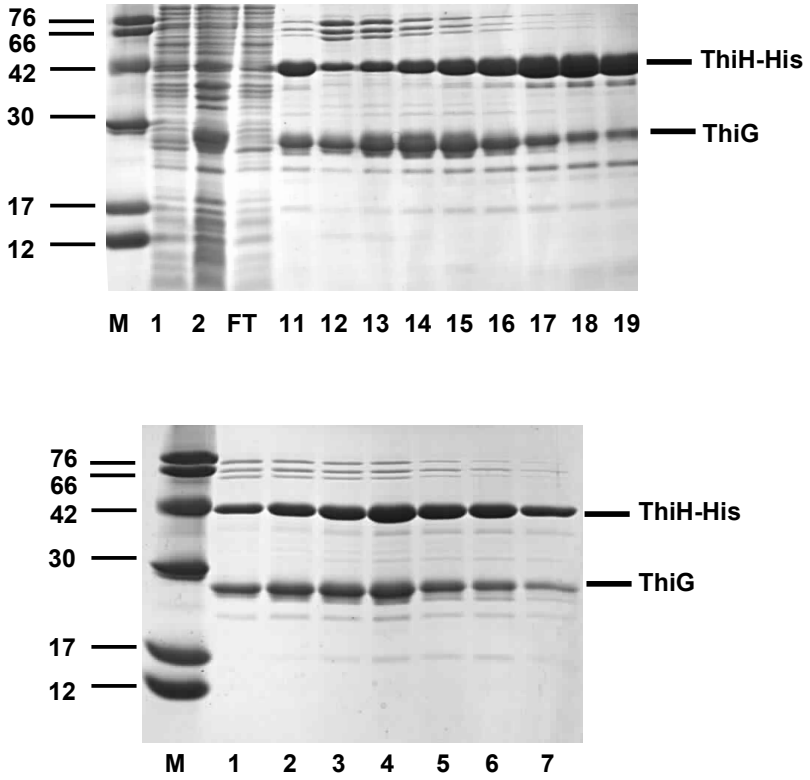


Figure 2.4 Coomassie Blue stained 15% SDS-PAGE gel of ThiGH-His purification from pRL1020/BL21(DE3) cells in 2YT medium. M= molecular weight marker (kDa). 1: supernatant; 2: re-suspended cell pellet; FT: flow through. The numbers correspond to the fractions eluted from each column; 11-19: nickel-chelating column, 1-7: S-75 gel filtration column (obtained by pooling fractions 11-16 from the nickel column and loading onto gel filtration column).

With the purpose of investigating the effect of nitrate addition in the production of *holo*-ThiH, investigations were initiated to determine the most suitable growth medium to use in pRL1020/BL21(DE3) expression under anaerobic conditions. Flint and co-workers reported that minimal medium was a good choice for *E. coli* growth under

anaerobic conditions[140]. A comparison of the total cell biomass produced using glycerol minimal medium (GMM) or 2YT growth using the same conditions was carried out.

A set of experiments were carried out in parallel (table 2.4). Sodium nitrate (40 mM, $NaNO_3$) and sodium nitrite (5 mM, $NaNO_2$) were added to 2YT and GMM medium. The yield of cells produced with 2YT growth was much larger than with GMM, but the yield of protein produced from GMM was higher then that from 2YT. The purified ThiGH-His samples from these growths were then analysed by the method of Fish to determine the iron content (method 14)[137]. Results show no improvement on the production of *holo*-ThiH using this method, indicating that the addition of nitrates was not assisting in the *in vivo* production of ThiH, therefore, this line of investigation was terminated.

Expression and purification experiments with pRL1021/BL21(DE3)

Studies on the effect of substrate addition were initiated using pRL1021/BL21(DE)3 expression system. Trial growths, under aerobic conditions, were carried out to assess the expression level of the proteins using GMM and 2YT medium (results not shown). Growths using 2YT medium resulted in a higher yield of protein, therefore all experiments were carried out using this growth medium.

Aerobic and anaerobic growth conditions were compared (table 2.4). Cells were grown initially aerobically in 3 L of 2YT supplemented with tyrosine (0.2 mM), ferrous sulfate ($FeSO_4$, 0.3 mM) and ampicillin (100 μg/mL), upon reaching an OD_{600}~0.6, protein expression was induced. After induction, the cells were transferred to gas tight bottles, grown anaerobically overnight and harvested anaerobically in gas tight centrifuge bottles. ThiGH-His was anaerobically purified by the optimised purification protocol (method 12a). During the elution of ThiH-His from the nickel-chelating column, two sets of equally concentrated fractions, named A and B and different in colour, were observed; these fractions were separately pooled and pool from batch B was desalted and concentrated. Batch A was intensely coloured and darker then batch B. SDS-PAGE

analysis revealed the presence of YqjI (23 kDa) protein in both batches, but slightly more abundant in batch A (figure 2.5).

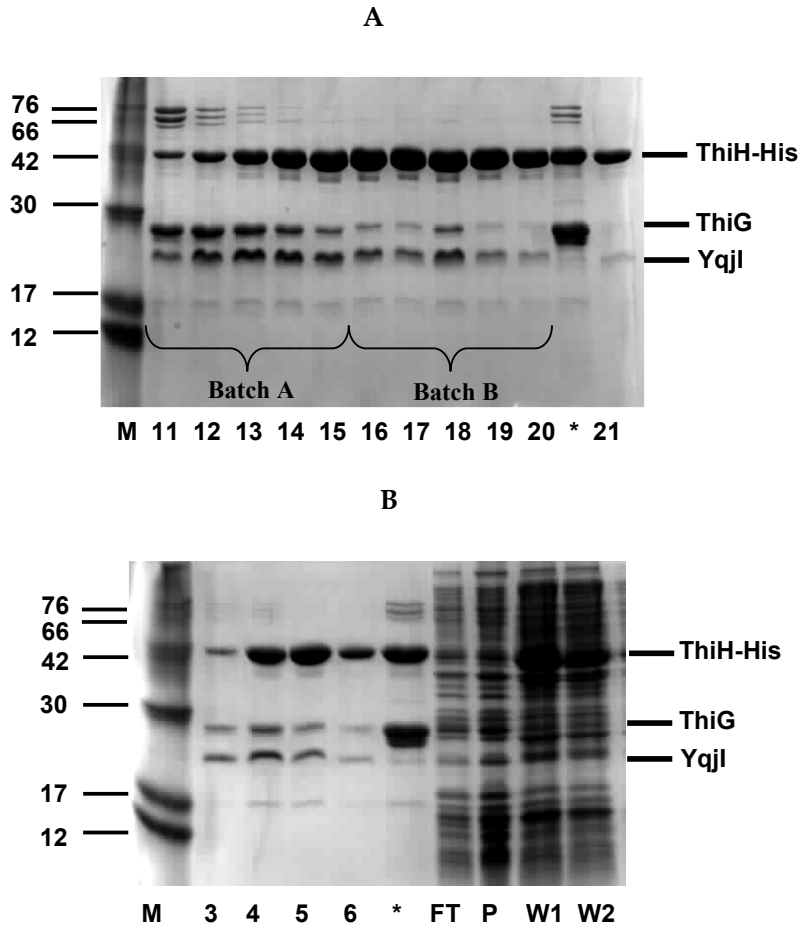


Figure 2.5 Purification of ThiGH-His from pRL1021/BL21(DE3) cells in anaerobic 2YT medium supplemented with iron and tyrosine. Coomassie Blue stained 15 % SDS-PAGE gels of proteins eluted from A: nickel-chelating column and B: gel filtration column (fractions 17-21, batch B); FT = flow through; W = washing; P = re-suspended pellet; M = molecular weight marker (kDa). The numbers correspond to fraction numbers.

R. Leonardi also reported the co-elution of this protein with ThiGH when isolated from pRL1021/BL21(DE3) expression system[58]. In these experiments however, no exogenous iron had been added to the growth medium, also, the growth was carried out aerobically. Leonardi further investigated this protein. N-terminal sequencing of the protein and a database search based on the knowledge of the first seven residues (SHHHEGL) identified a protein, YqjI containing eleven histidines in a sequence of forty residues, thus rationalising the high affinity for the nickel-chelating column. Although not characterised in detail, this protein shows a broad absorption maximum at 420 nm and may contain an iron cofactor resulting in the dark colour of the batch[58]. Therefore the additional iron may be associated with the presence of YqJI. The nature of this correlation is presently unclear as the function of the protein is thus far unknown. The iron content of isolated ThiGH was 2.6 Fe/ThiH from anaerobically cultured cells, supplemented with additional iron and tyrosine. Table 2.4 summarises the list of methods used as well as the results obtained from each experiment.

Studies on YqJI mutants have been continued by Mr. M. Challand. These experiments were designed to test whether *E. coli* cells with the genomic copy of YqJI deleted would provide a valuable system for production of *holo*-ThiH. However, the deletion of YqJI seems to adversely effect cell growth, as a typical 5L cell growth, using a –YqJI strain, produced only 13.3 g cell paste, and SDS PAGE analysis of protein expression suggested no improvement to the yield of ThiGH.

Table 2.4 Expression of ThiGH-His using a range of growth conditions and two expression systems.

pRL1020/BL21(DE3)

Cellular respiration	*Growth medium*	*Additives*	*Yield Cells (g/L)*	*Yield Protein (mg/g cell)*	*Iron Content◊ (ThiH:Fe)*	*Methods**
aerobic	2YT	-	7.6	5.2	1:1	11, 13
anaerobic	GMM	-	1.4	n/m	1:1	6.4.1, 6.4.4
	2YT	-	2.5	0.2	1:1	
		$NaNO_3$ $NaNO_2$	1.4	0.3	1:1	

pRL1021/BL21(DE3)

Cellular respiration	*Growth medium*	*Additives*	*Yield Cells (g/L)*	*Yield Protein (mg/g cell)*	*Iron Content◊ (ThiH:Fe)*	*Methods**
aerobic	2YT	-	7	3.5	1:2	11, 13
anaerobic	2YT	-	1	?	-	6.4.2, 6.4.4
		-	1.8	n/m	1:2	
		iron	1	0.72	1:1.2	
		tyrosine	1	n/m	-	
		tyrosine, iron	1	1.1	1:2.2	
			2.7	3.7	1:2.6	6.4.3, 13

◊ Iron content determined by the method of Fish (method 14)[137]; * Methods used for expression and purification; the numbers correspond to the experimental methods (general methods: 11, 13; experimental for chapter 2: 6.4.1, 6.4.2, 6.4.3, 6.4.4); n/m: not measured.

2.2.3 Improving purification methodology

2.2.3.1 Chromatography

i) Affinity chromatography: hexahistidine tag

ThiH from *E. coli* was successfully purified in a complex with ThiG using several expression plasmids (described above) that make use of a C-terminal hexahistidine-tagged ThiH (ThiH-His)[58]. This strategy has been employed for the isolation of many proteins containing an oxygen sensitive iron sulfur cluster[64,66,115,141]. Nevertheless, because the iron content of the purified protein was consistently lower then expected, an explanation for this result might be that the imidazole, used to elute proteins containing a His-tag, participates in the removal of the iron from the protein. Therefore other affinity chromatography methods, which do not involve elution with imidazole, were investigated.

ii) Strep-tag purification

The Strep-tag system was considered to be potentially valuable for metallo-enzymes such as ThiH, as an alternative to the His-tag method because it involves elution from an affinity column using dethiobiotin (DTB). This purification system is based on the interaction between the Strep-tag II affinity peptide (NH_2-WSHPQFEK-COOH)[142] and StrepTactin, a specifically engineered streptavidin[143] (figure 2.6, step 1). The elution is based on the displacement of the Strep-tag II protein by dethiobiotin (DTB, 2 mM), which is a specific competitor (figure 2.6, step 2). Dethiobiotin binds reversibly to streptavidin with an association constant (Ka) of 5×10^{13} M^{-1}, with a less tight binding constant then the natural ligand, biotin (1×10^{15} M^{-1}); this competitive elution is based on specificity, which allows the one-step purification providing a simple and efficient method[142,144].

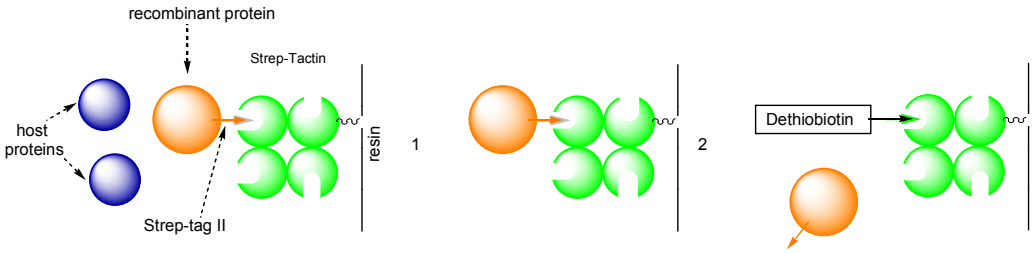


Figure 2.6 Step-tag purification system.

In order to use this method for ThiH purification, two plasmids were assembled: pFM700, encoding for *thiFSG* and a C-terminal Strep tagged *thiH* (ThiH-StagII) (figure 2.7) and pFM707, encoding for *thiFSGH*-StagII and the *isc* operon (figure 2.8) (method 6.4.5). Purified expression plasmids, pFM700 and pFM707, were used to transform *E. coli* BL21(DE3) cells (method 2) for cell culture and expression.

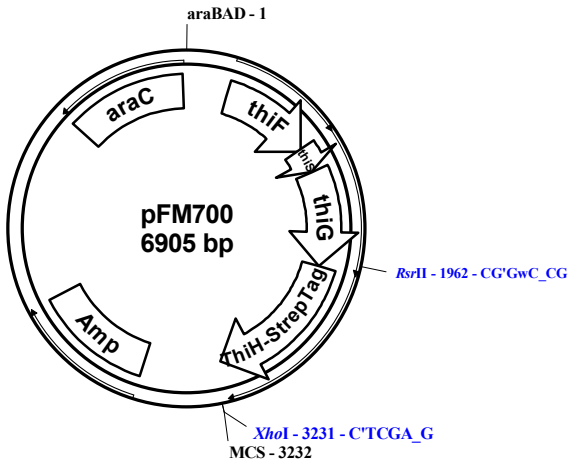


Figure 2.7 Map of pFM700. The restriction sites used to insert ThiH-StagII are indicated in blue (XhoI/NcoI). MCS = multiple cloning site. araBAD = operon promoter. Image created using pDraw32 DNA analysis software v1.0[145].

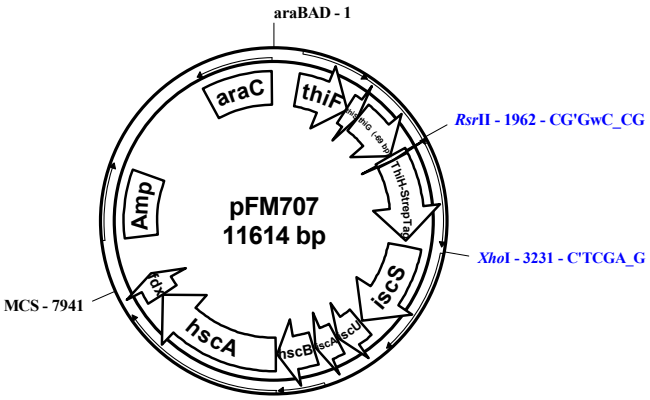


Figure 2.8 Map of pFM707. The restriction sites used to insert ThiH-StagII are indicated in blue (XhoI/NcoI). MCS = multiple cloning site. araBAD = operon promoter. Image created using pDraw32 DNA analysis software v1.0[145].

The cell paste obtained from the plasmid pFM707/BL21 (DE3) was dark brown (table 2.5), similar to that obtained with pRL1021 (both plasmids encoding Isc proteins). This dark brown colour is thought to be associated with the presence of [4Fe-4S] clusters. ThiGH-Stag obtained from these growths was purified with a Strep-tag II affinity column and the eluted fractions were analysed by SDS-PAGE gel (method 8).

Similar results were obtained from all the purifications using Strep-tagged proteins. As an example figure 2.9 shows the analysis of the fractions obtained from purification of pFM700/BL21(DE)3 grown aerobically.

Table 2.5 Expression of pFM700/BL21(DE3) and pFM707/BL21(DE3).

Expression plasmid	*Aerobic*		*Anaerobic*	
	Yield of cells g/L	*Colour of cells*	*Yield of cells g/L*	*Colour of cells*
pFM700	6.8	Light brown	3.6	Brown
pFM707	5.4	Dark brown	-	-
pRL1020	7.6	Light brown	2.5	Light brown
pRL1021	7.0	Dark brown	2.7	Dark brown

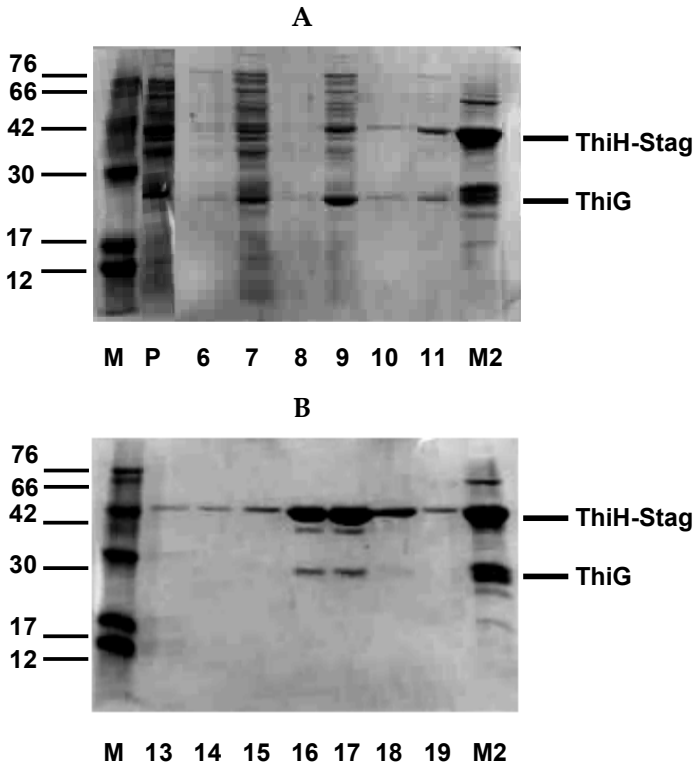


Figure 2.9 Coomassie Blue stained 15% SDS-PAGE gel of ThiGH-Stag purification expressed from pFM700/BL21(DE3) cells. M= molecular weight marker (kDa). P: re-suspended cell pellet; gel A: fractions 6-11 eluted during the wash with buffer A; gel B: fractions 13-19 eluted with buffer B (2.5 mM DTB); M2: purified ThiGH-His from pRL1020/BL21 (DE3). The numbers correspond to the fractions eluted from the column.

In all purification experiments it was observed that most of the ThiGH complex eluted directly through the column during the wash step (lanes 9, 10 and 11, figure 2.9), and only monomeric ThiH was eluted by dethiobiotin with very little ThiG associated. These observations may be explained if, due to the size of the ThiGH complex, the Strep-tag peptide attached to ThiH cannot efficiently bind to the resin; the quaternary structure of the ThiGH complex may cause the Strep-tag to be sequestered within the large complex and therefore is not able to reach into the rather deep biotin binding pocket of streptavidin (figure 2.10). Monomeric ThiH, however, binds more efficiently to the column, as observed by SDS Page gel analysis (figure 2.9). Figure 2.11 shows the hypothesised interaction of $(ThiGH)_6$ complex and monomeric ThiH to the StrepTactin resin.

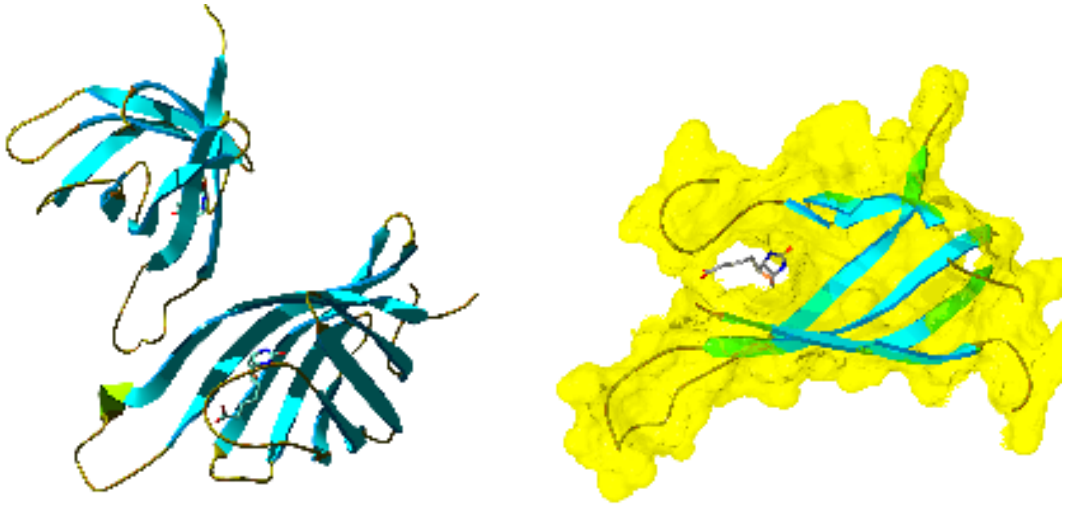


Figure 2.10 Crystal structure of avidin (1) and surface presentation of the biotin-binding region of avidin (2). Image created using POV-Ray version 3.6 and Swiss-Pdb Viewer Deep View version 4.0[146].

Iron analysis (method 14) of the purified sample shows that 1 mol of purified ThiH-StagII contains only 0.08 moles of Fe. This is a disappointing result, but it is consistent with hypothesis that *holo*-ThiH protein is stabilised by the formation of the complex with ThiG[58].

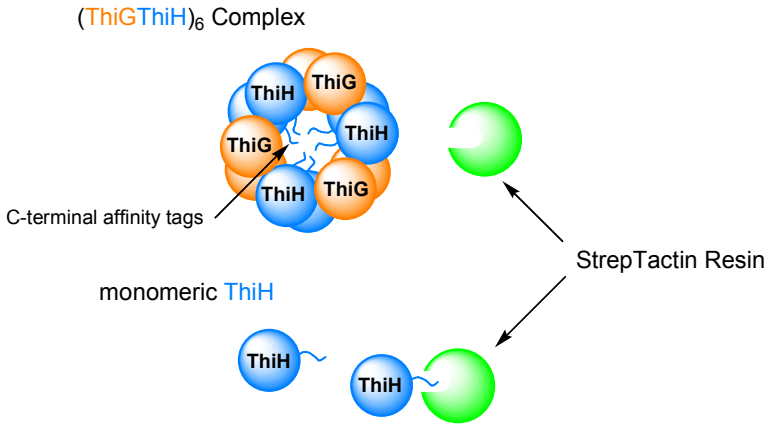


Figure 2.11 Schematic view of ThiGH complex and monomeric ThiH interacting with StrepTactin resin. For clarity only partial view of the structure is shown is the scheme.

One approach to solve this problem and improve the use of the Strep-tag system for ThiH purification would be the insertion of a flexible linker between the main ThiH

sequence and the strep-tag sequence, thus permitting more efficient binding, particularly of the ThiGH complex. Another alternative would be to add the Strep-tag to the N-terminus of ThiH, or to either terminus of ThiG. However, success in chemical reconstitution of ThiGH, isolated from pRL1020 and His-tag (described in section 2.2.4), made further developments of this method redundant and this approach was not continued.

iii) Ion exchange chromatography

The first attempts to purify ThiH were carried out aerobically using ion exchange chromatography[58]. This method was originally abandoned due to the presence of a 28 kDa protein, subsequently identified as ThiG.

Ion exchange chromatography was thought to be an alternative to the affinity chromatography since elution is based on a salt gradient, rather then imidazole, and therefore worthy of re-investigation. In this technique, separation is dependent upon the interaction between the charged solute molecules and the opposite charge immobilised ion exchange groups. These interactions can be controlled by varying conditions such as ionic strength and pH.

Q and S sepharose are commonly used resins in ion exchange chromatography. Q-Sepharose Fast Flow is a strong anion exchanger where the ion exchange group is a quartenary amine group ($–O–CH_2CHOHCH_2OCH_2CHOHCH_2N^+(CH_3)_3$). S-Sepharose Fast Flow is a strong cation exchanger. The ion exchange group is a methyl sulfonate ($-CH_2SO_3^-$). The elution of proteins bound to an ion exchange resin is typically accomplished by varying the concentration of salt in the elution buffer; also the ionic interaction between the protein and the resin can be manipulated as the charge of the protein is determined by its pI and the buffer pH.

The isoelectric point (pI) of ThiGH complex was estimated in order to assist on the choice of purification conditions, i.e. type of resin, pH. The pI of ThiG and ThiH was

calculated separately using Expert Protein Analysis System (ExPASy)[147]; $pI_{(ThiH)}$= 6.57 and $pI_{(ThiG)}$=5.36, as a result the average pI for the ThiGH complex is estimated to be 6.

ThiGH was expressed from pRL1021/BL21(DE3), aerobically in 2YT medium without the addition of supplements (method 7) and purified using Q-Sepharose at pH 8.0. The protein was eluted from the column using a gradient of salt (0-500 mM NaCl) (method 6.4.7). During the elution gradient three different coloured fractions were observed; lower concentration of salt eluted protein with light yellow colour, turning into greenish and after pinkish colour towards the end of the elution. ThiGH complex eluted in fractions 46-54, as shown by SDS-PAGE gel analysis (figure 2.12), together with a band tentatively identified as IscU (13.8 kDa), an iron sulfur cluster scaffold protein.

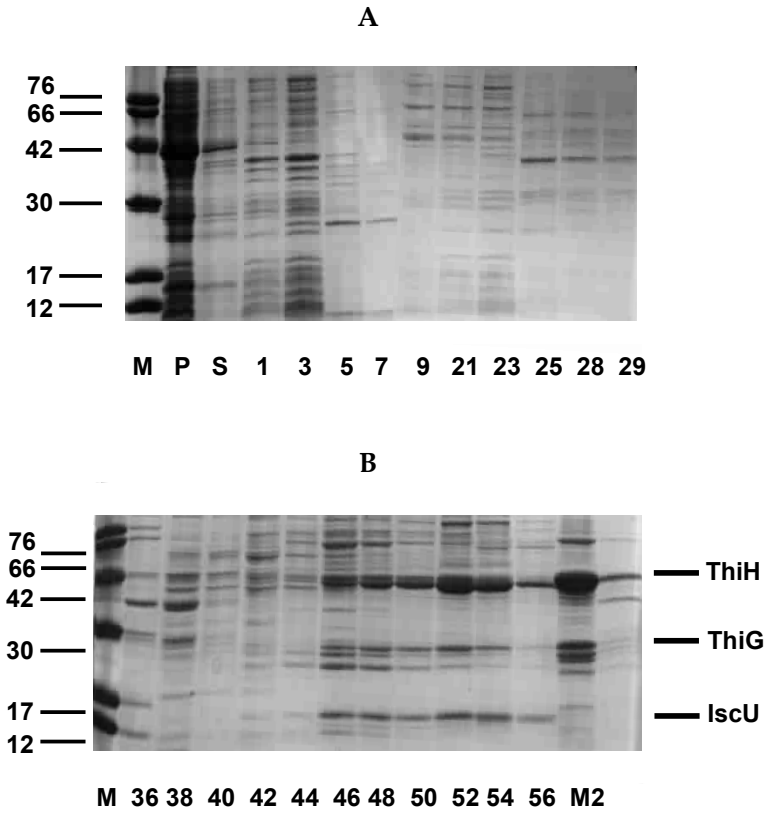


Figure 2.12 Coomassie blue stained 15% SDS-PAGE gel of ThiGH purification expressed from pRL1021/BL21(DE3) cells with Q-Sepharose. M= molecular weight marker (kDa); P= re-suspended pellet; S= supernatant; gel A: fractions 1-29 eluted during the wash using salt free buffer A; gel B: fractions 36-56 eluted using 500 mM NaCl buffer; M2= purified ThiGH-His from pRL1020/BL21(DE3) using nickel affinity chromatography. The numbers correspond to the fractions eluted from the column.

The UV-visible spectra of the different colour fractions (28, 48 and 54) were recorded (figure 2.13, 2.14 and 2.15).

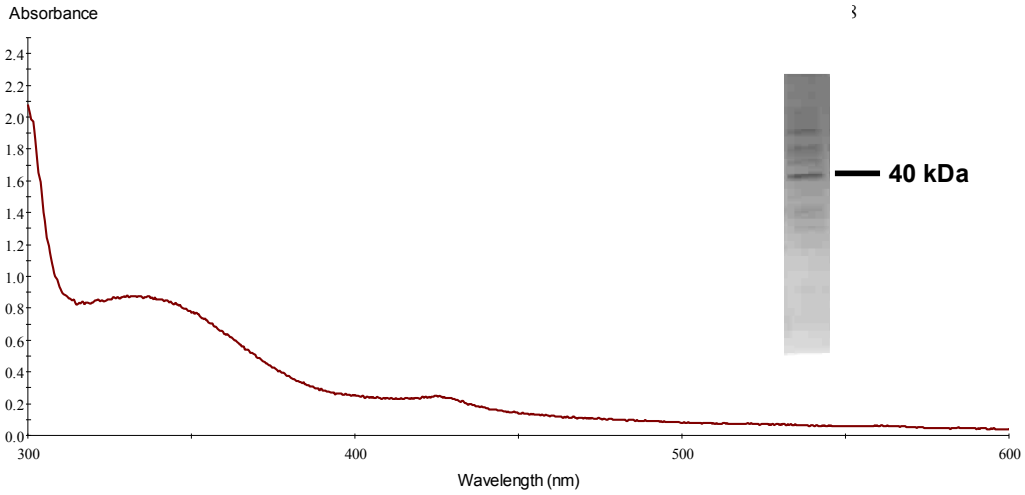


Figure 2.13 UV-visible spectra of fraction 28 (light yellow colour).

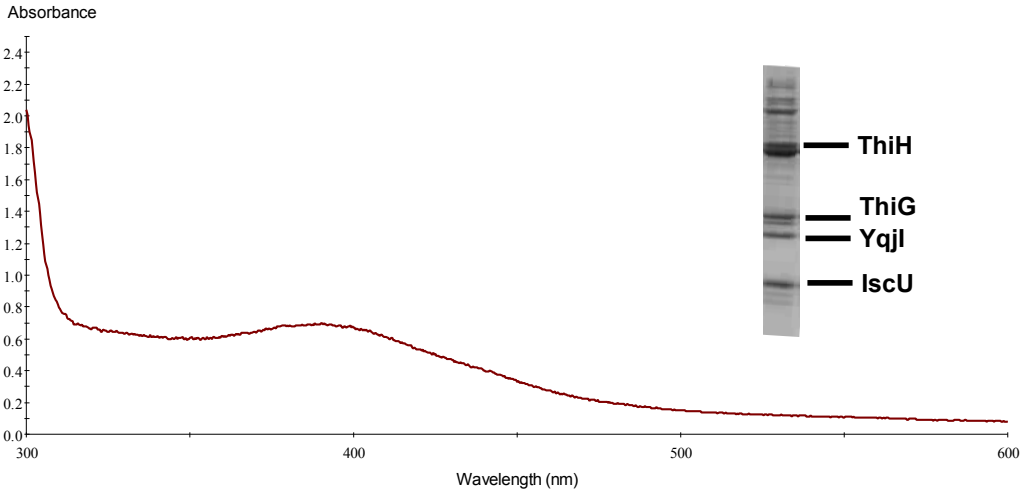


Figure 2.14 UV-visible spectra of fraction 48 (yellow-greenish colour).

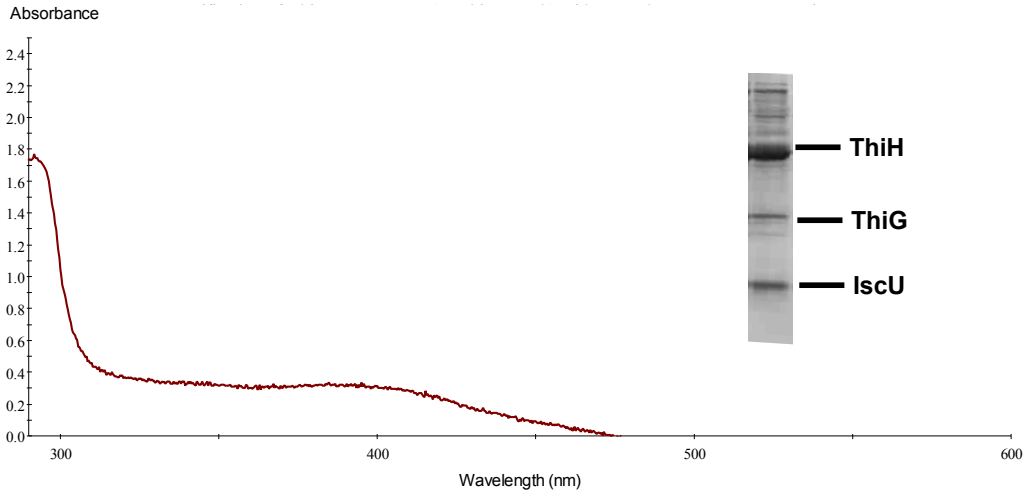


Figure 2.15 UV-visible spectra of fraction 54 (pinkish colour).

Fe-S cluster proteins are normally coloured because of their characteristic absorbance between 400 and 500 nm; more specifically they tend to be dark brown-reddish colour if containing [4Fe-4S] or [3Fe-4S] clusters[114,115,148,149]. UV-visible spectra analysis revealed that the shoulder at around 400 nm is probably associated with Isc proteins present in the samples. The spectra of fraction 28 has a shoulder at 410 nm which, initially was thought to be associated with an intact Fe-S cluster of ThiH but SDS-PAGE gel analysis revealed the presence of an unidentified 40 kDa protein, which might be IscS (~ 45 kDa). Studies by Mihara and co-workers[150] have shown that IscS has a maximum absorbance at 420 nm, but when reduced it loses the pyridoxal phosphate co-factor producing a peak at 335 nm; this is consistent with the observed spectra of fraction 28, the yellow colour of this fraction also indicates that it could be IscS. Fraction 48 shows a peak at ~400 nm, which could be attributed to a Fe-S cluster, but the presence of the 23kDa protein YqJI, together with the fact that fraction 54 has a lower absorbance at this wavelength (when YqJI is not present), leads to the conclusion that in fact this feature is due to the presence of this protein and not ThiH related.

Fraction 48 was further purified using S-sepharose resin at pH 6.7. The pH of the protein sample was adjusted before purification to 6.7. The protein was eluted from the

column using a gradient of salt (0-500 NaCl) (method 6.4.8). Figure 2.16 shows a SDS-Page gel analysis of the fractions obtained from this purification.

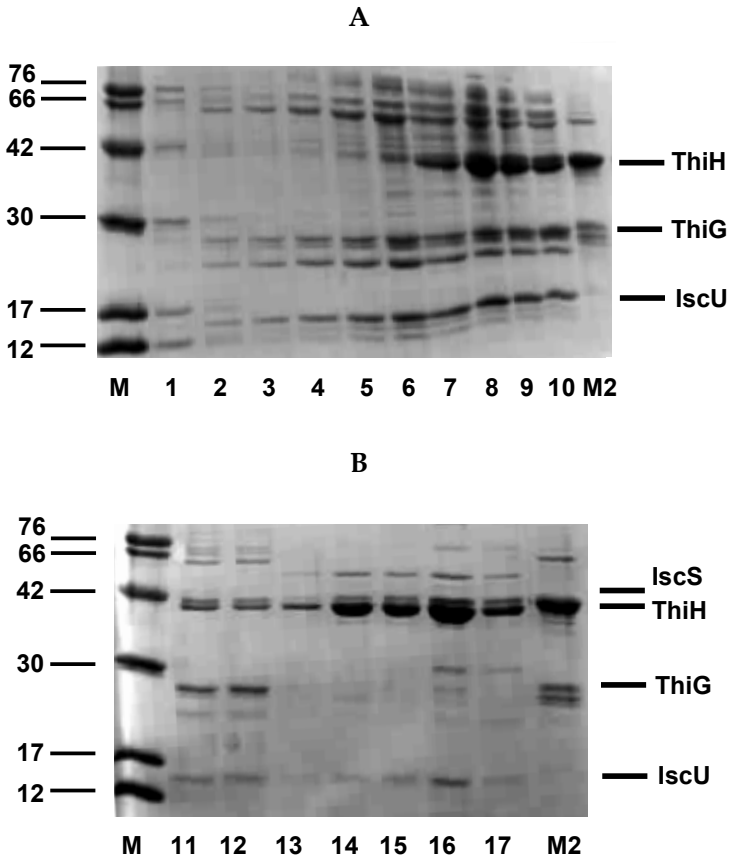


Figure 2.16 Coomassie Blue stained 15% SDS-PAGE gel of ThiGH purification from fraction 48 eluted from Q-Sepharose using S-Sepharose at pH 6.7. M= molecular weight marker (kDa); gel A: fractions 1-10 eluted during the wash with salt free buffer A; gel B: fractions 11-17 eluted with 500mM NaCl buffer; M2: purified ThiGH-His from pRL1020/BL21(DE3) using nickel affinity chromatography. The numbers correspond to the fractions eluted from the column.

During the beginning of the salt gradient, a small amount of ThiGH complex was eluted from the column (fractions 11 and 12), together with other proteins. Towards the end of the gradient, mainly monomeric ThiH (44 kDa) eluted together with the proteins IscU (13.8 kDa) and IscS (45 kDa) (fractions 14-17). This chromatographic method proved to be ideal for the isolation of monomeric ThiH, but due to previous observations of monomeric ThiH behaviour, i.e. high tendency to precipitate when not in a complex with ThiG, this purification method was not used for large scale

purifications. However, after a successful chemical reconstitution and isolation of stable *holo*-ThiH, using His-tag affinity chromatography (described in sections 2.2.4 and 2.2.5), this method could now be re-considered for isolation of *holo*-ThiH.

2.2.3.2 Modified purification buffers

Throughout this project, many variations to the cell growth medium and chromatography were attempted. Using nickel affinity chromatography, and expression from pRL1020/BL21(DE3), which proved to be so far the best expression system for ThiGH, a range of purification buffers were tested (method 6.4.9). The changes in purification buffers are discussed in the next paragraphs and listed in table 2.6. All the purified protein samples were analysed for iron content using the method of Fish (method 14).

(a) Tris buffer, pH 8.0: this buffer was used in the initial purification procedure developed by R. Leonardi (method 12)[58]. ThiGH complex was purified in a 1:1 ratio eluting consistently in fractions 9 to 23. The earlier eluted fractions contain higher concentration of ThiGH complex (9 – 17) and later fractions mainly contained monomeric ThiH (18 – 23). Only fractions containing high ratio of ThiH:ThiG complex were desalted and used for further studies. Even if this method was highly reproducible, purified samples of ThiGH not only contain low iron content but showed a high tendency to precipitate. To overcome this problem glycerol was added routinely to the samples to a final concentration of 25% (w/v). ThiGH complex could be concentrated to a maximum of 5-6 mg/mL. The iron content of the purified samples was 1:1 (mol Fe/mol ThiH).

(b) MOPS buffer, pH 7.7: since Tris buffer interferes with amino acid analysis, the purification buffer was replaced by MOPS buffer (method 13). The use of this buffer did not alter the behaviour of the protein during purification. ThiGH complex was still routinely obtained around the same fractions, and the ratio of ThiGH complex to monomeric ThiH was similar to that obtained whilst using Tris. However protein purified using this buffer, possibly assisted by the addition of an anaerobic cell lysis

step, was far more stable. In addition, the yield of protein was much higher then that of previous experiments. For this reason, the addition of extra glycerol to the final protein samples was not required; and also less glycerol was added to the purification buffers [10% (w/v) rather than 12.5% (w/v)]. The iron content of the purified samples using this buffer increased compared to the previous method, and was in average 1.4:1 (mol Fe/mol ThiH).

(c) MOPS buffer, pH 7.7 supplemented with dithionite: to improve anaerobicity, dithionite (final concentration 1 mM) was added to the purification buffers. ThiGH complex was isolated successfully and in a stable form. However, the iron content of the purified ThiGH samples was low, containing only 0.6 mol of Fe per mol of ThiH.

(d) MOPS buffer, pH 7.7supplemented with AdoMet: AdoMet might prevent oxidative degradation of the iron sulfur cluster of ThiH during purification, by competing with oxygen to bind the unligated iron of the cluster (section 2.2.2, scheme 2.1). Due to the significant amounts of AdoMet required, a low-cost source of AdoMet was identified. SAM-e® tablets are commercially available pills that contain AdoMet (also abbreviated as SAM). An extraction procedure was developed, followed by quantification of AdoMet present in the tablets by RP-HPLC (method 6.4.10). On average, each tablet contained approximately 35% (w/w) of pure AdoMet. Bearing this in mind, AdoMet was added accordingly to the purification buffers to final concentration of 0.5 mM. ThiGH complex was purified with high stability. The highest yield of protein was obtained using this method (8.5 mg protein/g cell) but unfortunately no improvement was observed on the iron content of the samples.

(e) MOPS buffer, pH 7.7 supplemented with AdoMet and tyrosine: the addition of substrates might also prevent oxidative degradation of the iron sulfur cluster[140]. AdoMet (final concentration 0.5 mM) and tyrosine (final concentration 1 mM) were added to the purification buffers. The protein obtained was highly stable and contained higher iron content, 1.5 mol of Fe per mol of ThiH.

(f) MOPS buffer supplemented with tyrosine: due to the successful isolation of ThiGH with increase of iron content and stability (e), tyrosine was added to the purification buffers; *1) pH 7.7; final concentration of tyrosine 1 mM*: ThiGH eluted using these purification buffers eluted normally, and a significant increase in iron content was observed (2.5 mol of iron per mol of ThiH). However, after storage at - 80 ºC, freeze and thawed, samples showed low stability resulting in slight protein precipitation. *2) pH 7.4, final concentration of tyrosine 2 mM*: ThiGH eluted using these purification buffers eluted normally, and as in previous experiments significant increase in iron content was observed (2.0 mol of iron per mol of ThiH). After storage at -80 ºC, freeze and thawed, samples showed low stability resulting in slight protein precipitation.

(g) MOPS buffer, variation to pH 7.0 and 6.5: variations on the pH of the purification buffers were carried out. ThiGH was purified using the novel method (method 13), and eluted as expected. However after storage at -80 ºC, freeze and thawed the purified protein samples showed very low stability and high tendency to precipitate. Nevertheless, the iron content of ThiGH was higher then when purified at pH 7.7 (pH 7.0: 1.8 mol of iron per mol of ThiH; pH 6.5: 1.7 mol of iron per mol of ThiH).

Table 2.6 Buffers used for purification of ThiGH-His from pRL1020/BL21(DE3).

Buffer	*pH*	*Additives*	*Yield of protein (mg/g cell)*	*Iron content (ThiH/Fe)*	*Protein stability**	*Method**
*Tris**	*8.0*	-	2.2	1:0.9	Medium	12
		-	2.8	1:1	Medium	12a
*MOPS***	*7.7*	-	5.2	1:1.2	High	13
		dithionite, 1 mM	4.1	1:0.6	Low	6.4.9
		AdoMet, 0.5 mM	8.5	1:1	High	
		AdoMet, o.5 mM Tyrosine, 1 mM	5.7	1:1.5	Medium	
		Tyrosine, 1 mM	5.1	1:2.4	Medium	
	7.4	Tyrosine, 2 mM	4.0	1:2	Low	
	7.0	-	4.3	1:1.8	Very low	
	6.5	-	3.6	1:1.7	Very low	

* 12.5% (w/v) glycerol; ** 10% (w/v) glycerol

As an example, figure 2.17 shows the SDS-PAGE gel analysis of ThiGH purification using MOPS buffer (pH7.7, 2 mM tyrosine).

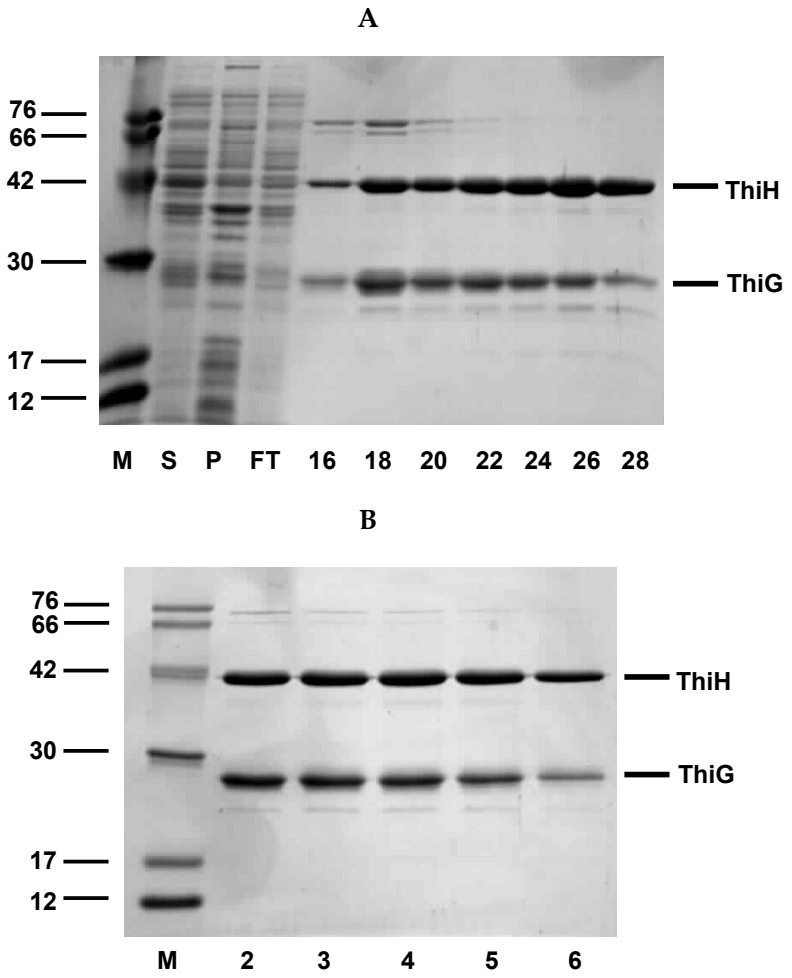


Figure 2.17 Example of a purification of ThiGH-His using modified purification buffers. Coomassie Blue stained 15% SDS-PAGE gel of ThiGH purification using MOPS buffer, pH 7.4, containing 2mM tyrosine. M: molecular weight marker (kDa); S: supernatant; P: re-suspended pellet; FT: flow through from nickel affinity column; gel A: fraction 16-28 eluted from nickel affinity column; gel B: fractions 2-6 eluted from gel filtration column. The numbers correspond to the fractions eluted from the column.

2.2.4 *In vitro* chemical reconstitutions of ThiH

In parallel with these purification studies, Dr. M. Kriek of our laboratory was investigating the chemical reconstitution of ThiGH. These experiments were critical to achieving active *holo*-ThiGH and are discussed in the following section.

Fe-S clusters can easily interconvert (figure 2.18), depending on the oxidising or reducing conditions[65,148,151,152]. They can undergo conversion from the free to the protein-bound form and also oxidative degradation[140].

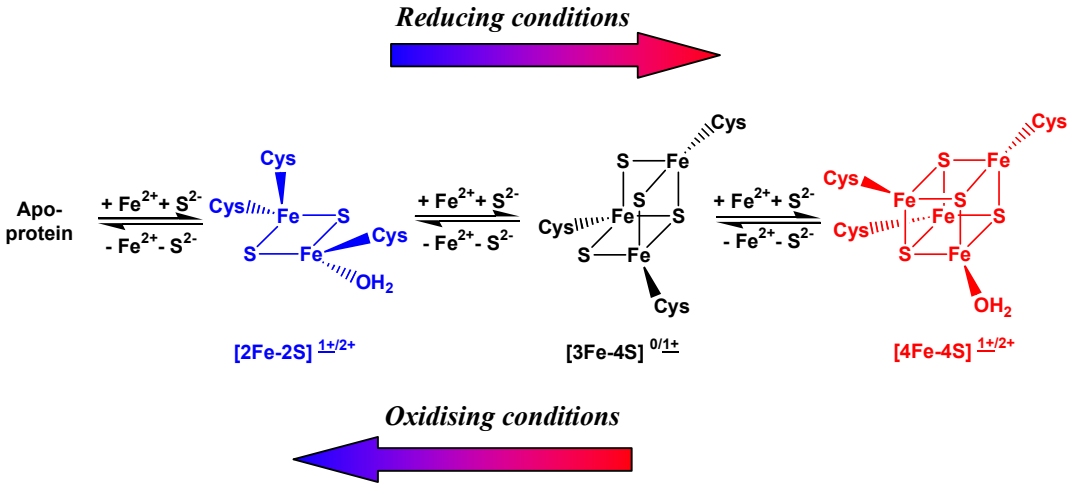


Figure 2.18 Interconversion between Fe-S clusters. Adapted from[58].

As discussed in section 2.2.2, the biogenesis of iron sulfur clusters requires a group of specific proteins (Isc proteins). However, protein-bound [2Fe-2S] and [4Fe-4S] clusters can be obtained *in vitro* by chemical reconstitution. Studies with BioB and LipA show that by incubating apo-protein with an excess of iron and sulfide (5-10 molar excess) under anaerobic conditions[62,153,154] it is possible to obtain fully reconstituted protein containing about 4Fe/polypeptide chain[65,151]. Chemically reconstituted ArrAE and MiaB, however, contain only 2-3 Fe/polypeptide chain[68,115]. Anaerobically purified ThiGH complex, from pRL1020/BL21(DE3), typically contains 1.0 ± 0.4 mol equivalent of Fe per mol of ThiH, but was occasionally higher (up to 2 equivalents of Fe).

An optimised protocol for chemical reconstitution of ThiH was developed by Dr. M. Kriek over many purifications and months. This optimisation involved the collaboration of the author who provided protein samples used for these experiments and final stages of optimisation. The optimised method can be summarised as follows: purified ThiGH complex was gently mixed with DTT solution (1 mL of 200 mM stock solution prepared in buffer D: 50 mM MOPS, 100 mM NaCl, 10% (w/v) glycerol) to a final concentration of 83 μM. After 15 minutes, 5 mol eq of $FeCl_3$ in water (40 μL, 10 mM) was added very carefully in small aliquots (5 μL), and after a further 10 minutes, 5 mol eq of Na_2S in water was added likewise (total added 40 μL, 10 mM). The protein solution was then incubated for another 2 hours, and the precipitated iron sulfide was removed by centrifugation. Because iron particles are very fine, further removal was carried out by applying the solution to a NAP-10 gel filtration column pre-equilibrated and eluted with buffer D (method 6.4.11).

The iron content was analysed by the method of Fish (method 14). Reconstituted ThiGH could be concentrated to a maximum concentration of 40 mg/mL using a Biomax 5000 molecular weight cut-off Ultrafree 0.5-mL centrifugal filter (method 6.4.11).

Whilst this method was being developed, several factors were identified as important:

- The incubation period with fresh DTT solution, should be around 15 minutes, as samples incubated for longer period show more tendency to precipitate upon the addition of $FeCl_3$.
- The $FeCl_3$ solution proved to be more effective when freshly prepared from black anhydrous $FeCl_3$, whereas the use of brown (wet) $FeCl_3$ did not result in a successful reconstitution, i.e. ThiGH precipitation.
- The addition of the $FeCl_3$ and Na_2S solutions have to be carried out very slowly, small aliquots added with intervals and gentle mixing, otherwise resulting in protein precipitation.

- The concentration process has to be carried out very carefully to avoid precipitation, allowing small intervals of mixing the solution whilst concentration is taking place.

After chemical reconstitution and gel filtration, the iron content was measured as 5.2 ± 1.3 mol equivalents of Fe per mol of ThiH, suggesting some residual non specifically bound iron may remain in these samples (method 14).

Reconstituted protein could be stored at -80 ºC without any effect on the stability of the protein. However, reconstituted and concentrated samples showed a slight tendency to precipitate when frozen and thawed. Therefore, concentration of the protein was only carried out when protein needed to be used for other experiments.

Comparative UV-visible spectra of purified ThiGH and purified/reconstituted ThiGH were recorded. This experiment was carried out using an Ocean Optics spectrophotometer attached to a cuvette holder placed inside the glove box (method 6.4.12).

Samples were thawed inside the glove box, diluted with anaerobic buffer D [50 mM MOPS, 100 mM NaCl, 10% (w/v) glycerol] to a final concentration of 1 mg/mL and spectra recorded at the range 250-700 nm, against buffer D. For reconstituted protein, the iron excess removed from samples by centrifugation (2 minutes, 12,000 rpm) and protein UV-vis spectra recorded. Figure 2.19 shows the UV spectra obtained.

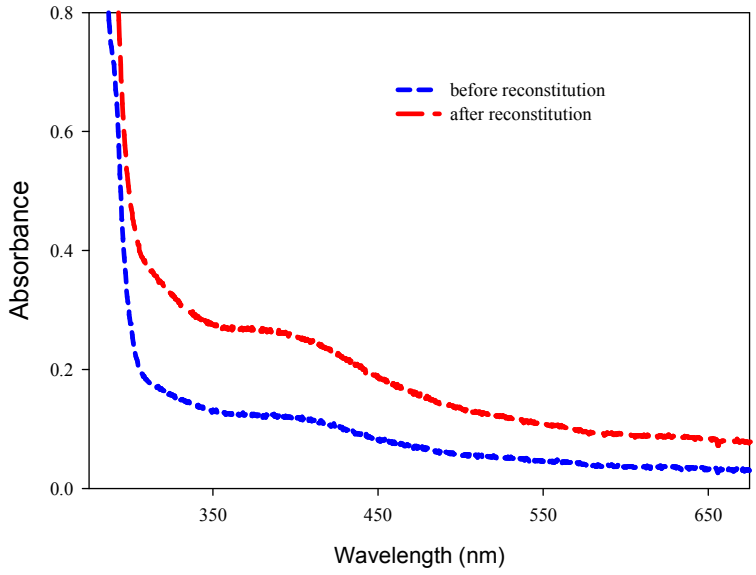


Figure 2.19 UV-visible spectra of purified ThiGH before and after chemical reconstitution.

The spectra exhibits a maximum absorption at 410 nm which is characteristic for sulfide-to-iron charge transfers within Fe-S centres[65,151], clearly more accentuated for reconstituted samples of ThiGH, indicating the successful reconstitution of the iron sulfur cluster. UV-vis spectra of ThiGH obtained are in agreement with those reported by R. Leonardi[24], however with a stronger absorbance between 300 and 450 nm. The extinction coefficient for reconstituted ThiGH at 390 nm was 4678 cm^{-1} M^{-1}/iron atom (assuming a 4Fe-4S cluster).

2.2.5 Analytical gel filtration chromatography

With the ultimate objective of solving the crystal structure of the ThiH or ThiGH complex, the separation of monomeric ThiH from ThiGH complex was investigated. Analytical gel filtration chromatography was carried out under strictly anaerobic conditions (method 6.4.13). ThiGH-His was expressed from pRL1020/BL21(DE3), purified (method 13), reconstituted, concentrated to 40 mg/mL and loaded onto a S-200 column pre-equilibrated with buffer D [50 mM MOPS pH 7.7, 200 mM NaCl, 5 mM

DTT, 10% (w/v) glycerol]. Protein elution was monitored at 280 nm. Fractions were collected and analysed by SDS-PAGE (method 8). Figure 2.20 shows the chromatogram (a) and the corresponding SDS-PAGE gels (b) obtained from the gel filtration. ThiGH complex eluted at t~120 min and monomeric ThiH eluted at t~160min.

The iron content of the eluted samples was measured (method 14). Interestingly, both ThiGH complex and monomeric ThiH samples had high iron content after the gel filtration step (table 2.7). This demonstrates the high stability of not only the ThiGH complex but also of ThiH, since hardly any protein precipitated on during the elution. In fact, monomeric ThiH was further concentrated to 60 mg/mL, which proves the efficiency and significant improvement of expression, purification, reconstitution and concentration methods, since the maximum previously obtained by R. Leonardi was 6 mg/mL, when in a complex[58].

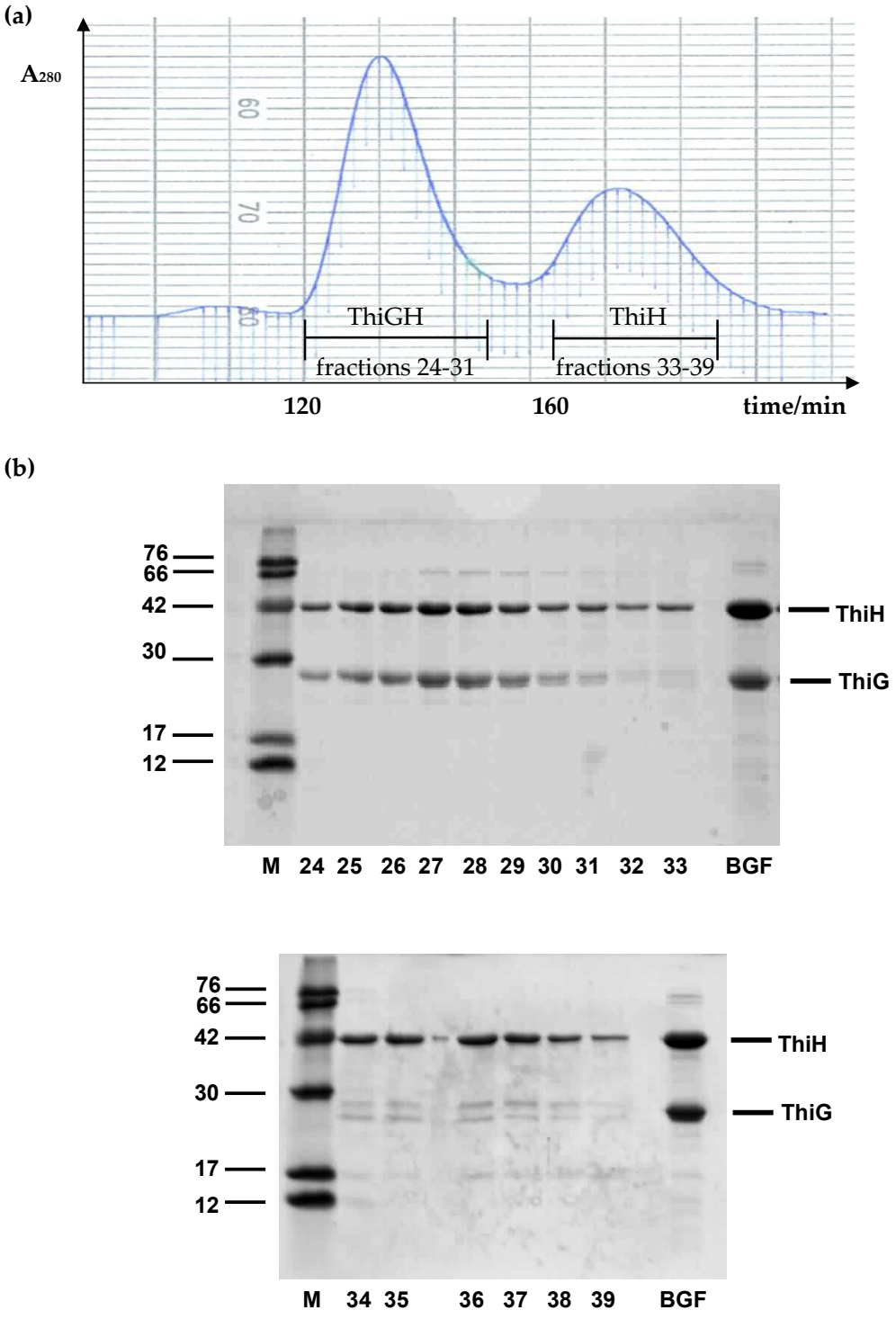


Figure 2.20 Analytical gel filtration analysis of ThiGH-His isolated from pRL1020/BL21(DE3). (a) chromatogram recorded at 280 nm; (b) Coomassie Blue stained 15% SDS-PAGE gel of fractions eluted from S-200 column. M: molecular weight marker (kDa); BGF: ThiGH initial sample before analytical gel filtration. The numbers correspond to the fractions eluted from the column.

Table 2.7 Iron content of the eluted samples from analytical S-200 column.

Protein present	*Fraction number*	*mg/mL protein*	*Iron content mol Fe/mol ThiH*
ThiGH	24	0.438	4.5
ThiGH	25	0.880	2.8
ThiGH	26	0.689	2.4
ThiH	34	0.443	3.8
ThiH	35	0.469	3.6
ThiH	36	0.396	3.5

2.2.6 Initial studies on the crystallisation of ThiH

After the successful optimisation of the chemical reconstitution and concentration of ThiGH and monomeric ThiH, studies were initiated on the crystallisation of both ThiGH complex and monomeric ThiH.

To achieve good diffraction data, it is necessary to obtain well ordered crystals. Locating and optimising the conditions required to produce diffraction quality crystals can be difficult due to the number of factors that can affect protein stability and crystal growth:, i.e., purity of protein, pH, ionic strength, counter ions, and temperature, nature of the precipitant.

Because there is not an unique method to crystallise proteins, a large number of screening experiments are prepared in a microplate using a finite number of crystallization conditions with varying salt, buffer, pH, precipitants and other chemical components[155,156].

The sitting drop vapour diffusion method has been successfully used for producing good quality crystals of a wide range of proteins and was chosen to initiate the crystallisation studies of ThiGH. In this method, a drop of protein is mixed with a precipitant and is slowly dehydrated in a sealed well, by equilibration with a reservoir

containing higher precipitant solution. Due to the vast number of parameters, the formulation of trial screens often derives from past successful crystallisation trials. For the ThiGH initial crystallisation studies, a trial screen for containing a range of precipitants and buffers was used[157].

ThiGH-His was expressed from pRL1020/BL21(DE3), purified (method 13), reconstituted, concentrated to 40 mg/mL. An analytical gel filtration chromatography was carried out under strictly anaerobic conditions (method 6.4.13) to separate ThiGH complex from monomeric ThiH. The iron content of the protein samples was measured (table 2.7). Fractions 24-26 of ThiGH complex were pooled and concentrated to 28.3 mg/mL. Fraction 34 of monomeric ThiH was concentrated to 60 mg/mL. These protein samples were then used to set up the crystallisation trial experiments. All crystallisation experiments were carried out under strict anaerobic environment, inside a glove box, in a controlled temperature room.

Screening for optimal crystal growth conditions of ThiGH and monomeric ThiH were prepared in 24 well plates. Precipitant solutions were prepared and dispensed onto the plate (1000 μL each). Protein solutions were completed by the addition of further reagents (for example AdoMet, tyrosine or Dxp). Drops were prepared by placing drops of protein solution (1.5 μL) onto cover slips (22 mm diameter) and then adding precipitant solution from the well (1.5 μL). The cover slip was then carefully inverted to cover the rim of each well of the plate. Table 2.8 shows the protein, buffer and additives on each tray.

Table 2.8 Protein, additives and buffers used for crystallisation

Tray nº	Protein/ volume added	Additives*/ volume added	Buffer
1	ThiGH/30 µL	AdoMet/3.5 µL Dxp/1.75 µL Tyrosine/1.75 µL	MOPS pH 7.7
2	ThiGH/90 µL	AdoMet/9.0 µL Dxp/4.5 µL Tyrosine/4.5 µL	Tris pH 8.8
3	ThiH/45 µL	AdoMet/4.5 µL Dxp/2.25 µL Tyrosine/2.25 µL	MOPS pH 7.7

* Stock solutions of 10 mM.

The trays were checked at weekly intervals for 3 months. Precipitates were commonly observed, but no reproducible crystallisation conditions were observed.

2.3 Summary and conclusions

Studies on improving the *in vivo* production of ThiH were carried out using a range of expression conditions and chromatography techniques.

Studies were carried out using two different expression plasmids (pRL1020 and pRL1021). Expression of ThiGH using pRL1021, showed the consistent co-expression of YqJI protein, and no significant improvement on the stability and iron content of ThiGH. Expression of pRL1021 with *E. coli* knock out (-YqJI) cells carried out by Mr. M. Challand showed no significant improvement on the production of monomeric ThiH. pRL1020 expression plasmid is currently used for production of ThiGH.

Addition of supplements to the growth medium showed no improvement on the production and iron content of ThiGH. Strep-tag purification method proved to be ineffective, but ion exchange chromatography, in particular using S-sepharose, proved

to be a good candidate for isolating *holo*-ThiH. Unfortunately at the time these experiments were carried out, due to previous observations of monomeric ThiH behaviour, i.e. high tendency to precipitate when not in a complex with ThiG, this purification method was discarded. ThiGH is routinely obtained from pRL1020/BL21(DE3) and isolated using His-tag affinity chromatography.

Variation of the purification buffer showed significant improvement on the yield of protein; from 2.2 mg/g cell obtained using tris buffer to 5.2 mg/g cell using MOPS buffer. The use of MOPS buffer and, possibly also the addition of anaerobic cell lysis step to the purification protocol, produced very stable ThiGH complex. Addition of supplements, in particular tyrosine, to the purification buffers produced protein with higher iron content, however the stability was low, resulting in protein precipitation after freezing and thawing.

For the first time, successful *in vitro* chemical reconstitution of ThiH was achieved (5.2 ± 1.3 mol equivalents of Fe per mol of ThiH). This allowed for the ThiGH to be concentrated to ~40 mg/mL, and further separation of ThiGH complex/monomeric ThiH showed that ThiH could be concentrated up to 60 mg/mL.

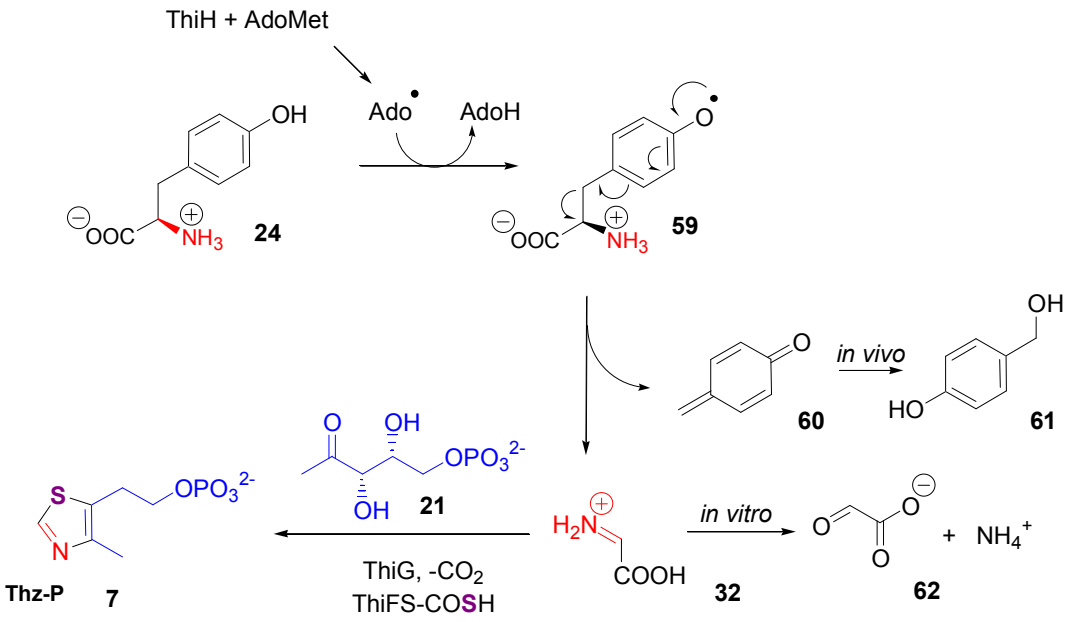
Initial studies on crystallisation of ThiGH complex and monomeric ThiH were unsuccessful.

CHAPTER 3

Investigating Thiazole Biosynthesis in *Escherichia coli*

3.1 Introduction

The study of the mechanism by which Thz-P is assembled in *E. coli* is one of the main focuses of this project. As introduced in chapter 1, the biosynthesis of this thiamine precursor has now been well characterised in the aerobe *B. subtilis* and a mechanism has been proposed[43,44,158]; the C2-N3 fragment of Thz-P derives from the oxidation of glycine to dehydroglycine, in a reaction catalysed by the flavoenzyme ThiO. This hydrolytically unstable intermediate is incorporated into the thiazole in a multistep reaction requiring the thiazole synthase (ThiG)[45,159] the sulfur donor ThiFS thiocarboxylate[27,40] and 1-deoxyxylulose 5-phosphate (Dxp). However, in aerobes such as *E. coli*, the steps leading to the formation of the thiazole moiety are still unclear. Dehydroglycine **32** (scheme 3.1) has been proposed as the first common intermediate for bacterial thiazole biosynthesis[160], this precursor is formed from tyrosine **24** in a ThiH dependent reaction[161,162]. This biosynthetic step therefore requires the cleavage of the C_α to C_β bond of tyrosine and release of the aromatic side chain (scheme 3.1). ThiH shows sequence similarity to the 'radical- *S*-adenosylmethionine (AdoMet)' family of proteins[59], including conserved ligands to an essential [4Fe-4S] cluster (section 1.4)[52] and has been shown to form a complex with ThiG (section 2.2)[24]. The proposed mechanism for Thz-P biosynthesis in *E. coli* is outlined in scheme 3.1.



Scheme 3.1 Original proposed mechanism for the formation of Thz-P in E. coli[58].

The [4Fe-4S]$^{1+}$ cluster of ThiH catalyses the reductive cleavage of AdoMet to generate Ado•. This highly reactive radical will then abstract a hydrogen atom from tyrosine **24** and form a tyrosyl radical **59** that will react further cleaving the Cα-Cβ bond. Dehydroglycine **32** is released together with the quinone methide **60**. Hydration of **60** leads to *p*-hydroxybenzyl alcohol **61**, a known *in vivo* thiamine by-product[113].

Initial studies on the development of an *in vitro* assay to measure ThiH activity, were carried out, attempting to monitor the formation of *p*-hydroxybenzyl alcohol (*p*-HBA) by HPLC. Despite extensive efforts, the formation of *p*-hydroxybenzyl alcohol could not be confirmed, but, interestingly, the presence of a reaction product with higher retention time than *p*-HBA was observed (results not shown). This work was carried out by the author. At this stage, the project was divided into two main routes of investigation: the production of *holo*-ThiH (described in chapter 2), and the development of a functional *in vitro* assay. These two lines of investigation were carried out in parallel, the first one by Dr. M. Kriek in the development of the *in vitro* assay; and a second route by the author in the production of *holo*-ThiH. During these investigations, the author and Dr. M. Kriek collaborated and a number of purified

samples of ThiGH were provided by the author for the development of the *in vitro* assays. In addition, extensive studies were carried out by the author to develop and optimise a reliable analysis method to detect the *in vitro* reaction products, which included: capillary electrophoresis, LC-MS (results not shown) and HPLC analysis. HPLC experiments were successfully applied and are described in detail in the next section.

The optimisation of the purification methodology and *in vitro* chemical reconstitution made possible the isolation of *holo*-ThiGH (section 2.2). This allowed for the development of an *in vitro* assay that not only allowed the measurement of the activity of the isolated protein, but ultimately facilitates the study of the mechanistic enzymology by which Thz-P is biosynthesised in *E. coli*.

The development of a functional *in vitro* assay was carried out by Dr. M. Kriek from our laboratory. After achieving the successful chemical reconstitution of ThiGH, electron paramagnetic resonance (EPR) studies were carried out to test the effect of the addition of substrates to the reconstituted *holo*-ThiGH samples (figure 3.1)[160].

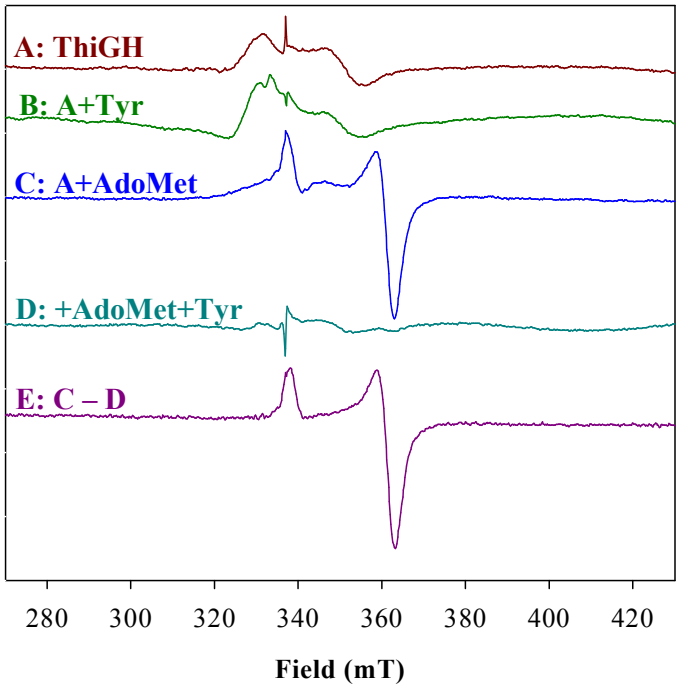
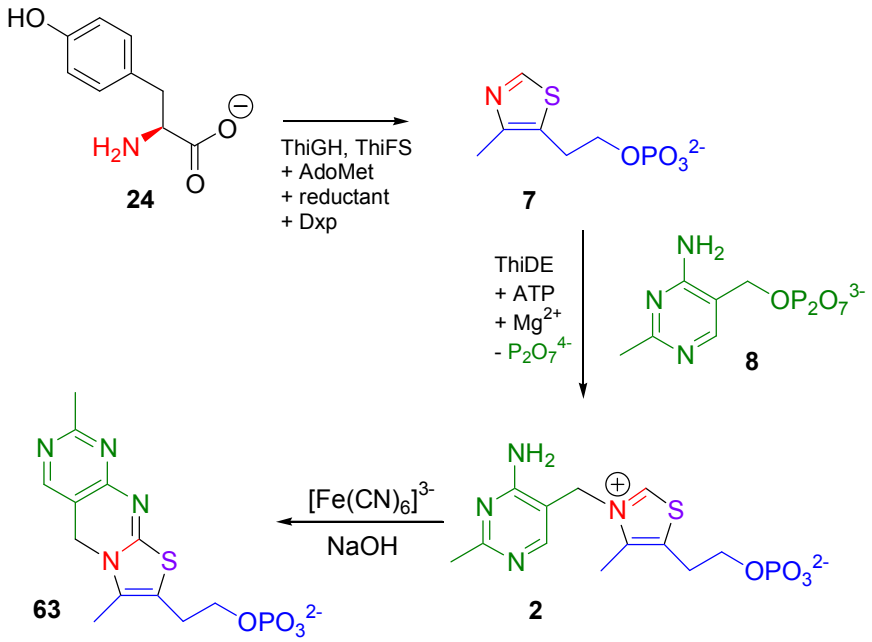


Figure 3.1 X-band EPR spectra of ThiGH. Adapted from[160].

The addition of tyrosine did not produce a marked change in the EPR spectrum (figure 3.1, B), whereas the addition of AdoMet had a major effect (figure 3.1, C). This observation can be explained by the reduction of the iron sulfur cluster of ThiH when binding to AdoMet. Changes of reduction potentials for $[4Fe\text{-}4S]^{2+/1+}$ have been observed for other members of the 'radical AdoMet' family such as anaerobic ribonucleotide reductase activating enzyme[163] and lysine aminomutase[69,164] (discussed in chapter 1). In comparison, a ThiGH sample which had both AdoMet and tyrosine added showed a much weaker EPR signal (figure 3.1, D). The rapid disappearance of the $[4Fe\text{-}4S]^{1+}$ cluster signal upon addition of tyrosine in the EPR experiments shows that, at least under these conditions, the reaction with tyrosine is faster than the rate at which the cluster can convert back to the reduced state[160]. These studies are consistent with the proposed mechanism for tyrosine cleavage (scheme 3.1), and previous studies by R. Leonardi using purified *E. coli* proteins have shown that ThiGH requires AdoMet and a reductant for activity[51].

With the objective of reconstituting thiazole synthase activity an *in vitro* assay was developed[160]. This assay consisted of reconstituted and concentrated ThiGH, anaerobically purified complex of ThiFS and a reducing system, [flavodoxin (FldA), flavodoxin reductase (Fpr) and NADPH]. Also, three substrates were shown to be essential: tyrosine, AdoMet and Dxp. Thz-P is highly polar and therefore difficult to detect by HPLC. An enzyme coupled reaction assay was carried out using 4-amino-5-hydroxymethyl-2-methylpyrimidine (**8**, Hmp), ATP and the enzymes ThiD and ThiE. The TP formed in the assays was converted to the thiochrome **63** [165], separated by HPLC and quantified using a fluorescence detector (scheme 3.2).



Scheme 3.2 In vitro assay for detection of TP formation[165].

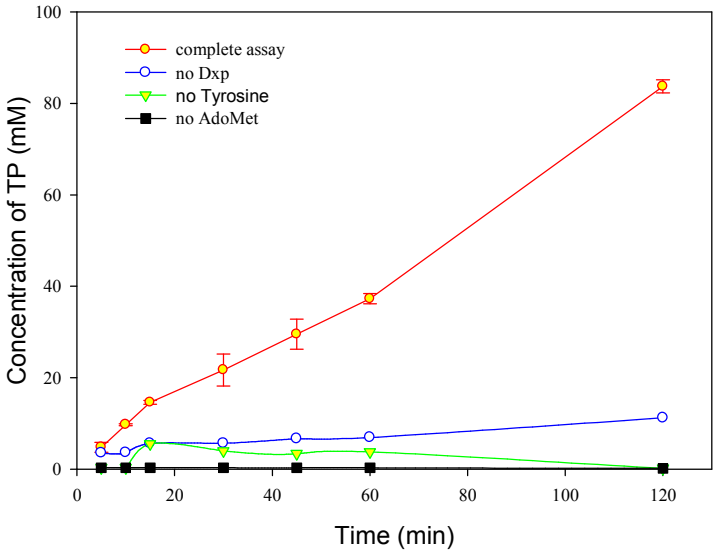


Figure 3.2 In vitro time course of thiamine phosphate formation: the TP formed in the assays was converted to the thiochrome, separated by HPLC and quantified using a fluorescence detector. Adapted from[160].

HPLC analysis of the fluorescent derivatives clearly shows that the purified ThiGH is active, since the control experiments, lacking one of the substrates, results in no formation of TP (figure 3.2).

Radiochemical studies of *in vitro* assays using ^{14}C labelled tyrosine were the first steps to elucidate the structure of the aromatic side-chain released from tyrosine cleavage. By using two TLC systems, two radiolabelled products with very different polarities were observed (figure 3.3). The highly polar component co-eluted with glyoxylate **62** and comparison of the less polar product with several aromatic standards showed that it co-eluted with *p*-cresol **63**. This experiment also demonstrated the absolute requirement for AdoMet as the negative control showed no activity.

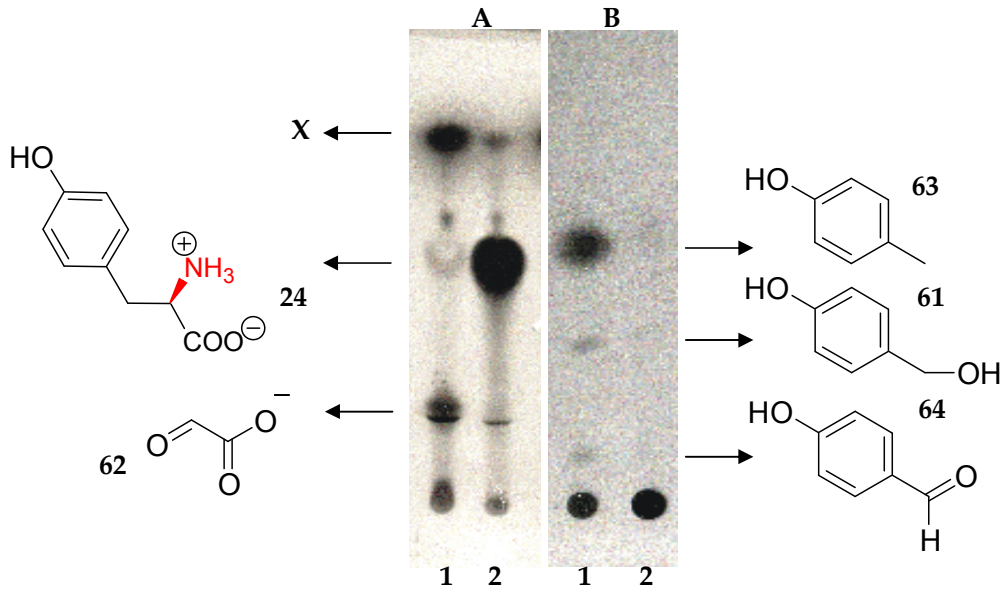


*Figure 3.3 Autoradiograms of a ThiH assay. Two TLC systems (plates A and B) were used to resolve the products. The complete assay (lane 1 on each plate) contained L-[U-^{14}C]-tyrosine, reconstituted ThiGH, AdoMet and a reductant. The negative control (lane 2) lacks AdoMet. Plate A: tyrosine **24**, glyoxylate **62**; plate B: p-cresol **63**, p-hydroxybenzyl alcohol **61**, p-hydroxybenzaldehyde **64**. Adapted from*[160].

Further studies with these assays using nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (GC-MS), confirmed that the products derived from tyrosine cleavage are *p*-cresol and glyoxylate[166].

3.2 Results and discussion

3.2.1 ThiH *in vitro* assay

The optimised *in vitro* assay protocol was developed by Dr. M. Kriek[160]. The following proteins and reagents were added, anaerobically, to a reaction mixture in the following order (with final concentrations): reconstituted and concentrated ThiGH (255 µM), tyrosine (0.51 mM), AdoMet (0.74 mM), NADPH (0.67 mM), FldA (62 µM), Fpr (10 µM). Negative control experiments were prepared using the same reagents but lacking

tyrosine. FldA, Fpr and NADPH were mixed and left for 15 min until the solution turned blue (resulting from the semiquinone radical form of Fpr). ThiGH, AdoMet and tyrosine solutions were mixed and after the reductant mixture was added. The reactions (200 μL aliquots each) were incubated at 37 ºC and stored at -80 ºC. For time course experiments reactions were incubated at 37 ºC for 5, 15, 30, 45, 60, and 120 min and subsequently stored at -80 ºC.

A standard method was developed for the analysis of these supernatants as follows. Samples were thawed and proteins were immediately precipitated with 20% perchloric acid (12 μL) and then cleared by centrifugation (14,000 rpm, 15 min), and the supernatants analysed by HPLC. This method was used throughout the experiments described in this chapter.

3.2.2 Identification of the intermediate and by-product derived from tyrosine

The formation of the two products, glyoxylate and *p*-cresol, observed by Dr. M Kriek on the autoradiography studies, was further confirmed by ^{13}C NMR spectra of reactions containing *U*-^{13}C-tyrosine, ThiGH, AdoMet and a reducing system (NADPH/FldA/Fpr) which were incubated at 37 ºC for two hours. Negative control assays were also carried out using the same reagents but lacking the reducing system. Protein was precipitated by addition of perchloric acid and removed by centrifugation. NMR analysis of the aqueous supernatant showed signals typical of glyoxylate [figure 3.4, (a)] signals at 173 and 86 ppm under these conditions, observed as doublets as a result of ^{13}C-^{13}C coupling.

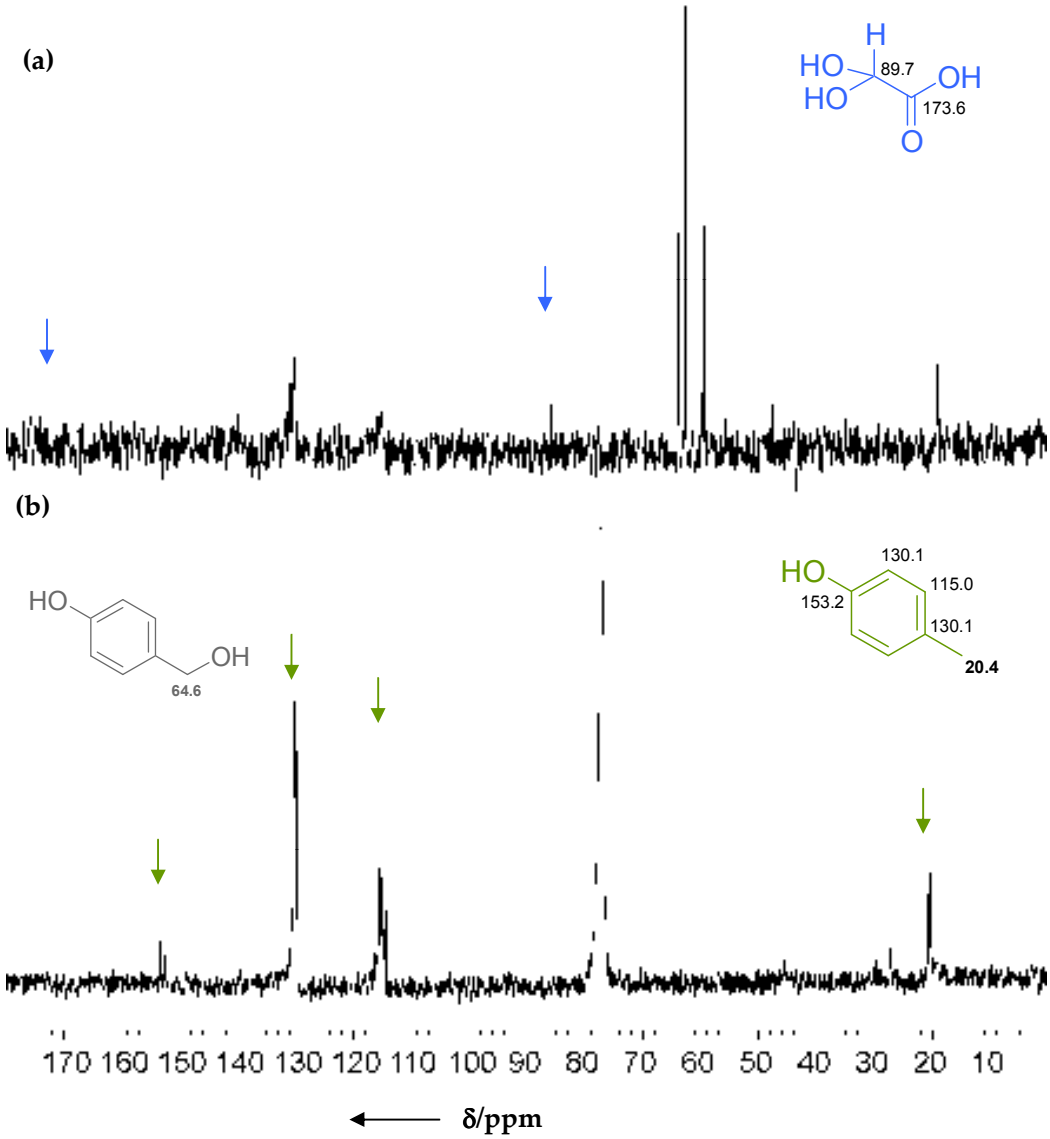


Figure 3.4 NMR spectra of a ThiH assay complete assay containing ^{13}C-tyrosine. (a) analysis of the assay supernatant after protein removal; (b) analysis of the supernatant extracted with $CDCl_3$.

An organic extract of the supernatant was prepared using $CDCl_3$ and analysed by ^{13}C NMR. The spectrum of the organic extract prepared from the supernatant is consistent with ^{13}C-labeled *p*-cresol [^{13}C NMR (100 MHz; $CDCl_3$) δ = 153.2 (t, J = 66 Hz, COH), 130.1 (m, 2CH, CCH_3), 115 (m, 2CH), 20.4 ppm (m, CH_3)] (figure 3.4, (b)).

The spectrum of the supernatant derived from the negative control confirms the presence of unreacted tyrosine (figure 3.5).

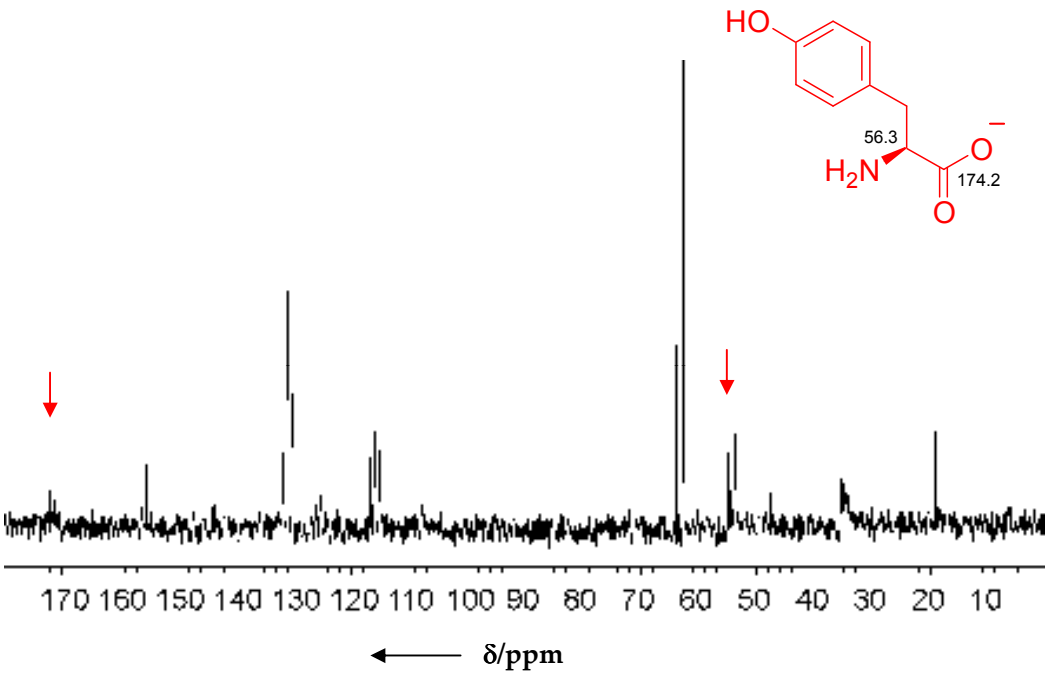


Figure 3.5 NMR spectra of a negative control ThiH assay containing ^{13}C-tyrosine.

The organic extract, containing the aromatic by product derived from the tyrosine side chain, was analysed by GC-MS (figure 3.6). A peak was observed with a retention time of 13.38 min, identical to that observed for a *p*-cresol standard. The expected mass for the *p*-cresol (calculated) was 108.2, and the observed mass was 108.2. As for the *p*-cresol derived from ^{13}C-tyrosine, the expected mass (calculated) was 115.14, and the observed mass was 115.2.

In conclusion, both NMR and GC-MS analysis clearly demonstrate that *p*-cresol and glyoxylate were produced from tyrosine during *in vitro* reactions.

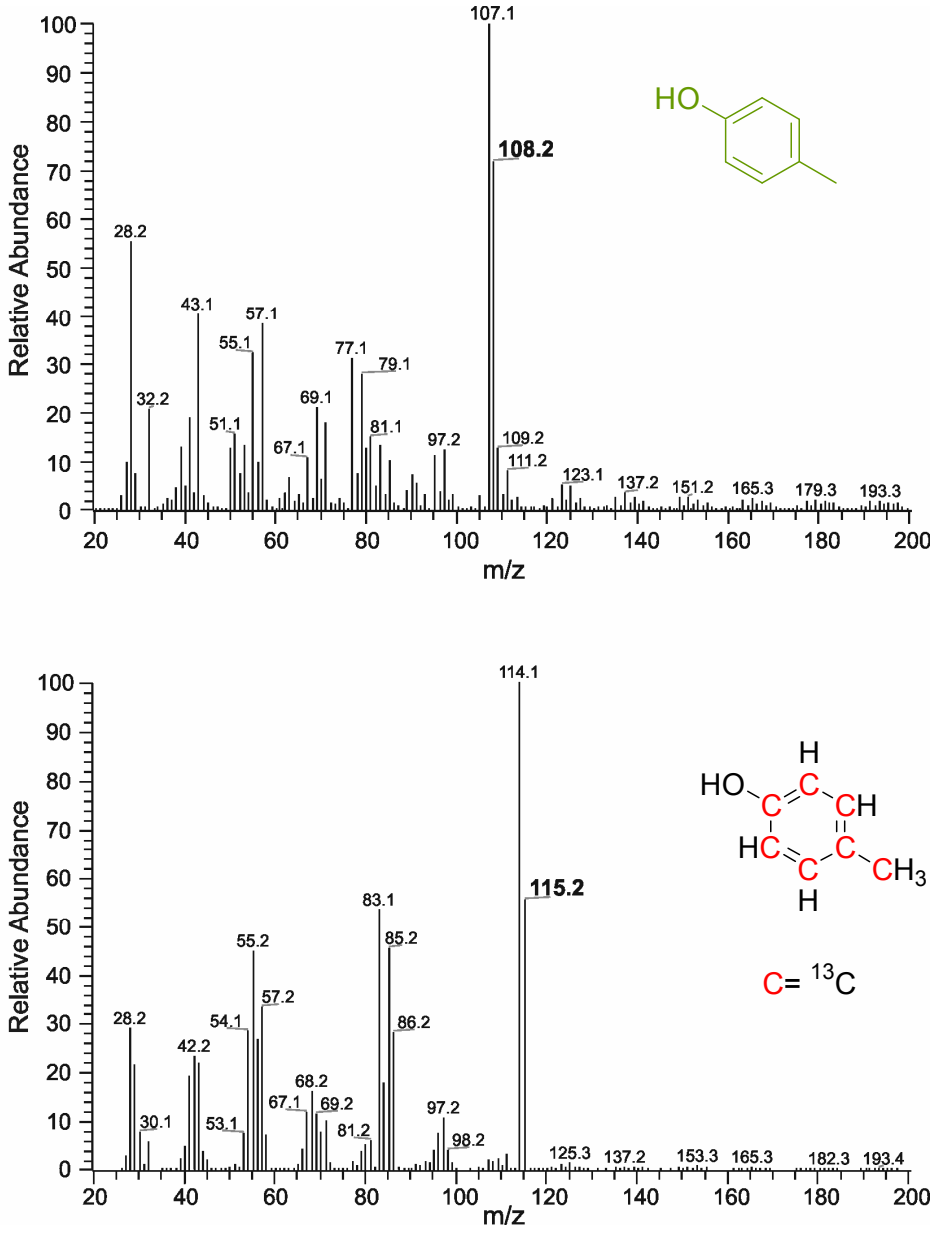


Figure 3.6 GC-MS analysis of a ThiH assay complete assay containing ^{13}C-tyrosine. (a) p-cresol, m/z = 108.2 (calculated mass 108.14); (b) U-^{13}C-p-cresol, m/z = 115.2 (calculated mass 115.14).

3.2.3 Detection and quantification of *p*-cresol, AdoH and tyrosine by HPLC

The development of an HPLC method, to investigate the reaction products of tyrosine cleavage, was initiated with an optimisation of the HPLC gradient conditions using standard solutions of: 5'-deoxyadenosine (AdoH), which is formed upon AdoMet reaction with the Fe-S cluster of ThiH; tyrosine; and *p*-cresol, the aromatic product released upon tyrosine cleavage.

A range of HPLC conditions were investigated to achieve optimal separation of the *in vitro* assay components and reaction products. Table 3.1 shows a brief summary of the principle methods investigated. These are described in more detail in experimental section (section 6.5.2).

Table 3.1 HPLC methods

Method nº	Column	Aqueous phase	Organic phase	Separates
1	Hypersil BDS C18	0.1% CH_3CO_2H	CH_3OH/ 0.1% CH_3CO_2H	HBA, *p*-cresol
2	Phenomenex C18 RP	0.1% TFA	CH_3CN/ 0.1%TFA	AdoMet, *p*-cresol
3	Synergy polar RP	0.1% HCO_2H	CH_3CN/ 0.1% HCO_2H	AdoH
4	Phenomenex ODS (3)	0.1% TFA	CH_3CN/ 0.1%TFA	AdoH, tyrosine
5	Phenomenex C18 RP	25 mM CH_3CO_2H/ 8 mM $C_8H_{18}O_3S$	50% CH_3CN/ 50% Aq. phase	AdoMet
6	Phenomenex C18 RP	0.1% CH_3CO_2H	CH_3OH/ 0.1% CH_3CO_2H	Tyrosine, AdoH, *p*-cresol

Separation of the main *in vitro* assay products was achieved by using method 6 which was used throughout the project to analyse all assay mixtures. Aliquots of standard solutions (40 μL) of AdoH, tyrosine and *p*-cresol were analysed by HPLC using UV-visible detection (λ_{ex} 223 nm, λ_{em} 280 nm). A range of concentrations of standards were used to construct calibration curves (section 6.5.2) for estimating the concentrations of tyrosine, AdoH and *p*-cresol. Concentrations were determined from experiments carried out in duplicate. The following retention times were observed for standards: tyrosine, R_t = 8 ± 0.3 min; AdoH, R_t = 11.8 ± 0.3 min; *p*-cresol, R_t = 16.7 ± 0.3 min. Unless otherwise stated all the data from the experiment described in this chapter were plotted using *SigmaPlot v 10.0.*

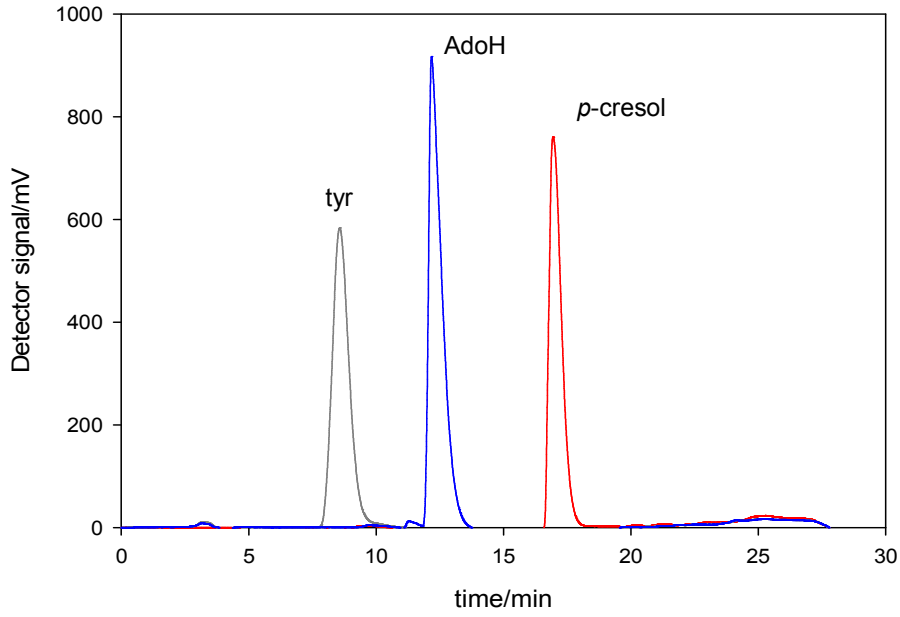


Figure 3.7 HPLC analysis with UV detection (280 nm) of standard solutions of tyrosine, AdoH and p-cresol. The chromatogram corresponds to 10 nmoles injected of each standard solution.

Assay mixtures containing purified and reconstituted ThiGH, AdoMet, tyrosine and FldA/Fpr/NADPH, were incubated at 37 ºC for one hour. ThiGH and other protein components were removed by acid precipitation and centrifugation. Aliquots of the resultant supernatants (40 μL) were analysed by reverse phase HPLC (RP-HPLC) (figure 3.7). A negative control experiment was carried out under the same conditions but without the addition of AdoMet.

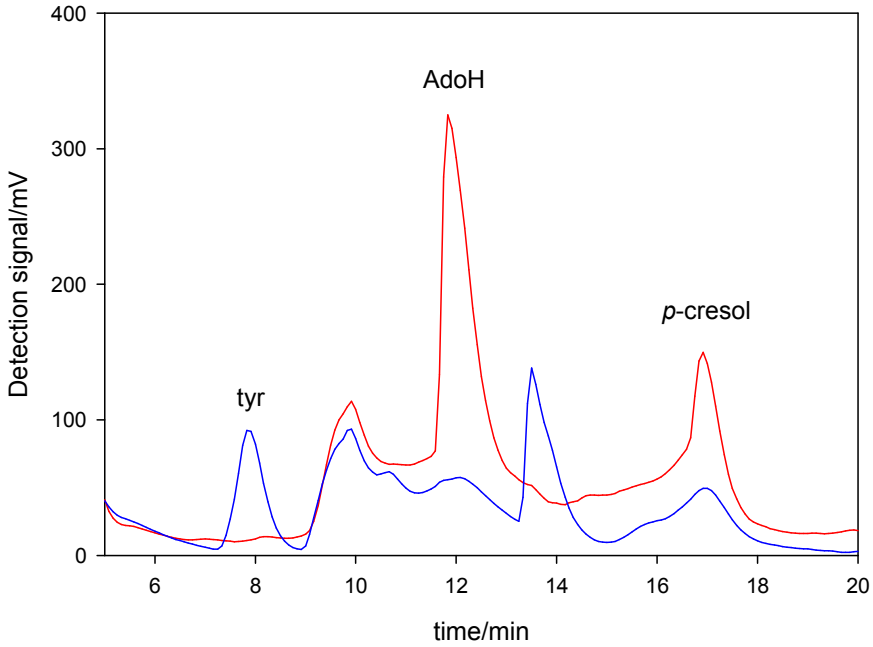


Figure 3.8 HPLC analysis with UV-vis detection of an in vitro assay: full assay (—); negative control (—).

Time course experiments were carried out. Assay mixtures were incubated at 37 ºC for 0, 15, 30, 60 and 120 minutes. After the removal of protein components, supernatants were analysed by RP-HPLC. Because the reactions were prepared inside an anaerobic glove box, it would be practically difficult to stop the assay immediately, i.e. to prepare a really reliable time zero as the first point; therefore, the time = 0 data point was calculated from a full assay to which no reducing system was added.

RP-HPLC analysis at 280 nm (figure 3.8) allowed the simultaneous measurement of AdoH and *p*-cresol formation in the *in vitro* assay. Tyrosine consumption was also measured. This HPLC method was optimised to allow for the products and substrates of interest to separate from other components present in the assay. The retention times for tyrosine (R_t = 8 ± 0.4 min), AdoH (R_t = 11.8 ± 0.4 min) and *p*-cresol (R_t = 16.7 ± 0.5 min) were consistent with those obtained from the standard solutions.

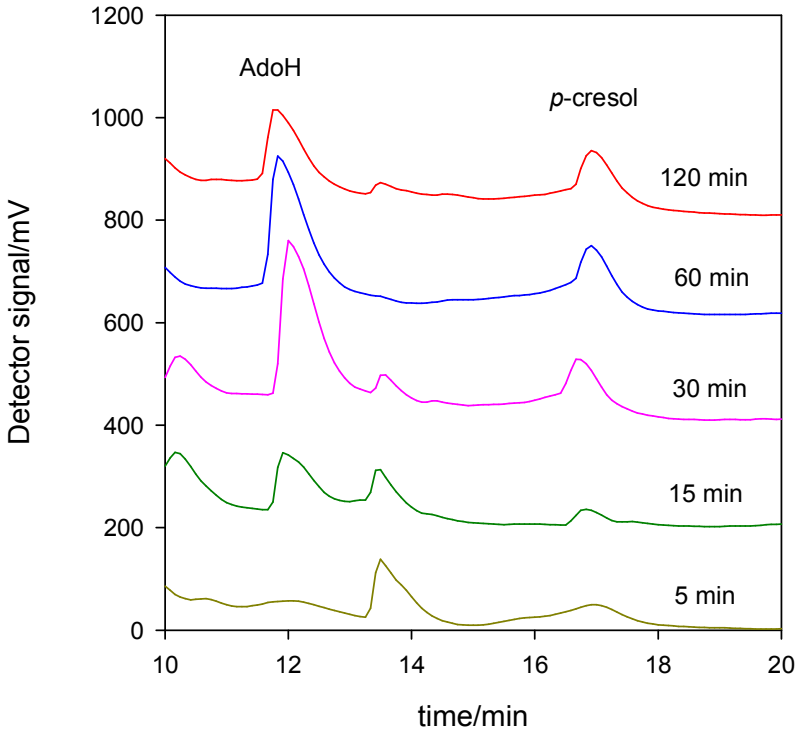


Figure 3.9 HPLC analysis with UV-vis detection of an in vitro assay time course. For clarity traces corresponding to incubation at 15, 30, 60 120 minutes are offset by 200, 400, 600 and 800 mV respectively.

The consumption of tyrosine, and AdoH and *p*-cresol formation was quantified by HPLC data by comparison of the peak areas with the calibration curves from standard solutions of each component. The preparation of the standard solutions and calibration curve plotted by HPLC analysis are described in detail in section 6.5.2. The concentrations measured for each time point are listed in tables 3.2, 3.3 and 3.4.

Table 3.2 Time course of tyrosine consumption. SD is standard deviation.

Time/min	[tyrosine]/μM			
	sample 1	sample 2	average	SD
0	154.8	128.1	137.9	14.7
15	72.7	76.5	62.6	20.8
30	0.0	30.3	26.1	24.3
60	0.0	0.0	0.0	0.0
120	0.0	0.0	0.0	0.0

Table 3.3 Time course of AdoH formation. SD is standard deviation.

Time/min	[AdoH]/μM			
	sample 1	sample 2	average	SD
0	0	0	0	0
15	134.7	126.0	130.3	6.1
30	155.1	161.6	158.3	4.6
60	182.7	189.2	185.9	4.6
120	165.0	162.1	163.5	2.1

Table 3.4 Time course of p-cresol formation. SD is standard deviation.

Time/min	[*p*-cresol]/μM			
	sample 1	sample 2	average	SD
0	0	0	0	0
15	18.4	33.4	25.9	10.6
30	86.5	113.5	100.0	19.0
60	130.0	136.2	133.1	4.4
120	85.0	143.9	114.5	41.7

In order to investigate the kinetics and stoichiometry of substrate consumption and product formation, the *p*-cresol, AdoH and tyrosine (table 3.5) formed in this assay were quantified by HPLC and plotted together (figure 3.10).

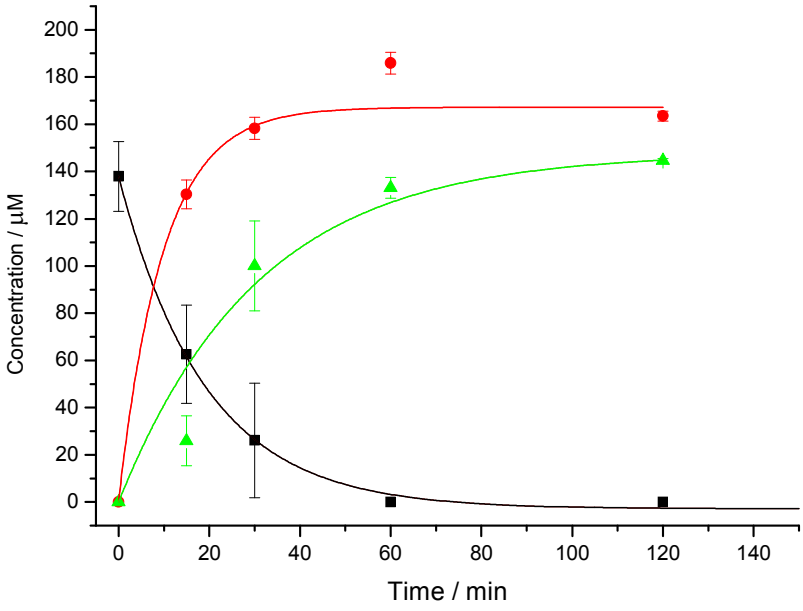


Figure 3.10 Time course of AdoH and p-cresol formation, and tyrosine consumption in a ThiGH assay. AdoH (—●) tyrosine (—■) and p-cresol (—▲).Error bars are derived from standard deviation of two data sets (tables 3.1, 3.2 and 3.3). The curves are fit to a first order exponential function.

The apparent first order rate constants of *p*-cresol and AdoH formation, and for tyrosine consumption, were calculated. Data from figure 3.10 were fitted, using *Origin v 6.0 analysis software*, to a first-order process to determine and calculate the final concentrations of products and substrate and the rate constant (equation 6.1).

$$y = A\,[1 - \exp(-Bt)] \qquad \textit{equation 6.1}$$

[y = final concentration of product or substrate; t = time, A = calculated final concentration, B = apparent first order rate constant (min^{-1})]

By comparison with standards of known concentrations, HPLC was used to measure the quantity of tyrosine and AdoMet utilized and the amount of AdoH and *p*-cresol produced during a time course experiment. The rate of formation of AdoH is

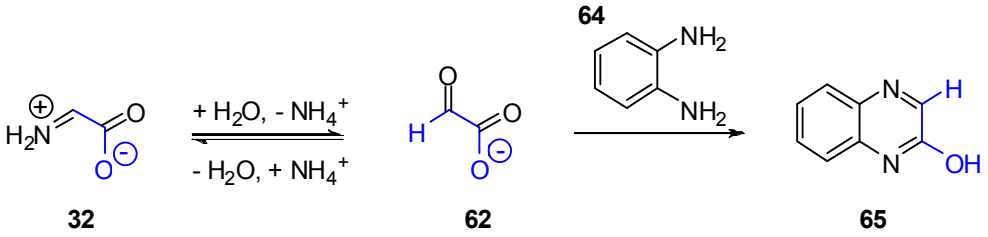
approximately two times faster then the rate of formation of *p*-cresol and the rate of consumption of tyrosine (table 3.5) and the reaction is shown to approach completion after 1 hour. The maximal expected concentration of AdoH was 160 μM, assuming one mole equivalent was generated per mole of tyrosine. However, the amount of AdoH was higher and reached a maximum of 186 μM. This result suggests that there may be some cleavage of AdoMet which does not lead to hydrogen abstraction from tyrosine. Similar results have been reported with *in vitro* assays of BioB and LipA, where for each mole of biotin generated approximately 2.9 moles of AdoH[167]; and in LipA assays where up to 2.4 mol equivalents of AdoH are formed with respect to the substrate[168]. Data from this set of assays indicated that the maximum production of *p*-cresol occurs at 60 min, with 133 μM being produced.

Table 3.5 Kinetic parameters for substrate utilisation and product formation.

Product	Rate constant/min	Final concentration/μM
Tyrosine	0.052 ± 0.003	0.0 ± 0.0
AdoH	0.102 ± 0.033	167.2 ± 6.0
p-cresol	0.032 ± 0.003	147.7 ± 1.9

3.2.4 Detection and quantification of glyoxylate by HPLC

In the absence of other components required for thiazole formation. dehydroglycine **32**, formed upon tyrosine cleavage by AdoMet, is hydrolysed to yield glyoxylate **62** and ammonia. The direct detection of glyoxylate by HPLC is difficult because such polar molecules are usually eluted with very short retention times. McNeill and co-workers[169] developed a very effective method to detect 2-oxoacids. The derivatisation of glyoxylate with *ortho*-phenylene diamine **64** (OPD), produces 2-quinoxalinol **65** which can be detected by fluorescence (λ_{ex} 340 nm, λ_{em} 420 nm) (scheme 3.3).



Scheme 3.3 Derivatisation of glyoxylate ***62*** *as fluorescent 2-quinoxalinol* ***65*** *with o-phenylene diamine* ***64****.*

The optimisation of the HPLC method was carried out using standard solutions of 2-quinoxalinol and derivatised glyoxylate. For clarity synthetic standards of 2-quinoxalinol are referred as 2-quinoxalinol and derivatised samples of glyoxylate, which also contain 2-quinoxalinol, are referred to as glyoxylate or derivatised glyoxylate. Aliquots of standard solutions (100 μL) of derivatised glyoxylate and 2-quinoxalinol were analysed by HPLC using fluorescent detection (λ_{ex} 340 nm, λ_{em} 420 nm). A range of concentrations of standards were used to construct calibration curves (section 6.5.3) for estimating the concentrations of 2-quinoxalinol and derivatised glyoxylate. Concentrations were determined from experiments carried out in duplicate. The following retention times were observed for standards: derivatised glyoxylate, R_t = 11.6 ± 0.5 min; 2-quinoxalinol, R_t = 11.6 ± 0.05 min (figure 3.11).

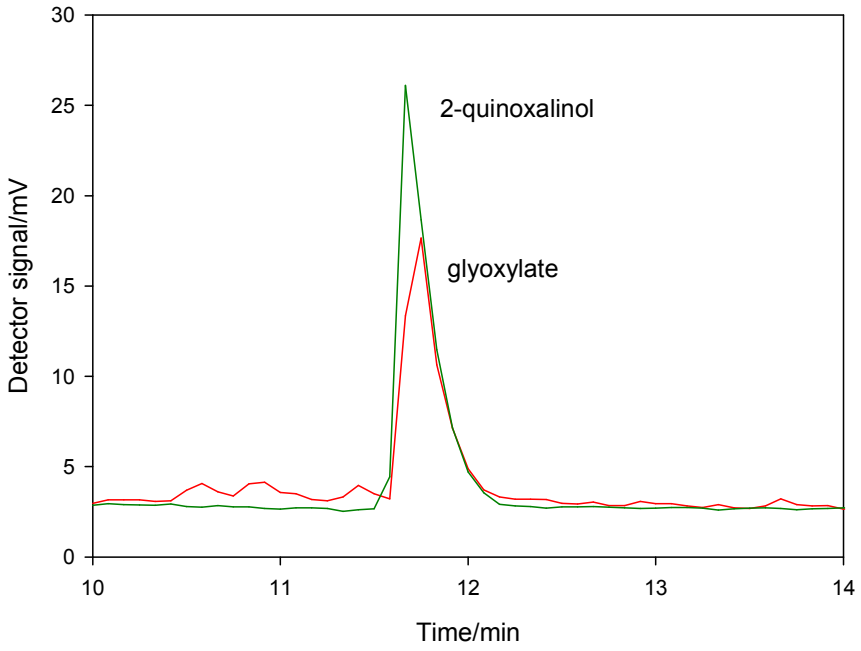


Figure 3.11 HPLC analysis with fluorescent detection of standard solutions of: derivatised glyoxylate (—), indicated as glyoxylate and 2-quinoxalinol (—).The chromatogram corresponds to 0.32 nmoles injected of each standard solution. The yield of derivatisation was 63.8% as estimated by the calibration curve from each standard.

The utilisation of substrates and the formation of products were measured by stopping assays a different time points. Thus, assay mixtures containing purified and reconstituted ThiGH, AdoMet, tyrosine and FldA/Fpr/NADPH, were incubated at 37 ºC for 0, 15, 30, 60 and 120 minutes. After the removal of protein components, supernatants were derivatised if necessary and then analysed by HPLC with fluorescence detection. From the analysis of the HPLC traces it is possible to see the increase in glyoxylate derivative through time (figure 3.12).

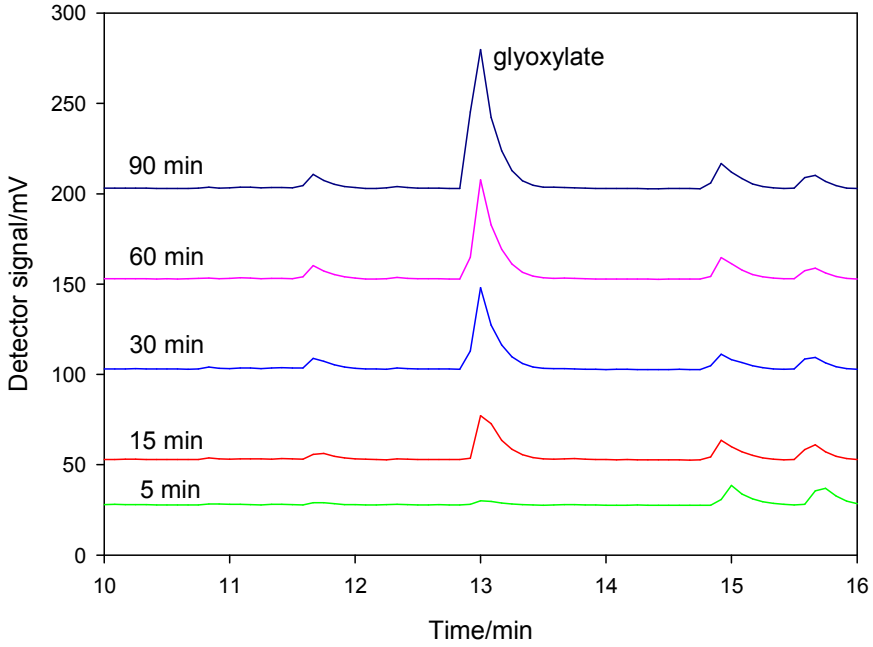


Figure 3.12 HPLC analysis with fluorescent detection of an in vitro assay time course. For clarity traces corresponding to incubation at 5, 15, 30, 60 and 90 minutes are offset by 25, 50, 100, 150, and 200 mV respectively.

For each measurement of glyoxylate in the *in vitro* assay, derivatised glyoxylate samples were also analysed. The derivatisation of the assay was always carried out in parallel with the derivatisation of the standards to ensure that the same experimental conditions were maintained and that the estimated concentrations were rigorous. In addition, standard samples of 2-quinoxalinol were analysed in parallel with both assay and glyoxylate standards in order to calculate the yield of derivatisation for each reaction.

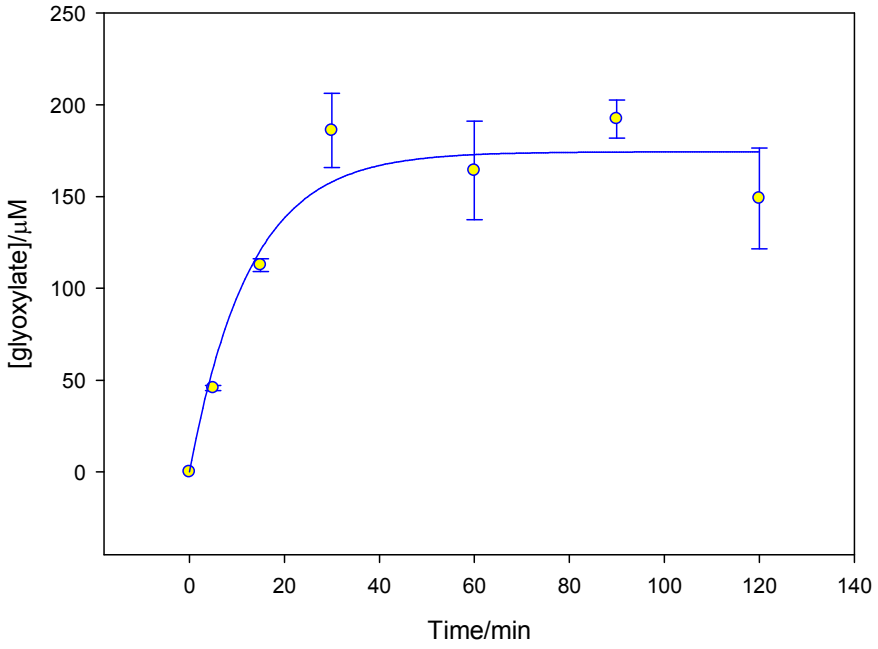


Figure 3.13 Time course formation of glyoxylate in a ThiGH assay. The curve shown here is the exponential growth and does not relate to the rate of the reaction. Error bars are derived from standard deviation of two data sets (table 3.6).

Table 3.6 Time course of glyoxylate formation. SD is standard deviation.

Time/min	[glyoxylate]/μM			
	sample 1	sample 2	average	SD
5	44.6	46.7	45.6	1.4
15	115.0	110.1	112.6	3.5
30	171.6	200.3	186.0	20.2
60	183.1	145.1	164.0	26.9
90	199.4	184.8	192.2	10.4
120	168.3	129.6	148.9	27.4

In order to investigate the kinetics and stoichiometry of the product formation, the glyoxylate and the *p*-cresol (table 3.7) formed in this assay were measured and plotted together (figure 3.14). The amounts of *p*-cresol and glyoxylate produced during a time course experiment were calculated by comparison with standards of known concentrations measured by HPLC.

Table 3.7 Time course of p-cresol formation. SD is standard deviation.

Time/min	[*p*-cresol]/μM			
	sample 1	sample 2	average	SD
5	23.1	23.4	23.3	0.17
15	56.5	79.1	67.7	15.9
30	126.5	115.9	121.2	7.52
60	167.7	147.7	157.7	14.2
90	162.1	169.2	165.6	5.04
120	162.1	182.3	172.2	14.3

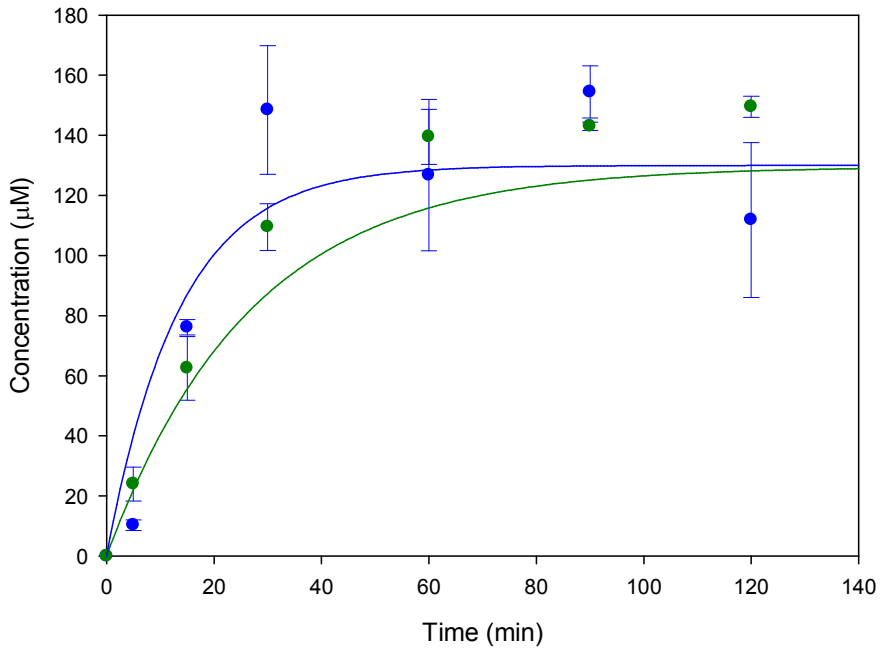


Figure 3.14 Time course formation of glyoxylate and p-cresol in a ThiGH assay. Error bars are derived from standard deviation of two data sets (table 3.4 and 3.5). Glyoxylate (—●), p-cresol (—●). The curves are fit to a first order exponential function.

The apparent first order rate constants of *p*-cresol and glyoxylate formation were calculated. Data from figure 3.14 were fitted, using *Origin v 6.0 analysis software*, to a first-order process to determine and calculate the final concentrations of products and the apparent first order rate constant (equation 6.1). Tyrosine cleavage reaction produces *p*-cresol and glyoxylate in an approximate 1:1 ratio over the time course and approached completion after 1 hour (table 3.8). Results are shown with associated standard errors.

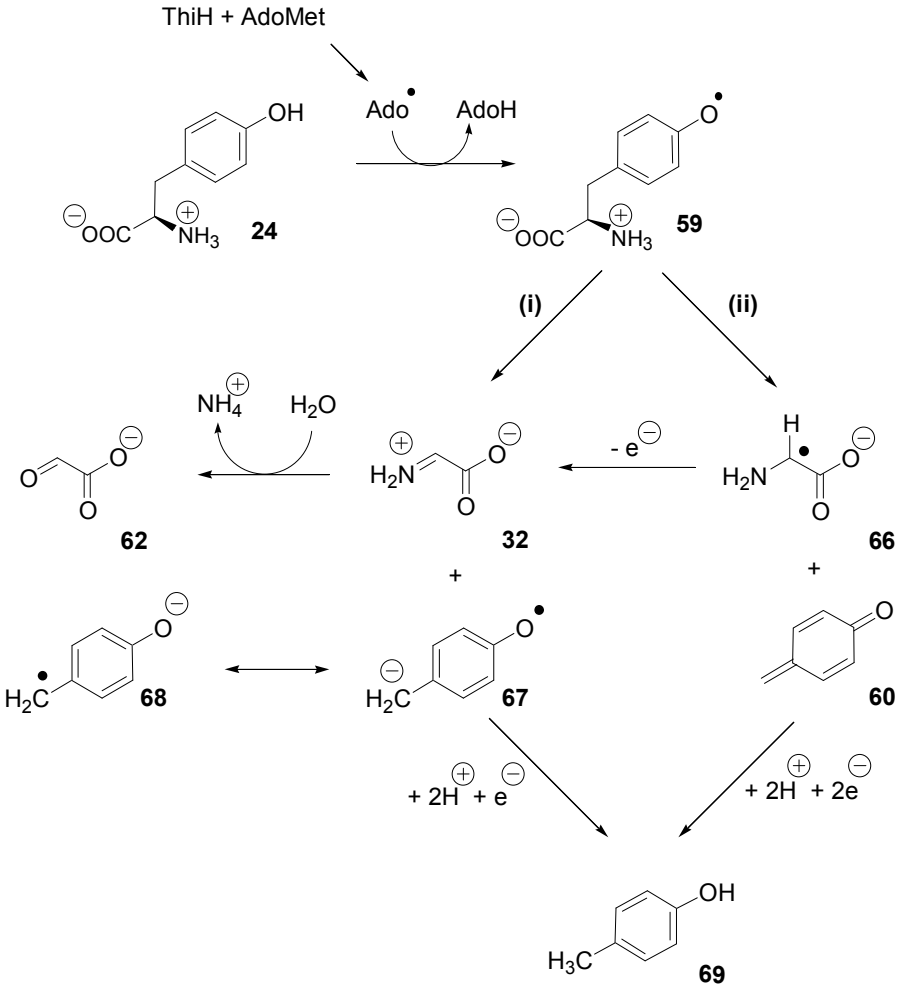
Table 3.8 Kinetic parameters for product formation.

Product	Rate constant [10^{-2}min^{-1}]	Final concentration [μM]
p-cresol	4.6 ± 0.7	176.2 ± 7.8
glyoxylate	6.3 ± 1.1	171.6 ± 7.9

3.3 Novel proposed mechanism for Thz-P biosynthesis

The investigations by Dr. M. Kriek provided significant evidence about the intermediate and by-product derived from tyrosine cleavage[166]. These results together with the HPLC studies demonstrate unequivocally that *p*-cresol and glyoxylate are produced from tyrosine during the ThiGH *in vitro* assay. This allowed the proposal of a novel mechanism for tyrosine cleavage in *E. coli* (scheme 3.4). Reductive cleavage of AdoMet generates the highly reactive 5'-deoxyadenosyl radical, which can abstract a hydrogen atom from tyrosine **24**. The resulting radical **59** can react via two pathways: (i) heterolytic pathway leads to dehydroglycine **32** and the stabilised radical anion **67** ↔ **68**; (ii) homolytic cleavage of the Cα-Cβ bond releases the radical intermediate **66** together with quinone methide **60**. To form *p*-cresol and dehydroglycine by either of these pathways requires further electron transfer steps, which can be possible if any of these intermediates is in close proximity to the [4Fe-4S] cluster in the active site of

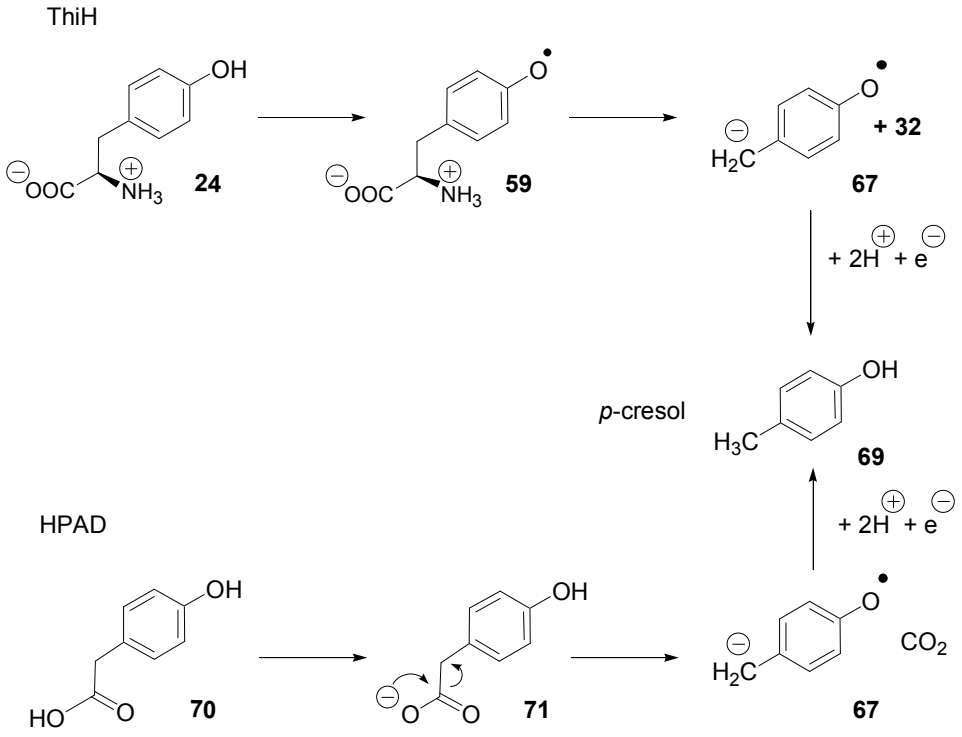
ThiH. DFT and *ab initio* calculations carried out by Dr. A. Croft[166], suggest that the homolytic mechanism [(ii), scheme 3.4] is thermodynamically more favourable. Increase of polarity improves the relative stability of the heterolytic pathway, but not enough for it to be identified as the preferred pathway. Further experimental data is required to elucidate this step.



Scheme 3.4 Novel proposed mechanism for tyrosine cleavage in E. coli[166].

The mechanism of tyrosine cleavage by ThiH shows similarities to the mechanism catalysed by the glycyl radical enzyme *p*-hydroxyphenylacetate decarboxylase (HPAD) from *Clostridium difficile*[170-172]. This organism produces *p*-cresol by the decarboxylation of *p*-hydroxyphenylacetate by HPAD. It seems likely that the reactions catalysed by

ThiH and HPAD use different carbon centred radicals to initiate mechanistically related cleavage reactions (scheme 3.5).



Scheme 3.5 Formation of p-cresol by ThiH and HPDA.

3.4 Summary and conclusions

The development of an *in vitro* assay, containing chemically reconstituted ThiGH, AdoMet, tyrosine and a reducing system (NADPH/Fld/Fpr), allowed the identification of the reaction products formed upon tyrosine cleavage. The aromatic side chain released upon tyrosine cleavage, initially thought to be *p*-hydroxybenzyl alcohol[113], was identified as *p*-cresol. The remaining fragment was hypothesised to be dehydroglycine. This intermediate hydrolyses to produce glyoxylate in the absence of Dxp and a sulfur donor, the products required to biosynthesise Thz-P. This intermediate was also detected.

The development of an HPLC method using UV-visible detection further confirmed the presence of *p*-cresol and also allowed for its quantification, the retention time was

found to be 16.7 ± 0.5 min. In addition, this method detected AdoH (R_t= 11.8 ± 0.4 min), the product of AdoMet reductive cleavage by the [4Fe-4S] cluster; and tyrosine consumption was monitored. Tyrosine (initial concentration of 160 μM) was fully consumed after 60 min, and 133 μM of *p*-cresol were formed. AdoH was also detected and although the maximum expected concentration was 160 μM, assuming one molar equivalent was generated per mol of tyrosine, the amount of AdoH was higher and reached a maximum of 186 μM, suggesting some uncoupled turnover. The reaction achieved completion after 60 min.

The intermediate of the tyrosine cleavage reaction was also detected as glyoxylate. Detection and quantification of glyoxylate in the *in vitro* assays was possible using a method devised for the detection of 2-oxoacids, in which glyoxylate is converted to the fluorescent 2-quinoxalinol derivative. The retention times were found to be R_t = 11.6 ± 0.5 min for derivatised glyoxylate, and R_t = 11.6 ± 0.05 min for 2-quinoxalinol.

The development of these two HPLC methods allowed the investigation of the kinetics and stoichiometry for the formation of the reaction products formed upon tyrosine cleavage. In addition the rate of consumption of tyrosine was also calculated and final concentration found to be 0.0 μM as it was expected. Uncoupled turnover of AdoMet was confirmed by these calculations. The rate of formation of AdoH seems to be higher (0.102 ± 0.033 min^{-1}) than the rate of tyrosine consumption (0.052 ± 0.003 min^{-1}). In another *in vitro* assay the rates of formation and stoichiometry of *p*-cresol and glyoxylate were calculated. The products are formed in an approximate 1:1 ratio over time (176.2 ± 7.8 μM *p*-cresol, 171.6 ± 7.9 μM glyoxylate), and the reaction is shown to achieve completion after 60 min, results that are in accord with the previous experiments. The fitted first order rate constants of formation were found to be 4.6 ± 0.7 [10^{-2} min^{-1}] for *p*-cresol, and 6.3 ± 1.1 [10^{-2} min^{-1}] for glyoxylate. The rates of formation of *p*-cresol seem to be approximately the same in different experiments.

CHAPTER 4

Expression of Microbial ThiH Genes in *Escherichia coli*

4.1 Introduction

One of the main goals of this project is the isolation of active *holo*-ThiH. The optimisation of the expression, purification and chemical reconstitution of ThiH from *E. coli* were described in detail in chapter 2. While these studies were in progress, other methods for the expression of *holo*-ThiH were being carried out in parallel, in particular the expression from different organisms. These studies are described in this chapter.

The production of ThiH from different microbes can be especially valuable because whilst the active site of a protein might be conserved throughout different species, the exterior structure may vary; the surface residues of the protein play an important role in protein-protein interactions[173]. This may be of particular importance in crystallisation studies, as during crystallisation these protein-protein interactions and the aggregation state of the protein are critical. While surveying for microorganisms containing the thiamine biosynthetic operon two factors were taken in account, the operon structure and the thermostability of the organism.

The operon structure can vary greatly across different types of organisms and this diversity affects the regulation of gene expression[174]. During evolution, organisms are thought to adapt in order to be more efficient in their biosynthetic functions. The expression of ThiH from organisms containing a different thiamine operon structure might produce protein with different characteristics.

The stability of an individual protein is greatly influenced by the thermophilicity of the source microorganism[175-177]. Hyperthermophiles grow optimally at temperatures

between 80 and 110 ºC. Enzymes from these organisms have developed unique structure-function properties of high thermostability and high activity at temperatures above 70 °C. A number of hyperthermophilic enzymes have been identified and characterised[178-181]. Enzymes from hyperthermophilic organisms can be efficiently purified after expression in a mesophilic host, such as *E. coli*. Studies on lipoyl synthase (LipA), a 'radical AdoMet' enzyme, by M. Kriek (unpublished results) showed that *Sulfolobus solfataricus* LipA is far more stable then *E. coli* LipA, which has a marked tendency to precipitate (unpublished results). These observations reinforce the theory that ThiH expressed from a hyperthermophile might be more stable, since *E. coli* presents similar stability characteristics to *E. coli* LipA.

Thermophilic enzymes are often considered to be advantageous for crystallisation studies, when compared to their mesophilic orthologs, due to their intrinsic stability and high solubility[176]. Moreover, due to their weak catalytic activity at low temperatures, enzyme-substrate complexes may have a sufficiently high lifetime for X-ray analysis[181]. The X-ray structures of a variety of enzymes expressed from hyperthermophilic organisms has been determined. Examples include the proteins from the histidine biosynthetic pathway such as the cyclase of imidazolglycerol phosphate synthase and N'-((5'-phosphoribosyl)-formimino)-5-aminoimidazol-4-carboxamid ribonucleotide isomerase, both derived from *Thermotoga maritima*[181]; the iron hydrogenase maturase HydE from *T. maritima*[182] and the IscU scaffold protein from the hyperthermophile *Aquifex aeolicus*[183]. Extending this logic to proteins from thiamine biosynthesis, the isolation of stable *holo*-ThiH from a hyperthermophile might assist in the determination of the crystal structure.

4.2 Results and discussion

4.2.1 Selection of microbial ThiH genes of interest

The structural motifs and elements of proteins are conserved in the genomic encoded sequence and across different genomes. The amino acid sequence of an enzyme can provide valuable mechanistic information and conserved catalytic residues; it can also reveal evolutionary relationships between enzymes, which can be revealed by a sequence alignment[184]. Therefore, a survey was carried out for organisms containing the thiamine biosynthetic operon (table 4.1). The sequence similarity between ThiH's was investigated through alignment of the sequences. All sequences used are listed in Appendix C. Sequence alignment allows the comparison of the degree of similarity between amino acids that occupy a particular position in the sequence of the protein.

Table 4.1 Microorganisms containing the thiH gene.

Organism	*Operon structure**	*Growth temperature*	*Respiration*
Escherichia coli	*CEFSGH*	37 ºC	facultative aerobe
Vibrio vulnificus	*CEFSGH*	43 ºC	facultative aerobe
Desulfovibrio vulgaris	*SGHFE*	37 ºC	obligate anaerobe
Thermotoga maritima	*H*	80 ºC	obligate anaerobe
Moorella thermoacetica	*H*	57 ºC	obligate anaerobe
Chlorobium tepidum	*SGHF*	47 ºC	obligate anaerobe
Campylobacter jejuni	*SFGH*	42 ºC	microaerophilic

* operon structure: describes the sequence of thiamine biosynthetic genes found in the genome of each organism. The genes reported here are those identified in the BLAST Genome Taxa Query Page[185].

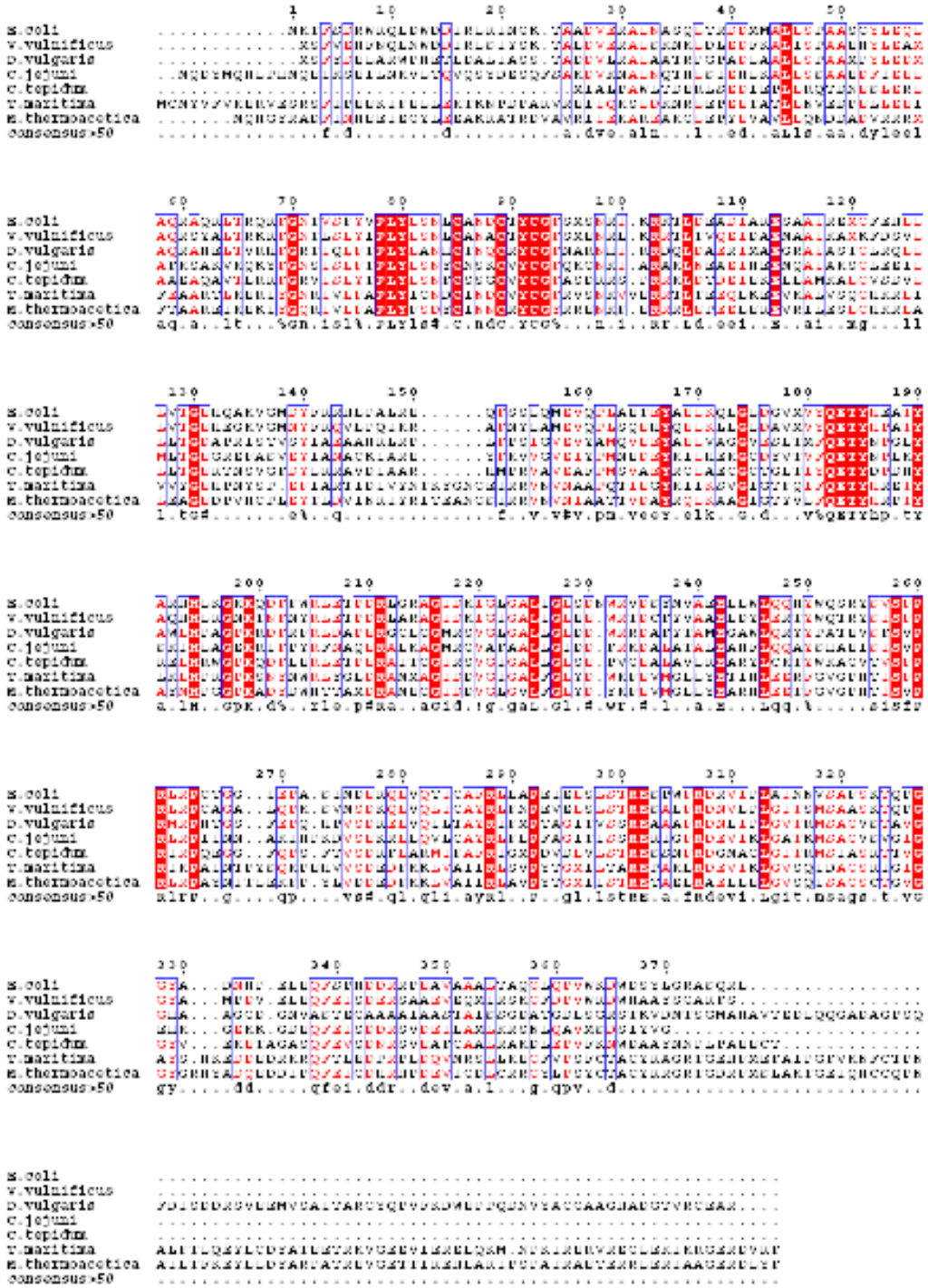


Figure 4.1 Sequence alignment of E. coli ThiH with sequences from V. vulnificus, D. vulgaris, C. jejuni, C. tepidum, T. maritima and M. thermoacetica.

Table 4.2 Partial view of the alignment of sequence of ThiH from E. coli with V. vulnificus, D. vulgaris, C. jejuni, C. tepidum, T. maritima and M. thermoacetica. Cysteine residues of the CXXXCXXC motif are in blue.

Organism	Residues	Sequence
E. coli	83-94	N L **C** A N D **C** T Y **C** G F
V. vulnificus	82-93	N L **C** A N A **C** T Y **C** G F
D. vulgaris	82-93	N H **C** T N Q **C** R Y **C** G F
T. maritima	95-106	N D **C** I N D **C** V Y **C** G F
M. thermoacetica	88-99	D Y **C** I N N **C** R Y **C** G Y
C. tepidum	57-68	N F **C** S S G **C** V Y **C** G F
C. jejuni	93-104	N Y **C** N S K **C** V Y **C** G F

The sequence alignment of the ThiH shows that all selected organisms contain a conserved CXXXCXXC motif, characteristic of the 'radical AdoMet' family (section 1.4).

Studies of expression and purification of ThiH were initiated using *D. vulgaris* (difference in operon structure) and *T. maritima* (hyperthermophile) and are described in the following sections.

4.2.2 Expression and purification of *Desulfovibrio vulgaris* ThiH

Desulfovibrio vulgaris is a gram-negative, anaerobic sulfur reducing bacteria that can be found in soil, animal intestines and fresh salt water[186]. The most commonly used strain, *D. vulgaris* Hildenborough, was discovered in clay soil near Hildenborough, Kent, United Kingdom, and is used as a model organism for studying energy metabolism of sulfate-reducing bacteria and bio-corrosion of metal[187]. *D. vulgaris* has an optimal growth temperature of 37 ºC and contains a *thiSGH* gene (ORF DVU2093).

Primer design and plasmid assembly

D. vulgaris thiSGH (2.381 kb) was amplified by polymerase chain reaction (PCR) using genomic DNA and two different primers (table 6.10, section 6.6.1). The PCR product (*thiSGH*) was restricted using *PciI* and *SacI* and cloned into pRL1020 on a *NcoI*/*SacI* fragment (method 6.6.1). The resultant vector was named pFM001 and the steps describing the amplification of the DNA segment as well as all the methods involved in the assembly of the plasmid are described in detail in section 6.6.1.

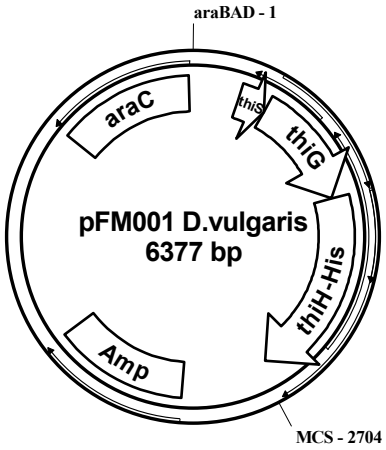


Figure 4.2 Map of pFM001, encoding for D. vulgaris thiSGH-His. MCS: multiple cloning site. araBAD: operon promoter. Image created using pDraw32 DNA analysis software v1.0 [145].

A second plasmid additionally containing the *E. coli isc* operon (encoding *iscSUA-hscAB-fdx*) downstream from the *thiSGH* operon, was assembled as these proteins have been shown to improve expression of enzymes in the *holo-* form (method 6.6.1)[120-122]. The co-expression of *E. coli* ThiH with Isc proteins carried out using *E. coli* thiamine biosynthetic genes and is described in section 2.2.2.

pRL1021, a plasmid bearing the *iscSUA-hscBA-fdx* operon, was restricted with *EcoRI*/*SacI* to yield the *iscSUA-hscBA-fdx* fragment which was inserted into similarly restricted pFM001 (method 6.6.1). The resulting plasmid was named pFM101 (figure 4.3).

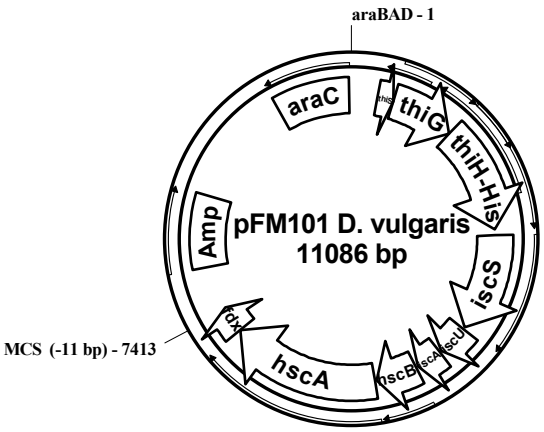


Figure 4.3 Map of pFM101, encoding for D. vulgaris thiSGH-His and E. coli iscSUA-hscBA-fdx. MCS: multiple cloning site. araBAD: operon promoter. Image created using pDraw32 DNA analysis software v1.0[145].

Expression and purification of D. vulgaris pFM001/BL21(DE3) and pFM101/BL21(DE3)

pFM001 and pFM101 were transformed into *E. coli* BL21(DE3) cells and grown aerobically in 2YT medium (method 6.6.2). Table 4.3 shows a comparison between the yield of cells and protein obtained from *D. vulgaris* (pFM001 and pFM101) and *E. coli* (pRL1020 and pRL1021).

Table 4.3 Yields of cells and protein obtained from E. coli and D. vulgaris plasmids.

Microorganism	*Expression vector*	*Yield of cells (g/L)*	*Yield of protein (g/L)*
E. coli	pRL1020	7.6	5.2
E. coli	pRL1021	7	3.5
D. vulgaris	pFM001	8	0.5
D. vulgaris	pFM101	5	1.3

The yield of cells obtained from *D. vulgaris* expression plasmids (pFM001, pFM101), was very similar to the yield obtained using pRl1020 and pRl1021.

The His-tagged proteins were purified from pFM001/BL21(DE3) and pFM101/BL21(DE3) (method 13) and SDS-PAGE analysis of the eluted fractions (method 8) shows that *D. vulgaris* ThiH-His co-elutes in a complex with ThiG, in a manner directly analogous to that observed for the *E. coli* proteins (figure 4.4 and 4.5). The predicted molecular weight of *D. vulgaris* thiamine proteins is higher then that of proteins derived from *E. coli* (table 4.4).

Table 4.4 Predicted molecular weight for thiamine biosynthetic proteins in E. coli and D. vulgaris.

	Proteins	*MW/kDa*
Escherichia coli	ThiH	43.0
	ThiG	26.9
	ThiS	7.3
	ThiF	27.0
Desulfovibrio vulgaris	ThiH	48.0
	ThiG	27.6
	ThiS	7.0
	ThiF	22.3

The yield of soluble ThiGH obtained from *D. vulgaris* expression plasmids was very low, comparing to that obtained from *E. coli* expression system (pRL1020 and pRL1021) (table 4.3).

E. coli can be a highly efficient host for recombinant protein expression and is certainly the most studied and optimised. It is widely used for protein expression of a number of genes from different organisms. Nevertheless, it has been observed that the expressions of foreign genes may interfere with the survival of *E. coli* cells and significantly affect the expression capability[188-190].

Therefore, an explanation for the low level of expression might be found in the codon bias of *D. vulgaris*. *E. coli* contains a number of low usage codons (listed in table 4.5),

when expressing genes from different organisms in *E. coli*, less tRNA is available to transfer the specific amino acids, limiting the rate of protein synthesis, which is observed as low soluble expression level of the protein[191]. Modifications of culture conditions (e.g. low growth temperature, changes in media composition, etc.) could potentially shift the codon usage bias enough to alleviate some codon usage-based expression problems. It has been reported however that the levels of most of the tRNA isoacceptors corresponding to rare codons remain unchanged at different growth rates[191,192]. Table 4.5 lists the frequency of usage of rare codons in *D. vulgaris*, showing that this organism uses some of the rare codons present in *E. coli*, in particular arginine (codons: AGG and CGG). A complete list of the codon bias is listed in Appendix C. A method that can potentially overcome this problem is the use of a Rosetta host strain. This strain has been shown to enhance protein synthesis[191].

Table 4.5 Usage of rare codons in D. vulgaris thiH.

Amino acid	*Codon*	*Frequency per thousand*	*number*
Arg	AGG	28.4	13
Arg	AGA	10.9	5
Arg	CGG	52.5	24
Arg	CGA	10.9	5
Gly	GGA	4.4	2
Ile	AUA	4.4	2
Leu	CUA	4.4	2
Pro	CCC	19.7	9

The strength of the promoter is another factor that might have affected the protein expression. Studies by R. Leonardi[58] using *E. coli* genes showed that pBAD derived plasmids expressed higher level of soluble ThiGH, comparing to pET derivatives (described in detail in section 2.2)[58]. Nevertheless, most studies using *D. vulgaris* genes, expressed in *E. coli* host, suggest that T7 RNA polymerase (pET vectors) is a more efficient promoter[190,193,194].

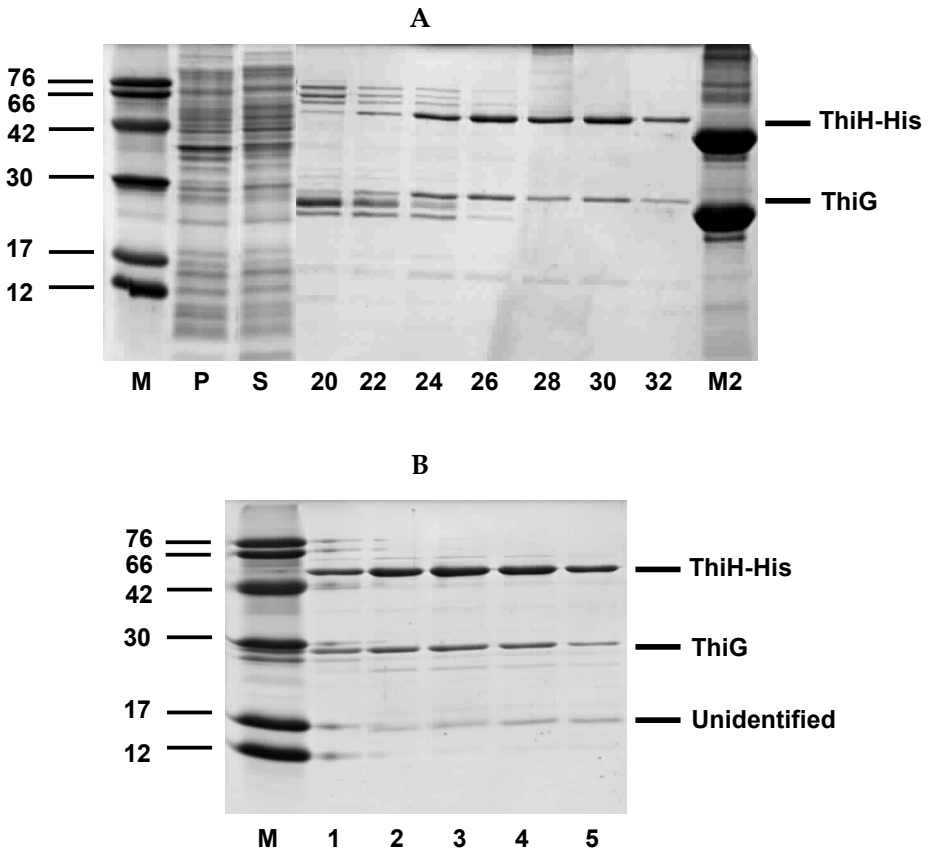


Figure 4.4 Purification of ThiGH-His from D. vulgaris pFM001/BL21(DE3). Coomassie Blue stained 15% SDS-PAGE gels of proteins eluted from A: nickel-chelating column and B: gel filtration column. P: re-suspended pellet; S: supernatant; M: molecular weight marker; M2: purified ThiGH-His from pRL1020/BL21 (DE3). Fractions 26-32 from the nickel-chelating column were pooled and gel filtered. The numbers correspond to the fractions eluted from the column.

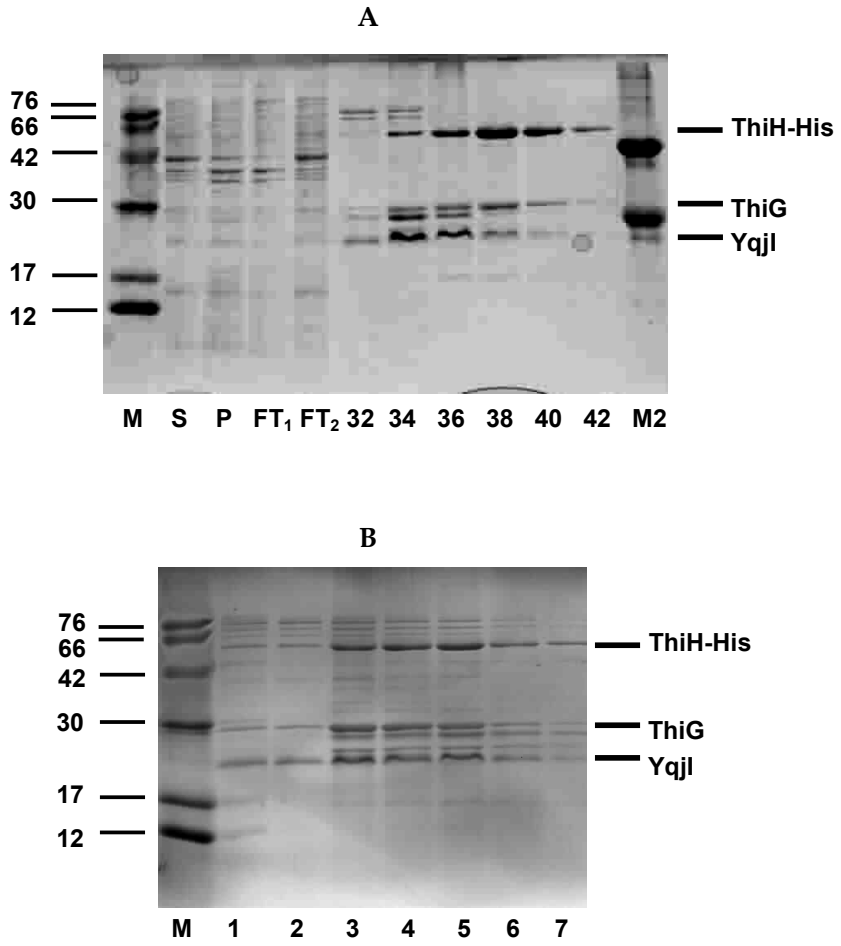


Figure 4.5 Purification of ThiGH-His from D. vulgaris pFM101/BL21(DE3). Coomassie Blue stained 15% SDS-PAGE gels of proteins eluted from A: nickel-chelating column and B: gel filtration column. P: re-suspended pellet; S: supernatant; M: molecular weight marker; M2: purified ThiGH-His from pRL1020/BL21 (DE3). Fractions 34-40 from the nickel-chelating column were pooled and gel filtered. The numbers correspond to the fractions eluted from the column.

The iron content of the purified ThiGH samples was determined by the method of Fish (method 14). An increase in iron content was observed in purified samples of ThiGH expressed from *D. vulgaris*: pFM001 yield 2 mol of Fe per mol of ThiH and pFM101 yield 4.5 mol of Fe per mol of ThiH. However, as it was observed with other plasmids containing the *isc* operon (section 2.2.2), ThiGH co-eluted with YqJI (23 kDa) (figure 4.5) and the presence of this protein has an effect on the overall iron measured from

these samples. Nevertheless, the increase of iron observed in ThiGH samples expressed from pFM001, suggests that the expression of ThiGH from this organism might result in a high proportion of *holo*-enzyme. Future expression studies are required to optimise the expression of ThiGH from *D. vulgaris* and determine if this expression system is a better source of *holo* protein.

4.2.3 Expression and purification of *Thermotoga maritima* ThiH

Thermotoga maritima is a rod-shaped bacteria belonging to the order Thermotogales, it can be found in heated sea floors and volcanoes. *T. maritima* was originally isolated at Volcano, Italy but it can also be found in the Portuguese island of Azores; it has an optimal growth temperature of 80 ºC. This organism metabolises many simple and complex carbohydrates and has potential as a renewable energy source[177,186]. Because *T. maritima* contains only a single *thiH* gene (ORF TM1267) and non of the other thiamine biosynthetic genes, there was a possibility that the assignment of TM1267 to be *thiH* was incorrect.

Primer design and plasmid assembly

T. maritima thiH (1.4 kb) was amplified by PCR (method 1), using genomic DNA and two primers (table 1.12, section 6.6.4). The PCR product (*thiH*) was restricted with *NcoI*/*SalI* and cloned into pBAD-HisA; on an *NcoI*/*SacI* fragment. The steps describing the amplification of the DNA segment as well as all the methods involved in the assembly of the plasmid are described in detail in method 6.6.4. The resulting plasmid was named pFM003 (figure 4.6).

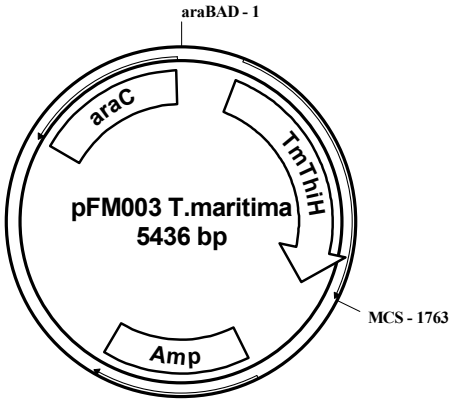


Figure 4.6 Map of pFM003, encoding for T. maritima thiH-His. MCS = multiple cloning site. araBAD = operon promoter. Image created using pDraw32 DNA analysis software v1.0[145].

A second plasmid containing the *isc* operon (encoding *iscSUA-hscAB-fdx*) downstream from the *thiH* operon, was assembled. During the planning of the assembly of this plasmid it was noticed that the *T. maritima* PCR product contained a *XhoI* restriction site (figure 4.7). This represented a problem since the *iscSUA-hscBA-fdx* operon is typically removed using the *XhoI*/*EcoRI* restriction sites. Another option to this method was the ligation of the PCR product, restricted with *NcoI*/*SalI*, to a plasmid restricted on a *NcoI*/*XhoI* site, since *SalI* complements the *XhoI* site (figure 4.7). pMK024, a plasmid bearing the *isc* operon, contained two *NcoI* sites, one of which would make impossible the correct restriction. Therefore mutagenesis on the *NcoI* site of interest was carried out. The new plasmid was named pFM024 (figure 4.7) (method 6.6.3). This made possible the restriction of pFM024 with *NcoI*/*XhoI* and insertion of the *NcoI*/*SalI* restricted PCR product. This resulting plasmid was named pFM103, and due to the ligation of *SalI* and *XhoI* sites, result in the subtraction annulment of a restriction site (figure 4.7).

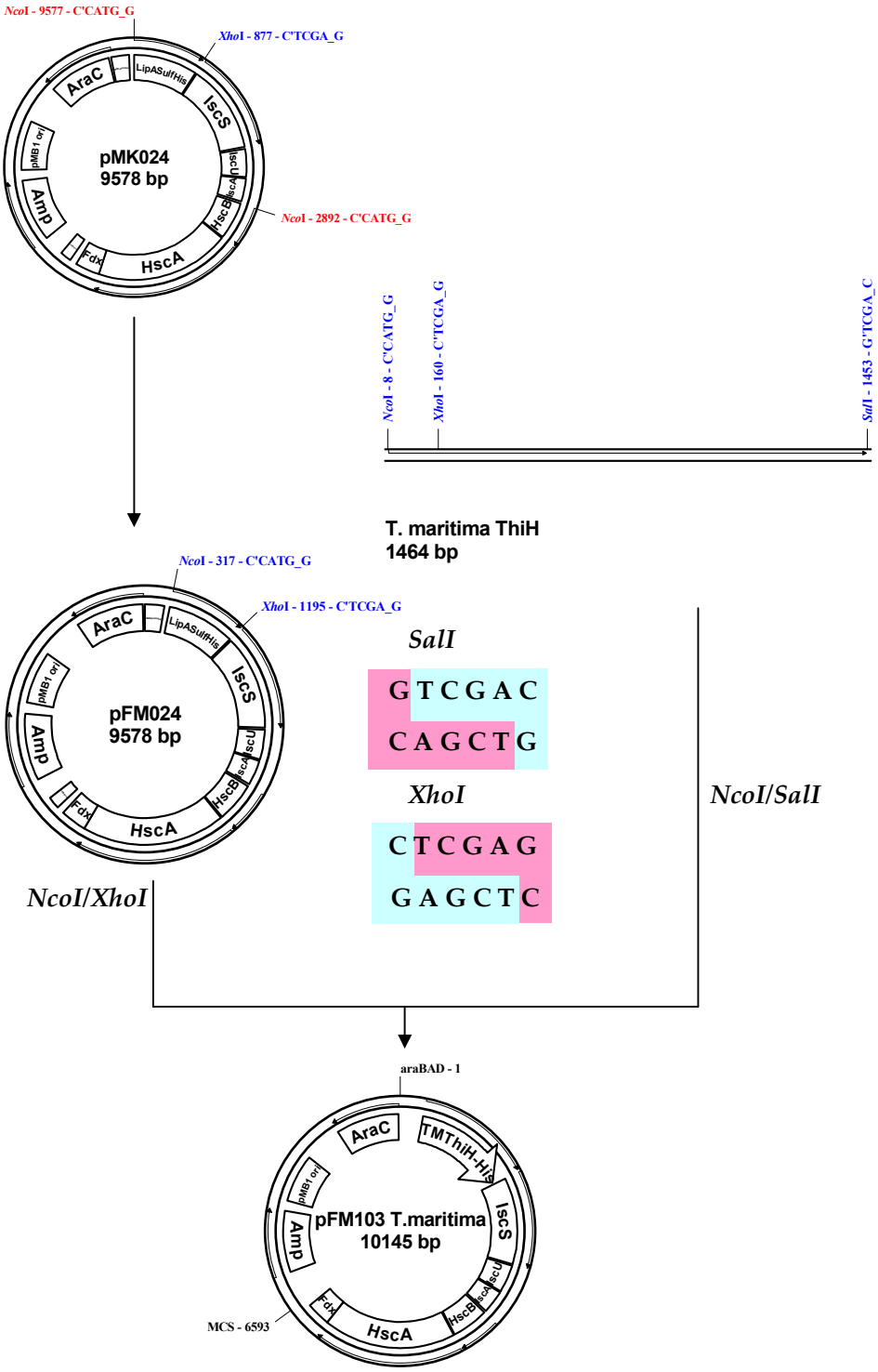


Figure 4.7 Assembly of pFM103 encoding for T. maritima thiH-His and E.coli iscSUA-hscAB-fdx. MCS = multiple cloning site. araBAD = operon promoter. The figure shows the mutagenesis of the NcoI site of pMK024 to give pFM024 and the sequence of the restriction sites. Image created using pDraw32 DNA analysis software v1.0[145].

Expression and purification of T. maritima pFM003/BL21(DE3) and pFM103/BL21(DE3)

pFM003 and pFM103 were transformed into *E. coli* BL21(DE3) cells and grown aerobically in 2YT medium (method 6.6.5). A comparison between the yield of cells and protein obtained from *T. maritima* (pFM003 and pFM103) and *E. coli* (pRL1020 and pRL1021) is shown in table 4.6. The yield of cells obtained from *T. maritima* expression plasmids (pFM003, pFM103), was very similar to the yield obtained using pRL1020 and pRL1021.

Table 4.6 Yields of cells and protein obtained from E. coli and T. maritima plasmids.

Microorganism	*Expression vector*	*Yield of cells (g/L)*	*Yield of protein (g/L)*
E. coli	pRL1020	7.6	5.2
E. coli	pRL1021	7.0	3.5
T. maritima	pFM003	6.0	1.2
T. maritima	pFM103	6.4	1.0

pFM003/BL21(DE3) and pFM103/BL21(DE3) were purified (method 13) and SDS-PAGE analysis of the eluted fractions (method 8) shows that ThiH-His co-elutes with a 28 kDa protein, assumed to be *E. coli* ThiG (figure 4.8 and 4.9). The molecular weight of *T. maritima* ThiH-His is 55.6 kDa.

The iron content of the purified ThiH samples was measured (method 14) and an increase was observed; pFM003 yield 2 mol of Fe per mol of ThiH and pFM103 yield 4 mol of Fe per mol of ThiH. However, as it was observed with plasmids containing the *isc* operon (section 2.2.2), ThiH co-eluted with YqJI (23 kDa) (figure 2.2.2). The presence of this protein has an effect on the overall iron measured from these samples.

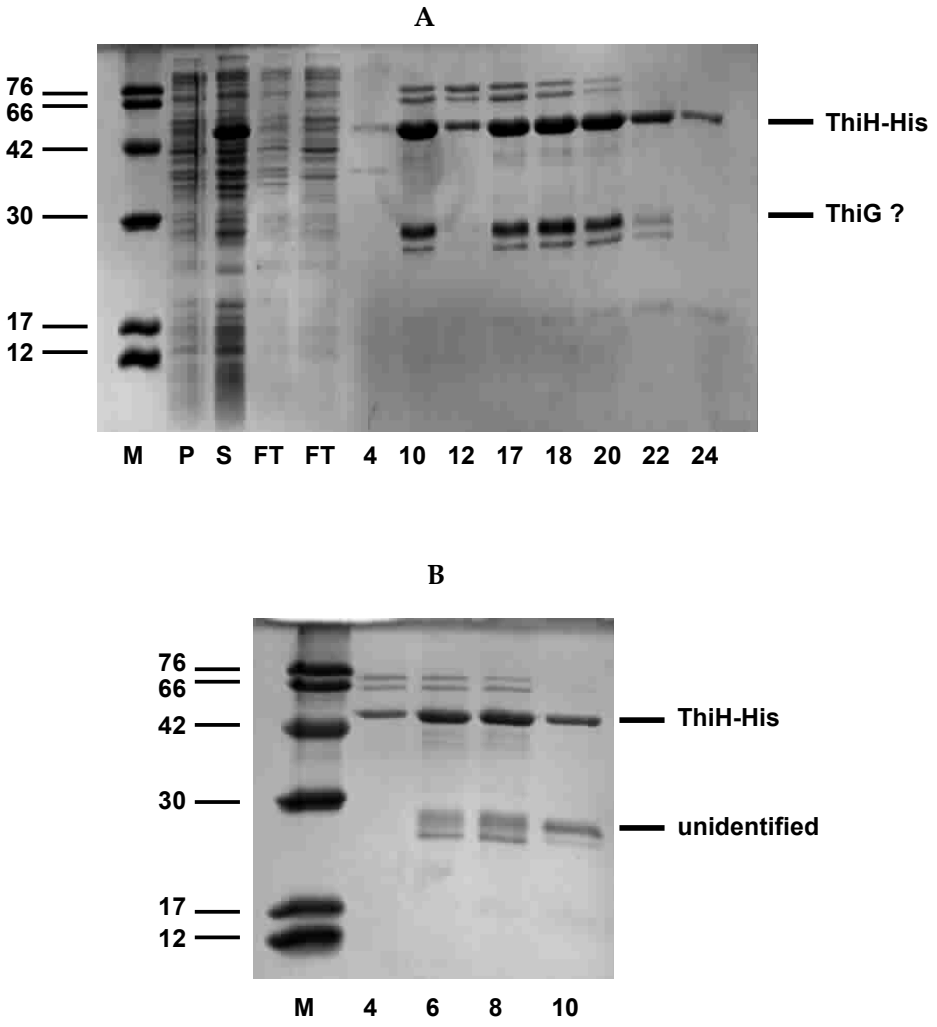


Figure 4.8 Purification of ThiH-His from T. maritima pFM003/BL21(DE3). Coomassie Blue stained 15% SDS-PAGE gels of proteins eluted from A: nickel-chelating column and B: gel filtration column. P: re-suspended pellet; S: supernatant; FT: Flow through; M: molecular weight marker. Fractions 16-20 from the nickel-chelating column were pooled and gel filtered. The numbers correspond to the fractions eluted from the column. The protein band at ~28kDa labelled as unidentified corresponds to the expected molecular weight of ThiG.

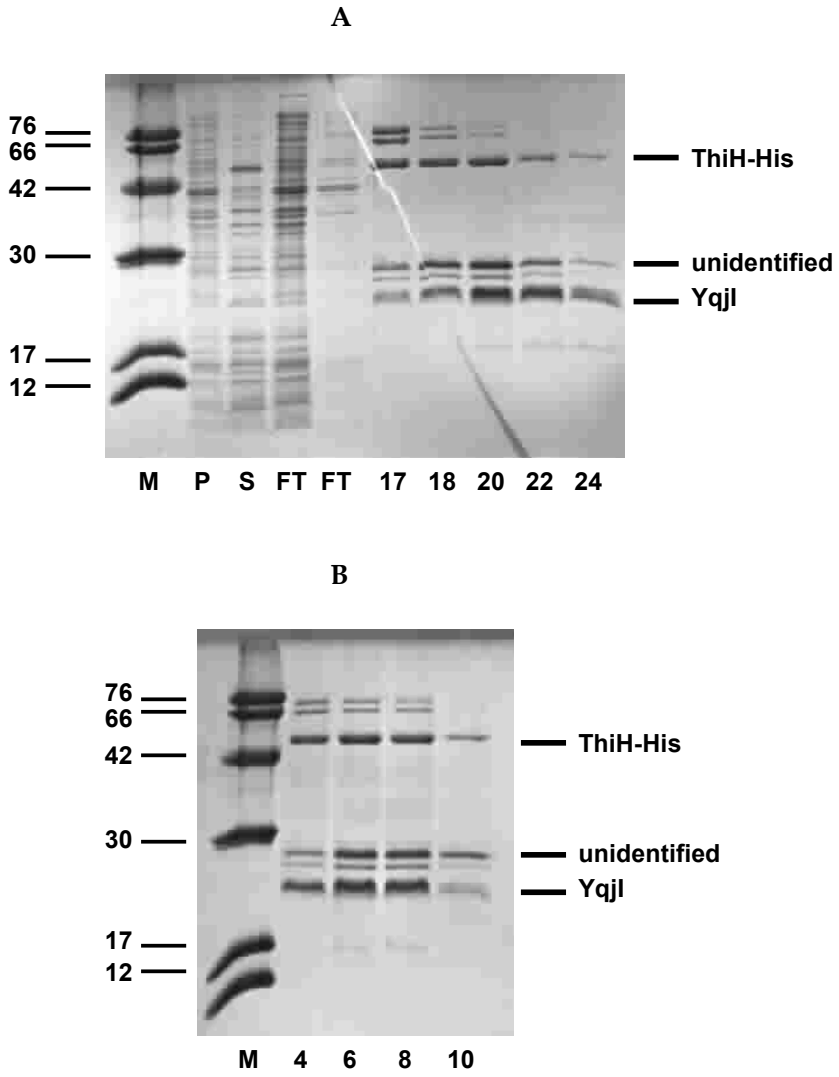


Figure 4.9 Purification of ThiH-His from T. maritima pFM103/BL21(DE3). Coomassie Blue stained 15% SDS-PAGE gels of proteins eluted from A: nickel-chelating column and B: gel filtration column. P: re-suspended pellet; S: supernatant; FT: Flow through; M: molecular weight marker. Fractions 17-22 from the nickel-chelating column were pooled and gel filtered. The numbers correspond to the fractions eluted from the column.

The co-elution of what appears to be ThiG with ThiH (figure 4.8 and 4.9) was confusing due to the absence of a *thiG* gene in the expression plasmids. An explanation for this might be that ThiH associates with the gene encoding for ThiG present in the *E. coli* chromosome during cell growth and protein expression.

The observed insolubility of ThiH might be caused by the inadequate choice of promoter and/or expression host (*E. coli*). Studies using hyperthermophiles, have shown that the use of stronger promoters, such as plac, ptac, or T7 RNA polymerase, results in higher levels of expression when using host *E. coli* cells. Another difficulty encountered in expressing archeal genes in *E. coli* is the difference of codon usage between *E. coli* and the expressed gene, similar to the *D. vulgaris* experiments. Hyperthermophiles in particular tend to adopt different strategies for stability of the proteins amongst them are: CG-rich codons; ratio of charged amino acids compared to uncharged amino acids; ionic interactions and amino acid preferences[175]. Table 4.7 lists the frequency of usage of rare codons in *T. maritima*, showing that this organism uses arginine, glycine and isoleucine codons with significantly higher frequency. A complete list of codons is listed in Appendix C. A method that can potentially overcome this problem is the use of a Rosetta host strain[191].

Table 4.7 Rare codon usage in T. maritima thiH.

Amino acid	*Codon*	*Frequency per thousand*	*Number*
Arg	AGG	21.1	10
Arg	AGA	52.7	25
Arg	CGG	0	0
Arg	CGA	0	0
Gly	GGA	31.6	15
Ile	AUA	29.5	14
Leu	CUA	2.1	1
Pro	CCC	10.5	5

4.3 Relationship between ThiH and HydG

It was later noticed that the *T. maritima* genomic DNA (ORF TM1269) used in our studies and originally assigned in the databases as ThiH[195] was identified by Rubach and co-workers[196] to be the iron only hydrogenase maturation enzyme HydG.

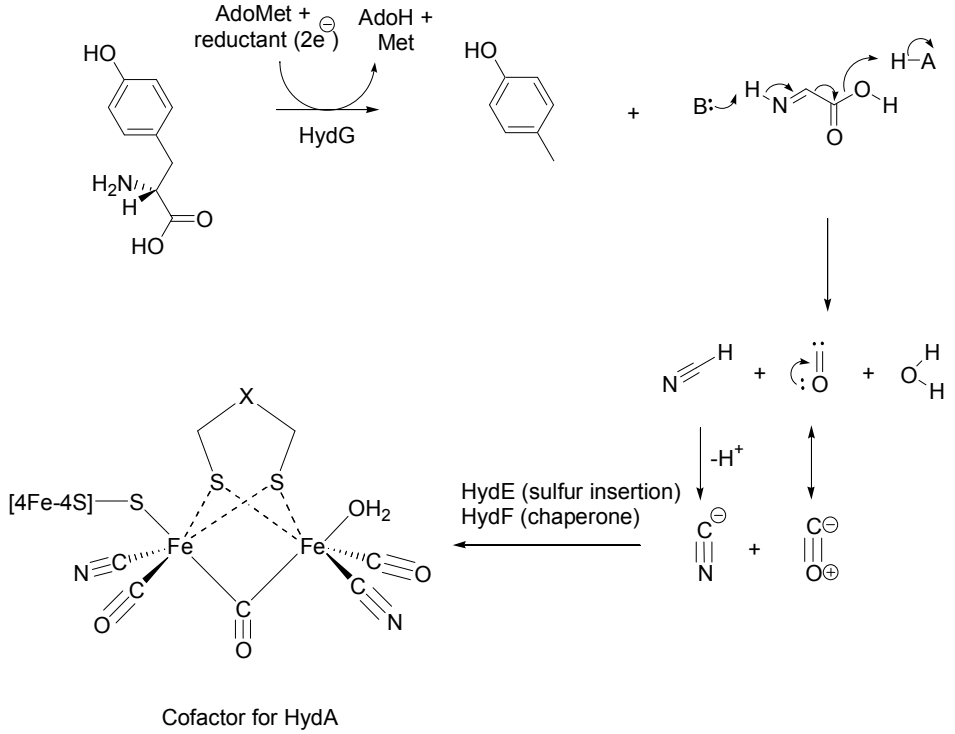
Hydrogenases are enzymes that catalyse the reversible reduction of protons to molecular hydrogen (H_2). These enzymes enable the organism to use either H_2 as a reducing source, or protons as terminal electron acceptors for the electrons produced during metabolism. Over recent years, there has been significant interest in hydrogenases, since they have the potential to generate hydrogen as a bio fuel[194].

There are three phylogenetically unrelated types of hydrogenases: [NiFe]-hydrogenases. [FeFe] hydrogenases and [Fe]-hydrogenase[197]. All of these enzymes contain unique active sites with iron coordinated to cyanide or carbon monoxide ligands[198].

The assembly of the cofactor found in the iron only hydrogenase (HydA), shown in figure 4.10, is proposed to require a chaperone (HydF) and two enzymes HydE and HydG, which show sequence similarity to the 'radical AdoMet' superfamily of enzymes (discussed in section 1.4).

HydE shows similarity to the sulfur insertion enzymes BioB and LipA, which may hint that this enzyme is involved in the biosynthesis of the bridging dithiolane ligand[199]. The crystal structure of HydE from *T. maritima* was recently elucidated by Nicolet and co-workers[182], although the nature of the substrate remains unknown.

The second protein, HydG, shows a high degree of sequence similarity to ThiH. We hypothesise that HydG catalyses a reaction closely related to ThiH, cleaving tyrosine and forming *p*-cresol and dehydroglycine. However, the proposed pathways diverge at this point, as dehydration of this dehydroglycine would yield the cyanide and carbon monoxide required for the HydA cofactor (scheme 4.1).



Scheme 4.1 Proposed mechanism for the formation of the HydA cofactor.

The dehydration of dehydroglycine could potentially to be initiated by HydG via the radical-anion **66** (section 3.3, scheme 3.3), or another active site residue acting as a base to catalyse the dehydration reaction.

Preliminary studies on the activity of HydG have been undertaken within the Roach group. Working with recombinant HydG, Mr. M. Challand has detected the presence of *p*-cresol in HydG samples (results not shown).

The microorganisms listed in table 4.1 were aligned (figure 4.1), and from the phenogram (figure 4.10) it is possible to identify to closely related families. HydG and ThiH were re-assigned to each of these organisms.

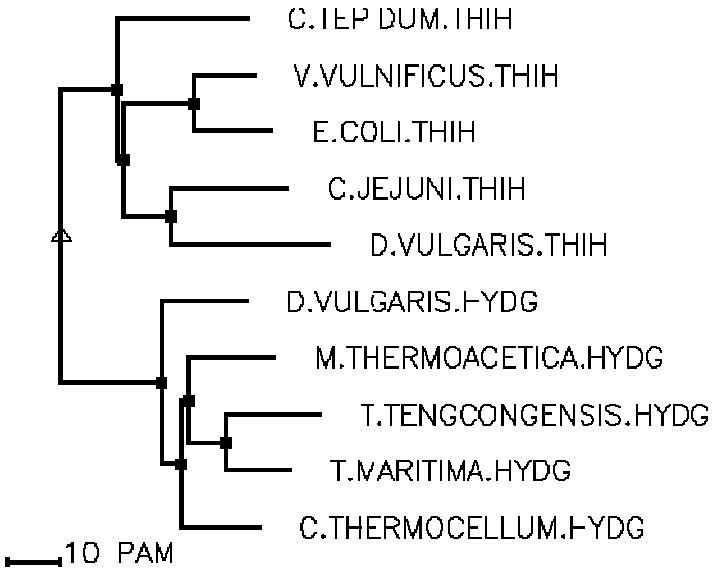


Figure 4.10 Phenogram representing sequence homology between ThiH and HydG for selected organisms. Image created using Multialin[200]

4.4 Summary and conclusions

In order to study novel methods for the production of *holo*-ThiH, several plasmids were assembled using genomic DNA from two microorganisms, *D. vulgaris* and *T. maritima.* The selection of these organisms was based on the difference in operon structure (*D. vulgaris*) and thermostability of the organism (*T. maritima*).

Studies using *D. vulgaris* plasmid pFM001, encoding for ThiSGH-His, resulted in a low level of soluble ThiGH, judging from the yield of protein obtained. This indicates that either the expression system was not appropriate, i.e. the chosen host (*E. coli*) was unsuitable for the expression of these genes; or that the strength of the promoter was not appropriate. The co-expression of *thiSGH*-His from *D. vulgaris* with the *E. coli isc* operon was also investigated, but these studies proved to be ineffective since the protein YqJI co-eluted with the ThiGH complex. This had been previously observed in

studies using *E. coli* (section 2.2). However, the increase of iron content observed in ThiGH expressed from pFM001, suggests that the *D. vulgaris* operon might produce more stable form of *holo*-ThiGH.

The investigations using *T. maritima* produced unexpected results. Initial assignment of *T. maritima* (TM1269) strain indicated the presence of a *thiH* gene. Therefore, a plasmid was assembled encoding for *T. maritima thiH*-His, pFM003, and expressed in *E. coli* BL21(DE3) cells. The co-elution of ThiH with a second protein, which appears to be ThiG was confusing since the *thiG* gene was not present in the expression plasmid. An explanation for this result might be that ThiH associates with ThiG present in the *E. coli* cells during *in vivo* production. The co-expression with the Isc proteins, pFM003, was unsuccessful, since ThiH co-eluted with YqJI. This was also observed in studies using *D. vulgaris* and *E. coli*. The low solubility of ThiH might be associated with unsuitable choice of promoter, but most likely is caused by the significant difference of codon usage that exists between *T. maritima* and *E. coli*. The use of Rosetta cells, which contain rare codons, to express *T. maritima* genes, might produce a better yield of soluble protein.

The *T. maritima* (ORF TM1267) gene initially assigned as thiH was later found to be an iron only hydrogenase maturation protein, HydG, suggested to be a member of the 'radical AdoMet' family. Finally, a mechanism was proposed in which this enzyme participates in a reaction very similar to that catalysed by ThiH (described in chapter 3) resulting in the formation of *p*-cresol. Preliminary studies by Mr. M. Challand have identified *p*-cresol as a product from HydG assays, lending strong support to this hypothesis.

CHAPTER 5

Conclusions

Investigations on thiazole biosynthesis (Thz-P), a thiamine precursor, were initiated by expressing and isolating the enzymes involved in this reaction. The isolation of active *holo*-ThiH, an enzyme required for thiazole synthesis, was essential for the reconstitution of *in vitro* activity. ThiH has been classified as a member of the 'radical-AdoMet' family, and contains an essential iron sulfur cluster. Due to the oxygen sensitivity of this cluster, the isolation of this enzyme in the *holo*-form is challenging.

Studies on the development of novel expression and purification techniques were carried out with the aim of improving *in vivo* production of ThiH. Expression studies confirmed that pRL1020, a plasmid assembled by R. Leonardi, remains the best expression system for the production of soluble ThiGH. Studies using addition of supplements to the growth medium showed no improvement on the production and iron content of ThiGH; and the use of alternative affinity chromatography methods, to replace nickel affinity chromatography, proved to be ineffective. Nevertheless, the use of ion exchange chromatography produced a relatively pure sample of monomeric ThiH. This technique might potentially be an alternative to affinity chromatography purification. Variations of the buffers used during purification showed significant improvement on the yield of protein; from 2.2 mg/g cell obtained using Tris buffer to 5.2 mg/g cell using MOPS buffer. The use of MOPS buffer and the addition of anaerobic cell lysis step to the purification protocol produced very stable ThiGH complex.

For the first time, successful *in vitro* chemical reconstitution of ThiH (5.2 ± 1.3 mol equivalents of Fe per mol of ThiH) was achieved. Once reconstituted in this manner, it

was possible for the ThiGH complex to be concentrated to ~40 mg/mL and to be used for *in vitro* experiments. Chemical reconstitution and concentration procedures followed by gel filtration chromatography allowed for the separation of the ThiGH complex from monomeric ThiH, resulting in samples with high iron content. The stability of monomeric *holo*-ThiH, which could be concentrated to 60 mg/mL, is proof of the improvement achieved by changing the purification methodology and the chemical reconstitution procedure. This achievement is likely to prove useful during future crystallisation studies of ThiH, which will give detailed information about the structure of the active site of this protein, and insights into the mechanism of tyrosine cleavage.

The successful production of *holo*-ThiH assisted the development of a functional *in vitro* assay used to study the mechanism of tyrosine cleavage. The *in vitro* assay contained chemically reconstituted ThiGH, AdoMet, tyrosine and a reducing system (NADPH/Fld/Fpr). The aromatic side chain released upon tyrosine cleavage, initially thought to be *p*-hydroxybenzyl alcohol, was identified as *p*-cresol. The remaining fragment was identified as glyoxylate, the *in vitro* hydrolysis product of dehydroglycine.

The development of an HPLC method using UV-visible detection further confirmed the presence of *p*-cresol and also allowed for its quantification, the retention time was found to be 16.7 ± 0.5 min. In addition, this method detected AdoH (R_t= 11.8 ± 0.4 min), the product of AdoMet reductive cleavage by the [4Fe-4S] cluster; and tyrosine consumption was also monitored by this method. The intermediate produced by the tyrosine cleavage reaction, dehydroglycine, is thought to undergo rapid hydrolysis upon release from the ThiGH complex, but the hydrolysis product, glyoxylate was also detected. Quantification and detection of glyoxylate in the *in vitro* assays was achieved using a method for devised for the detection of 2-oxoacids, in which glyoxylate is converted to the fluorescent 2-quinoxalinol derivative. The retention times were found to be R_t = 11.6 ± 0.5 min for derivatised glyoxylate, and R_t = 11.6 ± 0.05 min for 2-quinoxalinol.

The development of these two HPLC methods allowed the investigation of the kinetics and stoichiometry for the formation of the reaction products formed upon tyrosine cleavage. In addition the rate of consumption of tyrosine was also calculated. Uncoupled turnover of AdoMet was observed, with 1.3 mol equivalents of AdoH being produced relative to *p*-cresol. The rate of formation of AdoH is found to be higher (0.102 ± 0.033 min^{-1}) than the rate of tyrosine consumption (0.052 ± 0.003 min^{-1}). In another *in vitro* assay the rates of formation and stoichiometry of *p*-cresol and glyoxylate were calculated. These tyrosine derived products are formed in an approximate 1:1 ratio over time (176.2 ± 7.8 μM *p*-cresol, 171.6 ± 7.9 μM glyoxylate), and the reaction is shown to reach completion after 60 min. The fitted first order rate constants of formation were found to be 4.6 ± 0.7 [10^{-2} min^{-1}] for *p*-cresol, and 6.3 ± 1.1 [10^{-2} min^{-1}] for glyoxylate. The rates of formation of *p*-cresol seem to be approximately the same in different experiments. To explain these observations, a novel mechanism for tyrosine cleavage was proposed. Further studies using tyrosine analogues might provide information on the kinetics of this reaction and give further information about the mechanism of Thz-P formation.

In order to study novel methods for the production of *holo*-ThiH, several plasmids were assembled using genomic DNA from two microorganisms, *D. vulgaris* and *T. maritima.* The selection of these organisms was based on the difference in operon structure (*D. vulgaris*) and thermostability of the organism (*T. maritima*). Studies using *D. vulgaris* resulted in a low level of soluble ThiH but slightly higher iron content (in comparison with ThiH produced from *E. coli*). Careful analysis of the codon usage from *D. vulgaris* genes suggest that the use of Rosetta host cells might assist in the production of ThiH. Despite its original assignment as a ThiH, the sequence from *T. maritima* (ORF TM1267) has subsequently been proposed to form part of the iron only hydrogenase cluster biosynthesis pathway. During these studies, *T. maritima* HydG was expressed and isolated and future experiments will investigate the reaction catalysed by this enzyme. Due to the sequence similarities between ThiH and HydG, a mechanism sharing many common steps has been proposed for HydG. The observation of Mr. M. Challand of *p*-cresol formation in HydG assays supports this model.

CHAPTER 6
Experimental Methods

6.1 Materials

The following table (table 6.1) summarises all materials and reagents used throughout this project.

Table 6.1 Materials and Reagents.

Materials	Supplier
Reagent grade chemicals *(unless otherwise stated)*	Sigma-Aldrich
Molecular biology reagents	Promega or Qiagen
Restriction Enzymes	Promega or New England Biolabs
Taq DNA polymerase T4 DNA ligase Molecular weight markers for DNA electrophoresis	Promega
Pfu Turbo DNA polymerase	Stratagene
QuikChange® Multi Site-Directed Mutagenesis Kit	
dNTPs	MBI Fermentas
Oligonucleotides	Sigma-Genosys
Electrophoresis grade agarose	Bio-Rad
Yeast-extract	Oxoid (Fisher)
Bacto tryptone	
Bacteriological agar	
β-(+)-L-arabinose	Avocado
$K_3[Fe(CN)_6]$	
Molecular weight markers for SDS-PAGE	BDHMerck
Acetic acid	

Potassium acetate	
Potassium dihydrogen phosphate	
Dipotassium hydrogen phosphate	
TLC plates (aluminium backed silica Gel-60 (F-254, 250μm) and RP-18 F254)	
Ampicillin	Melford Laboratories
Anhydrous $FeCl_3$	Acros
Perchloric acid (20%)	Fluka
Polyacrylamide-bis polyacrylamide (30% w/v, 37 :5 :1)	Amresco
Glycerol	Fisher Scientific
HPLC grade solvents	
L-[U-^{14}C]tyrosine	GE Healthcare
S-adenosyl-*L*-[*methyl*-^{14}C] methionine	
PD-10 and NAP-10 gel filtration columns	
Q and S- Sepharose High Performance	AP Biotech
Superdex 75 (S-75)	
Superdex 200 (S-200)	
Chelating Sepharose Fast Flow resins	
Pressure cells (10, 50 and 250 mL)	Amicon
PM30 and PM10 ultrafiltration membranes	Millipore
Biomax Ultrafree filter membranes (5k NMWL)	
Sterile 0.22 um filters	
Analytical and semi-preparative HPLC columns	Roche, Fisher Scientific and Phenomenex
pBAD-HisA	Invitrogen
E. coli One Shot cells	
E. coli BL21(DE)3	Novagen
pFLAV and pFLDR[201]	Gift from Dr. K. S. Hewitson (Oxford University)

6.2 Equipment

Centrifugation: Samples (2-250 mL) were centrifuged at 4 ºC in an Avanti J-25 centrifuge (Beckman). For samples of a volume less than 2 ml a microCentrifugette 4221 (ALC) at room temperature or a Biofuge Fresco (Heraeus) at 4 ºC were used.

Determination of pH: Mettler Delta 340 pH meter connected to a Mettler Toledo Inlab 413 Combination Electrode was used for pH determination and adjustment. The pH meter was calibrated at pH 7.0 and 10.0 before use and stored in a saturated solution of KCl.

PCR, Sterilisation and Incubation: A Progene Thermal Cycler (Techne) was used for PCR amplifications. Media and heat stable solutions required for microbiological experiments were sterilised in a PriorClave steam autoclave (Priorclave Ltd) at 121 ºC for 25 min. Bacterial cultures (solid and liquid) were incubated in an Innova 4400 Incubator Shaker (New Brunswick Scientific) or in an Innova 4230 Refrigerated Incubator Shaker (New Brunswick Scientific).

Cell Lysis by Sonication: Cells suspensions were lysed by sonication in ice/cold water baths using either a Soniprep 150 sonicator (Sanyo) or a VC130 sonicator (Sonics and Materials, Newtown, CT). Cell lysates were cleared in 250-mL gas-tight polycarbonate centrifuge bottles (Beckman-Coulter, High Wycombe, UK).

Anaerobic Purifications, protein reconstitutions and assays: all experiments requiring anaerobic conditions were carried out under a nitrogen atmosphere in an anaerobic glove box with less than 2 ppm of O_2 (Belle Technology, Portesham, UK). The glove box was maintained at 19 ºC in a temperature controlled laboratory.

UV-visible Spectroscopy: absorbance readings and UV spectra were recorded aerobically on a Lamba 2 spectrophotometer (Perkin-Elmer) and anaerobically using a Ocean Optics (Duiven, The Netherlands) USB2000 spectrophotometer equipped with a light

Mini-D2-GS source connected by P-400-2-UV/SR optical fibers to a cuvette holder inside the glove box.

NMR: ^{1}H and ^{13}C NMR spectra were recorded on a Bruker DXP 300 operating at 75 MHz. For the samples extracted into organic solvents, the supernatants were extracted, combined and the organic phase was analysed by ^{13}C NMR spectrometry on a Bruker DPX 400 operating at 100 MHz.

FPLC, HPLC and LC-MS: A Pharmacia-LKB fast performance liquid chromatography (FPLC) unit was used for aerobic purifications at 4 ºC; an Amersham Biosciences Acta fast protein liquid chromatograph was used for anaerobic protein purifications at 19 ºC.

A Gilson 321 HPLC work centre equipped with a dual wavelength UV-visible detector and a Shimadzu RF-10Axl fluorescence detector was used for analytical HPLC methods; chromatograms were collected and analyzed using the Gilson Unipoint software (Gilson). Mass spectrometry analysis was carried out using one of three systems: (1) Gilson HPLC coupled to a Thermo Finnigan Surveyor MSQ single quadrupole mass spectrometer with electrospray ionization. The data were collected and processed using the XCalibur software system, and multiple charged species were analyzed by Promass Deconvolution (Thermo, Altrincham, UK); (2) VG Platform single quadrupole; (3) Waters ZMD quadrupole mass spectrometer.

Gel Densitometry Analysis: a Syngene Gene Genius imaging system and related software were used to estimate the relative amounts of proteins analysed by SDS-PAGE and revealed by Coomassie staining (appendix B).

6.3 General Experimental Methods

Standard sterile techniques were applied throughout microbiological experiments. Growth media and heat stable solutions were autoclaved, whilst heat labile solutions were filter sterilised through 0.22 μm filters. Unless otherwise stated, growth media

were supplemented with ampicillin (100 μg/mL) and protein expression was induced by the addition of arabinose [20 % (w/v), 10 mL/L]. Assembled plasmids are listed in appendix A. The composition of media and other relevant solutions is given in appendix B.

Protein purifications and assays were all carried out inside an anaerobic glove box at 19 ºC. Buffers, solutions, containers (bottles, jars, tubes, beakers etc.) and equipment required during the purification were degassed overnight inside the glove box. For the optimised purification procedure (method 13), purification buffers were degassed outside the glove box with a nitrogen line tightly connected to the bottles, sealed and placed inside the glove box overnight.

Columns and the HPLC were placed outside the box but the buffer lines were introduced in a sealed gastight fashion into the box so that the anaerobic buffers could be pumped through the system and used to elute the proteins; the column eluate was fed back into the box to permit anaerobic fraction collection. For experiments using, FPLC system, the equipment was placed inside the glove box. Protein was fed to the column, placed outside the glove box, using gastight lines, and fed back to the box to allow anaerobic collection of the protein fractions. Protein elution was monitored at 280 and 420 nm.

Method 1: PCR amplification of genes of interest

Polymerase chain reactions (PCR; 50 μL) were set up in sterilised tubes (200 μL) using the conditions reported in table 6.2; however, in order to minimise the extension of partially annealed primers to non-specific sites on the genomic DNA, the DNA polymerase was added only after 3 min incubation at 95 ºC (table 6.3).

Table 6.2 Conditions for PCR reaction.

Component	Stock concentration	Volume	Total quantity
Pfu turbo buffer	10 ×	2.5 μL	1 ×
dNTPs	2 mM	10 μL	20 nmol
DNA template (pRL1020)	10 or 100 ng/μL	1 μL	10 or 100 ng
Primers	10 μM	2.5 μL/each	0.025 nmol
Pfu turbo DNA polymerase	0.25 U/μL	5 μL	1.25 U
+/- DMSO		5 μL	
Sterile H_2O		to 50 μL	

* Two reaction mixtures containing different amounts of DNA template (100 and 10 ng) were prepared for the amplification of target genes.

PCR reactions were processed in a thermocycler as reported in table 6.3, then analysed on 1 % agarose gels by gel electrophoresis.

Table 6.3 PCR amplification reaction cycle conditions.

Nº of cycles	T/ºC	Time/min
1	95	3
	80 (addition of the polymerase)	3
30	95 (denaturation)	1
	50 (annealing)	1
	72 (elongation)	1-1.5/kb
1	72	10
	4	as required*

*At the end of the last cycle, reaction mixtures were stored at 4 ºC until removed.

PCR products were purified using Wizard® PCR Preps DNA Purification System (Promega), as per manufacturer's instructions. Plasmid DNA was isolated using

QIAprep Spin Miniprep Kit (Quiagen), as per manufacturer's instructions. Sterile water (25-45 μL) was used to elute the purified plasmid DNA.

Method 2: Preparation of Competent Cells and Transformation

E. coli competent cells were either purchase (One Shot) or prepared by the rubidium chloride method[202]. Purchased competent cells were transformed according to the manufacturer instructions and in the case of the competent cells prepared in our laboratory the following general method as been employed successfully.

Aliquots of competent cells (50-200 μL) contained in 1.5 mL microcentrifuge tubes were defrosted on ice for 10 min before adding, with gentle mixing, 6-10 μL of a ligation reaction (method 7) or 1 μL of purified plasmid DNA. The tubes were then incubated on ice for 30 min and then heat-shocked for 45 seconds in a bath maintained at 42 ºC. Cells were returned to the ice for 2 min, and room temperature SOC medium (800-950 μL) was added. The resulting cell suspensions were incubated at 37 ºC (230 rpm) for 1 hour before being plated on selective 2YT or supplemented DM agar plates (appendix B). These plates were incubated overnight at 37 ºC. Positive colonies (those containing the correct construct) were identified as follows: well isolated single colonies were picked from plates with a 10 μL sterile tips and placed into a 50 mL falcon tube containing 10 mL of 2YT medium supplemented with ampicillin (100 μg/mL). Cells were grown overnight at 37 ºC for further investigation or storage.

Method 3: Glycerol freeze preparation

Well isolated single colonies were picked from plates with a 10 μL sterile tip and placed into a 50 mL falcon tube containing 10 mL of 2YT medium supplemented with ampicillin (100 μg/mL). The cell culture was incubated overnight at 37 ºC. Glycerol (125 μL) was added to the cell culture (500 μL), mixed vigorously, and stored at -80 ºC.

Method 4: Plasmid DNA isolation

Plasmid DNA was isolated using Wizard Plus Minipreps DNA Purification System (Promega), following manufacturer instructions.

Method 5: Restriction digestion (analytical and preparative)

Analytical restriction digestion of plasmid DNA isolated from 5 mL bacterial culture (method 10) used the following conditions:

*Table 6.4 Conditions for analytical plasmid DNA restriction digestion.**

Component	Concentration	Volume	Total quantity
Plasmid	50-75 ng/μL	5 μL	250-375 ng
Buffer 4	10 ×	1 μL	1 ×
BSA	1 mg/mL	1 μL	1 μg
Enzyme 1	10 U/μL	0.5 μL	5 U
Enzyme 2	10 U/μL	0.5 μL	5 U
Sterile H_2O		2 μL	

* Total volume: 10 μL.

Preparative restriction digestion of plasmid DNA (25 μL, 100-300 ng/μL) isolated from 10 mL bacterial culture (method 10) used the following conditions:

Table 6.5 Conditions for preparative plasmid DNA restriction digestion.

Component	Concentration	Volume	Total quantity
Plasmid	20 μL	25 μL	2-6 μg
Buffer 4	10 ×	5 μL	1 ×
BSA	1 mg/mL	5 μL	10 μg
Enzyme 1	10 U/μL	2.5 μL	40 U
Enzyme 2	10 U/μL	2.5 μL	40 U
Sterile H_2O		10 μL	

Reaction mixtures were incubated at 37 ºC for 1-2.5 h and then analysed on a 1 % agarose gel by gel electrophoresis.

Method 6: Purification of digested fragments

Completed preparative digestion reactions (63 μL) were loaded onto a 1 % low melting point agarose gel to separate the fragments of interest. The required bands were excised and the DNA recovered from the agarose using the QIAquick Gel Extraction Kit (Quiagen), as per manufacturer's instructions. Agarose gel electrophoresis analysis was used to estimate the concentration of the purified fragments by comparison with DNA markers of known concentration.

Method 7: *Ligation into an expression vector*

Samples of suitably restricted and purified plasmid and insert DNA were ligated under the following conditions:

*Table 6.6 Conditions for the ligation of a DNA fragment into an expression vector.**

Component	Concentration	Volume	Total quantity
Ligase buffer	10 ×	5 μL	1 ×
Plasmid	50 ng/μL	1 μL	50 ng
Insert	25 – 100 ng/μL	x μL*	
T4 DNA ligase	3 U/μL	1 μL	3 U
H_2O		to 10 μL	

* x varied according to the size of both the vector and the insert, and was calculated to give 3:1 and 1:3 insert:vector molar ratios.

Ligation reactions were incubated overnight at 4 ºC and then used to transform TOP10 competent cells (method 2). Positive colonies were then identified by PCR screening and plasmid DNA isolation followed by analytical restriction digestion.

Method 8: SDS-PAGE analysis[203]

Protein purity was determined by analysis of 15 % SDS-PAGE gels revealed by Coomassie blue staining (appendix B). For each sample, small amounts (20 μL) of cleared lysate and re-suspended cell pellets, were mixed with 2 X SDS-PAGE loading buffer (20 μL, appendix B), heated at 90 ºC for 5 min to ensure the proteins were denatured and then a sample loaded onto a SDS-PAGE gel.

Method 9: Estimation of protein concentration[204]

Protein concentration was routinely determined by the method of Bradford. Briefly, Bradford reagent (1 mL) was added to a protein sample (20 μL) and incubated at room temperature for 5 min, and then the A_{595} was measured using water as a control. The sample was diluted if A_{595} exceeded 1.0. Appropriate correction factors for proteins of unknown concentration were obtained from calibration curves constructed with BSA standards.

Method 10: Small scale expression experiments

Cells from glycerol freezes [pRL1020/BL21(DE3) or pRL1021/BL21(DE3)] were used to inoculate small scale cultures in 2YT medium (10 mL) supplemented with ampicillin (100 μg/mL), which were incubated overnight at 37 ºC. This culture was then used as a 1% inoculum for 200 mL of 2YT growth media containing 200 μL of ampicillin (100 μg/mL). Cells were grown aerobically in an orbital shaker (37 ºC, 180 rpm) and when the OD_{600} reached ~0.6, protein expression was induced by the addition of a filter-sterilised arabinose solution [20 % (w/v), 10 mL/L] and returned to the shaker at 27 ºC. After 5 hours the cell cultures were harvested by centrifugation (12000 rpm, 15 min). Cell paste was weighed and re-suspended in lysis buffer [50 mM Tris-HCl pH 8.0, Glycerol 20% (w/v)] into 50 mL falcon tubes and stored at -80 ºC until further use.

Method 11: Aerobic large scale expression experiments

Cells from a glycerol stock were used to inoculate starter cultures in 2YT medium (5×10 mL) supplemented with ampicillin (100 μg/mL), which were incubated overnight at 37 ºC. These cultures were then used as 1% inoculum into fresh 2YT medium (5 L) aliquoted in 4 L flasks (~ 1.25 L of medium/flask), and cells grown at 37 ºC. When the OD_{600} reached 0.4-0.6, protein expression was induced by the addition of a filter-sterilised arabinose solution [20 % (w/v), 10 mL/L], after which cells were transferred to a shaker at 27 ºC. Cell growth continued until the OD_{600} = 1.8-2.3, cells were harvest by centrifugation (JA-14 rotor, 12000 rpm for 20 min at 4ºC) and cell pellet stored at -80 ºC until further use.

Method 12: Initial anaerobic purifications of ThiGH-His[58]

ThiH was anaerobically purified using the following buffers: *A* - 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 50 mM imidazole, 12.5% (w/v) glycerol; *B* - 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 50 mM imidazole, 12.5% (w/v) glycerol; *C* - 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 500 mM imidazole, 12.5% (w/v) glycerol; *D* - 50 mM Tris-HCl pH 8.0, 5 mM DTT, 12.5% (w/v) glycerol.

To the still frozen cell paste (27-30 g), lysozyme (0.1 mg/mL), benzonase (10 U/mL) and PMSF (100 mM stock solution in *i*-PrOH, 1 mM final concentration) were added under aerobic conditions; the cells were then introduced inside the glove box and re-suspended in anaerobic buffer A (90 mL) stirring continuously for ~30 min. The suspension was then withdrawn from the anaerobic box and rapidly lysed, on ice, by sonication (15-20 times 30 second bursts, 30 second rest between each burst). The lysate was returned to the box, allowed to degas (10 min), and after the addition of PEI (0.2 % v/v) cleared by centrifugation in gastight centrifuge tubes (JA-14 rotor, 12000 rpm, 30-40 min). The brown supernatant was applied to a Chelating Sepharose column (25 mL) previously charged with nickel sulfate (25 mL, 0.2 M) and equilibrated with anaerobic buffer A (10 column volumes). The column was washed with buffer B (4 column volumes), then buffer A (2 column volumes) and proteins were eluted at 5 mL/min

with an increasing gradient of buffer C from 0 to 50% over 4 column volumes, collecting 5 ml fractions. High molecular weight impurities were normally eluted within the first 60-65 mL of the gradient, whilst the ThiGH-His complex was eluted in a broad peak between 45 and 200 ml (towards the end of the gradient). To avoid protein precipitation, the most concentrated and brown fractions (typically fractions 15-20), were immediately pooled and applied (up to 30-35 mL) to an S-75 gel filtration column (3 cm I.D. x 15 cm), previously equilibrated with anaerobic buffer D. Proteins were eluted isocratically at 2 mL/min and collected in 10 mL fractions. The most concentrated fractions were pooled, glycerol added to a final concentration of 25% (w/v) and proteins concentrated by ultrafiltration (30 kDa MWCO) to 5-6 mg/mL, as the ThiGH-His complex precipitates above 7 mg/mL. Proteins were stored at -80 ºC in this buffer (buffer E).

Aliquots of protein (50 μL) were removed from the fractions collected during the elution from the nickel-chelating and the gel filtration columns, and subsequently analysed by SDS-PAGE (method 8). Due to the excellent reproducibility of this purification procedure, highly pure ThiGH-His was routinely obtained, in spite of the fact that the purity was not checked prior to pooling and concentrating the fractions. This was judged to be necessary to allow the highly unstable complex to be rapidly exchanged into a buffer in which it was more stable.

Method 12a: Introduction of anaerobic cell lysis

To the still frozen cell paste (27-30 g), lysozyme (0.1 mg/mL), benzonase (10 U/mL) and PMSF (100 mM stock solution in *i*-PrOH, 1 mM final concentration) were added under aerobic conditions; the cells were then introduced inside the glove box and re-suspended in anaerobic buffer A (90 mL) stirring continuously for ~30 min. The suspension was then lysed, on ice, by sonication (30 min, amplitude: 70, pulser: 2 sec), and after the addition of PEI (0.2 % v/v) cleared by centrifugation in gastight centrifuge tubes (JA-14 rotor, 12000 rpm, 30-40 min). The purification protocol is described above (method 12)

Method 13: Optimised anaerobic purifications of ThiGH-His

Cells were grown on a large scale as described in method 11. ThiGH was anaerobically purified using the following buffers: *A* - 50 mM Mops pH 7.7, 200 mM NaCl, 50 mM imidazole, 10% (w/v) glycerol; *B* - 50 mM Mops pH 7.7, 500 mM NaCl, 50 mM imidazole, 10% (w/v) glycerol; *C* - 50 mM Mops pH 7.7, 200 mM NaCl, 500 mM imidazole, 10% (w/v) glycerol; *D* - 50 mM Mops pH 7.7, 200 mM NaCl, 5 mM DTT, 10% (w/v) glycerol.

To the still frozen cell paste (27-30 g), lysozyme (0.1 mg/mL), benzonase (10 U/mL) and PMSF (100 mM stock solution in *i*-PrOH, 1 mM final concentration) were added under aerobic conditions; the cells were then introduced inside the glove box and re-suspended in anaerobic buffer A (90 mL) stirring continuously for ~30 min. The suspension was then lysed, on ice, by sonication (30 min, amplitude: 70, pulser: 2 sec), and after the addition of PEI [final concentration: 0.2 % (v/v); pH 8.0, adjusted with HCl] cleared by centrifugation in gastight centrifuge tubes (JA-14 rotor, 12000 rpm, 30-40 min). The brown supernatant was applied to a Chelating Sepharose column (25 mL) previously charged with nickel sulfate (25 mL, 0.2 M) and equilibrated with anaerobic buffer A (10 column volumes). The column was washed with buffer B (4 column volumes), then buffer A (2 column volumes) and proteins were eluted at 5 mL/min with an increasing gradient of buffer C from 0 to 50 % over 4 column volumes, collecting 5 mL fractions. High molecular weight impurities were normally eluted within the first 60-65 mL of the gradient, whilst the ThiGH-His complex was eluted in a broad peak between 45 and 200 mL (towards the end of the gradient). To avoid protein precipitation, the most concentrated and brown fractions (typically fractions 15-20), were immediately pooled and applied (up to 30-35 mL) to a S-75 gel filtration column (3 cm I.D. x 15 cm), previously equilibrated with anaerobic buffer D. Proteins were stored at -80 ºC. Aliquots of protein (50 μL) were removed from the fractions collected during the elution from the nickel-chelating and the gel filtration columns, and subsequently analysed by SDS-PAGE (method 8). The elution of ThiGH from the columns, using this purification method, was highly reproducible so it allowed the rapid transfer of protein samples to the next purification step without SDS-Page gel

analysis. This rapid purification is highly advantageous as ThiGH complex shows great tendency to precipitate in buffers other than the final storage buffer.

Method 14: Iron analysis[137]

The amount of iron present in purified ThiGH was analysed by the methods of Fish and was determined in triplicate. For the iron analysis, the following solutions were prepared: reagent A, 0.142 M $KMnO_4$ in 0.6 N HCl, obtained by mixing equal volumes of 1.2 M HCl and 0.284 M $KMnO_4$ in H_2O (445 mg in 10 mL), and reagent B, 6.5 mM ferrozine, 13.1 mM neocuproine, 2 M ascorbic acid, 5 M ammonium acetate in H_2O, prepared by first dissolving the ammonium acetate (4.85 g) and ascorbate (4.4 g) in H_2O (12.5 mL) followed by ferrozine and neocuproine (40 mg each). Reagent A was freshly prepared every time, whilst reagent B was stored in the dark and was used for no longer than 3 weeks. Standard solutions were prepared on a range of concentrations (0 – 72 nmoles Fe^{2+}) from a stock solution of $FeSO_4 \cdot 7\ H_2O$ in water (500 μg in 10 mL, 180 μM). This dilution method has been previously described by Leonardi[58].

ThiGH-His samples (300 μL, 5-6 mg/mL) were diluted to 1 mL with water and incubated, together with the standards, for 2 h at 60 ºC after the addition of reagent A (500 μL). At the end of the 2 hour incubation, samples were allowed to cool down to room temperature before adding reagent B (100 μL). When the purple colour was completely developed (15-20 min) the A_{562} of the samples was measured. If necessary, protein precipitate was removed by centrifugation before reading the absorbance. The amount of iron present in the ThiGH-His samples was estimated from standard calibration curves constructed in parallel (figure 6.1).

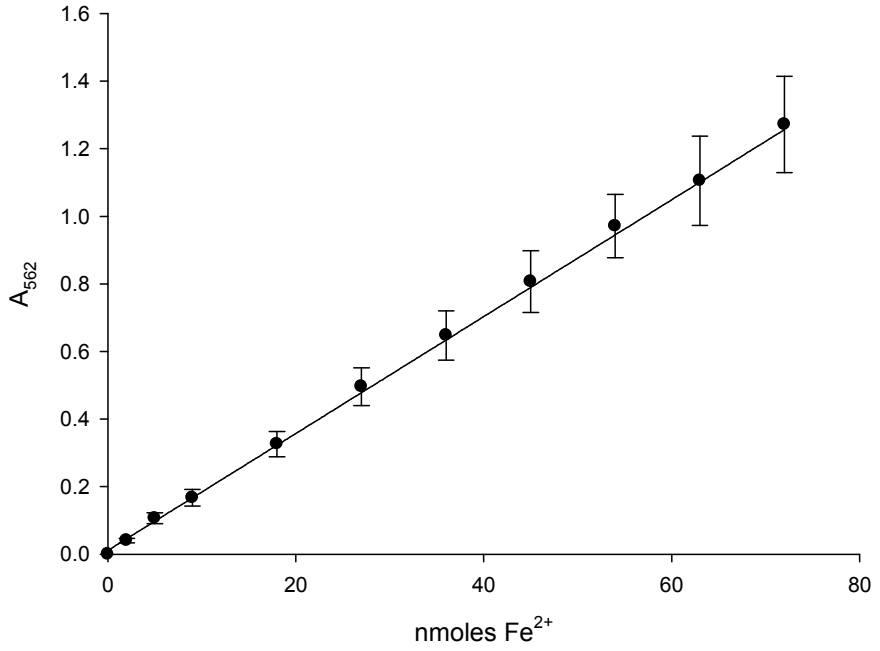


Figure 6.1 Iron standard curve. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.9985, y0=1.094E^{-2}, a=1.730E^{-2}.

6.4 Experimental for chapter 2

6.4.1 Small scale anaerobic expression experiments using pRL1020

Cells from glycerol freezes (pRL1020/BL21(DE3)) were used to inoculate small scale cultures in 2YT medium (10 mL) supplemented with ampicillin (100 μg/mL), incubated overnight at 37 ºC. This culture was used as a 1% inoculum for 1 L of growth medium [(1) Glycerol Minimal Medium; (2) 2YT Medium; (3) 2YT Medium supplemented with sodium nitrate (40 mM, $NaNO_3$) and sodium nitrite (5 mM, $NaNO_2$) depending on the experiment]. To each of them 1 mL of ampicillin (100 μg/mL) was added. Cells were grown aerobically in an orbital shaker (37 ºC, 180 rpm) and when the OD_{600} reached ~0.6 (or ~0.3 when glycerol minimal medium was used), protein expression was induced by the addition of a filter-sterilised arabinose solution [20 % (w/v), 10 mL/L] and returned to the shaker at 27 ºC. After 1 hour, growth media was transferred into gas tight bottles and placed inside the glove box to degas for ~30 min. The bottles were

then withdrawn from the box and left stirring overnight in a cold room at 4 ºC. The following morning bottles were returned to the glove box and the cell culture was anaerobically transferred into gas tight centrifuge bottles. Cells were harvest by centrifugation (JA-14 rotor, 12000 rpm, 20 min, 4 ºC). Cell paste was placed inside the glove box, weighed, re-suspended in lysis buffer [50 mM Tris-HCl pH 8.0, Glycerol 20% (w/v)] into 50 mL falcon tubes and immediately frozen in liquid nitrogen. The cell suspension was stored at -80 ºC until further use.

6.4.2 Small scale anaerobic expression experiments using pRL1021

Cells from glycerol freezes (pRL1021/BL21(DE3)) were used to inoculate small scale cultures in 2YT medium (10 mL) supplemented with ampicillin (100 μg/mL), incubated overnight at 37 ºC. This culture was then used as a 1% inoculum for 1 L of 4 different growth medium: (1) 2YT medium; (2) 2YT supplemented with tyrosine (2 mM); (3) 2YT supplemented with ferrous sulfate ($FeSO_4$, 0.3 mM) and (3) 2YT supplemented with tyrosine (2 mM) and iron sulfate ($FeSO_4$, 0.3 mM); to each of them 1 mL of ampicillin (100 μg/mL) was also added. Cells were grown aerobically in an orbital shaker (37 ºC, 180 rpm) and when the OD_{600} reached ~0.6 (or ~0.3 when glycerol minimal medium was used), protein expression was induced by the addition of a filter-sterilised arabinose solution [20 % (w/v), 10 mL/L] and returned to shaker at 27 ºC. After 1 hour, growth media was transferred into gas tight bottles and placed inside the glove box to degas for ~30 min. The bottles were then withdrawn from the box and left stirring overnight in a cold room at 4 ºC. The following morning bottles were returned to the glove box and the cell culture was anaerobically transferred into gas tight centrifuge bottles. Cells were harvest by centrifugation (JA-14 rotor, 12000 rpm, 20 min, 4 ºC). Cell paste was placed inside the glove box, weighed, re-suspended in lysis buffer [50 mM Tris-HCl pH 8.0, Glycerol 20% (w/v)] into 50 mL falcon tubes and immediately frozen in liquid nitrogen. The cell suspension was stored at -80 ºC until further use.

6.4.3 Large scale anaerobic expression experiments using pRL1021

Cells from the glycerol freeze of interest were used to inoculate starter cultures in 2YT medium (5×10 mL) supplemented with ampicillin (100 μg/mL), incubated aerobically overnight at 37 ºC. These cultures was then used as 1 % inoculum into fresh 2YT medium (3 L) 2YT supplemented with tyrosine (2 mM) and ferrous sulfate ($FeSO_4$, 0.3 mM) and placed in a fermentor; cells were grown at 37 ºC, and constantly purged with nitrogen to provide an anaerobic atmosphere. When the OD_{600} reached 0.4-0.6, protein expression was induced by the addition of a filter-sterilised arabinose solution [20 % (w/v), 10 mL/L], cells were cooled down to 27 ºC. Cell growth continued until the OD_{600} = 1.8-2.3, and after nitrogen pressure was applied to the air port of the fermentor, and a harvest line was used to transfer the cell culture to previously degassed centrifuge bottles inside an anaerobic glove box. The cells were harvest by centrifugation (JA-14 rotor, 12000 rpm for 8 min at 4 ºC), transferred to the glove box and cell pellet collected and immediately stored at -80 ºC.

6.4.4 Small scale anaerobic purifications of ThiGH

ThiGH was anaerobically purified from cell past (~3 g cells) using a purposely made small nickel column (figure 6.2) at 19 ºC. The purification buffers were: *A* - 50 mM Tris-HCl pH 8.0, 12.5% (w/v) glycerol; *B* - 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 10 mM imidazole, 12.5% (w/v) glycerol; *C* - 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole, 12.5% (w/v) glycerol; *D* - 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 50 mM imidazole, 12.5% (w/v) glycerol; *E* - 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 500 mM imidazole, 12.5% (w/v) glycerol.

Nickel columns were prepared by transferring 4 mL of Chelating Sepharose Fast Flow resin into a small column (figure 6.2); excess buffer was removed and the resin was charged by re-suspending 8 ml of nickel sulfate (0.2 M). The column was then washed with buffer A (3× 2 mL). Meanwhile the frozen cell paste (3 g) was introduced inside the glove box and re-suspended in 9 mL of anaerobic buffer A and left to stir for 15 min. The suspension was then lysed, on ice, by sonication (20 min, amplitude: 70,

pulser: 2 sec), and cleared by centrifugation in microcentrifuge tubes (30 min). The brown supernatant was applied to the column and left to stand for 10 min. The flow through was collected and re-applied to the column repeating the previous step. The column was then washed with increasing gradient of imidazole buffer B, C and D (2×10 mL each). Washings were collected. The protein was eluted by the addition of 6×1 mL of high imidazole buffer E, mixing well each time and left to stand for 5 min before collecting. From each fraction collected during the purification aliquots (50 μL) were removed and subsequently analysed by SDS-PAGE (method 8).

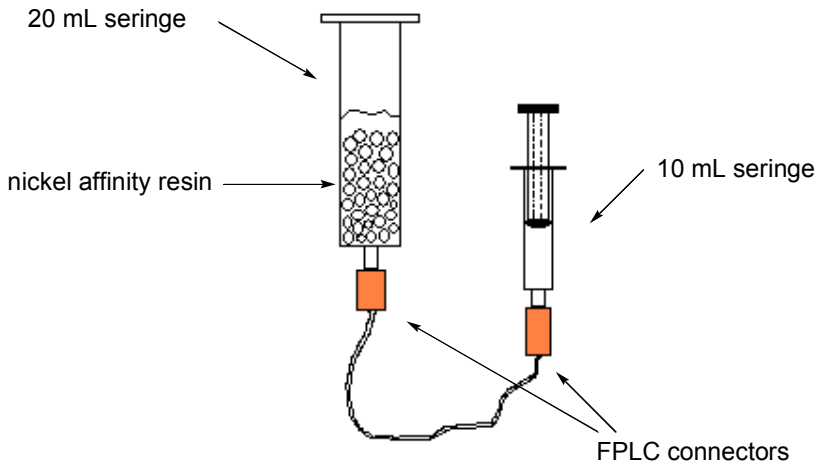


Figure 6.2 Design of the small nickel column used in small scale purifications.

6.4.5 Assembly of pFM700 and pFM707

Unless otherwise specified all genes were amplified using *Pfu* turbo DNA polymerase (method 1) and an appropriate pair of oligonucleotide primers.

The following general strategy was employed to assemble all the expression plasmids mentioned in chapters 2 and 4. The genes of interest were amplified by PCR (method 1) and cloned either into pBAD-HisA or pBAD derivative plasmid. Reaction mixtures were then transformed into *E. coli* TOP10 One Shot competent cells (method 2) and plated onto selective 2YT solid medium. Positive clones were identified by DNA isolation followed by analytical restriction digestion (method 5). Once isolated, positive

clones were stored as glycerol freezes (method 3), and their plasmid DNA isolated and transformed into *E. coli* BL21 (DE3) expression strains.

ThiH-StrepTag (1.2 kb) was amplified by PCR (method 1) using two different sets of primers: ThiH-StrepTag F and ThiH-StrepTag R to introduce the Strep-tag to the C-terminal end of ThiH. The sequences for the oligonucleotides used in this reaction are listed in table 6.7.

*Table 6.7 Sequences of oligonucleotides used in polymerase chain reaction.**

Primers	**Sequence 5′ → 3′**
ThiH-StrepTag F*	CGT CAA CAT GGC GAA GGC AT
ThiH-StrepTag R*	CCC TCT CGA GTC ACT TTT CGA ATT GTG GAT GAC TCC AAA CTA GTC TTT GCG AGG CGC

*****Primers for PCR amplification were synthesised by Shannon Campbell (Southampton University, UK). ThiH-StrepTag F: forward primer; ThiH-StrepTag R: reverse primer.

The PCR product (ThiH-Stag) was restricted with *RsrII*/*XhoI* and cloned into pRL1020 on a *RsrII*/*XhoI* fragment. Both digested pRL1020 and PCR product were loaded onto a 1% agarose gel (appendix A) and analysed by gel electrophoresis (figure 6.3).

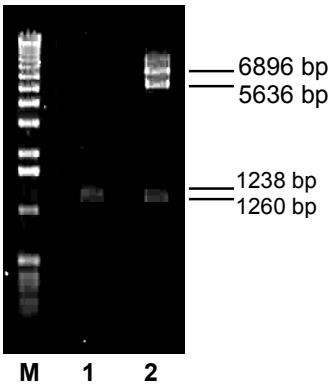


Figure 6.3 Identification of restriction digestion of PCR product and pRL1020. Lane 1: ThiH-Streptag (1238 bp); Lane 2: pRL1020 restricted with RsrII/XhoI (1260 bp: ThiH-His, 5636 bp: pRL1020 backbone (ThiFSG section), 6896 bp: pRL1020 linearised with XhoI); M: molecular weight marker.

The PCR product was ligated into pRL1020 backbone (ThiFSG section) using T4 ligase (method 7). Ligation reactions were incubated overnight at 4 ºC and then used to transform *E. coli* TOP10 competent cells (method 2). Positive colonies were then identified by analytical restriction digestion using *Stu*I and *BstB*I (figure 6.4, method 5). The resultant vector was named pFM700.

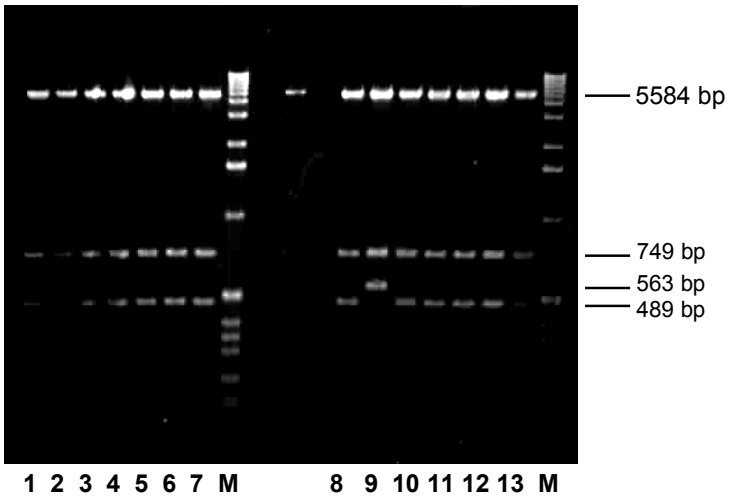


Figure 6.4 Identification of positive clones containing pFM700 by plasmid DNA isolation and restriction. Lanes 1-8 and 10-13: restriction of pFM700 with StuI/BstBI (5584, 749 and 489 bp); lane 9: pRL1020 restriction with StuI/BstBI (5584, 749 and 563 bp); M: molecular weight marker.

pRL1021, a plasmid bearing the *iscSUA-hscBA-fdx* operon, was restricted with *RsrII*/*XhoI* to remove the ThiH-His fragment and the ThiH-Streptag fragment, equally restricted with *RsrII*/*XhoI*, inserted (method 7). The ligation reactions were incubated overnight at 4 ºC and then used to transform *E. coli* TOP10 competent cells (method 2). Positive colonies were identified by analytical restriction digestion using *XhoI*/*BstBI*/*NcoI* (figure 6.5, method 5). The restriction site *BstBI* is unique for this new vector which was named pFM707.

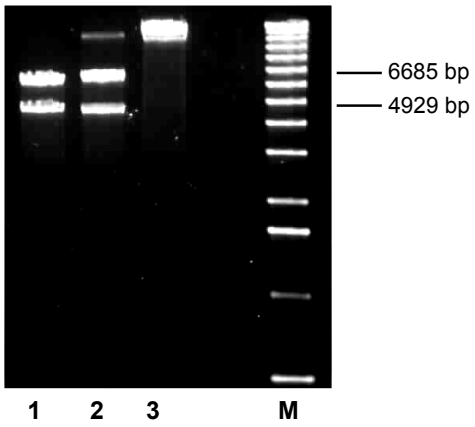


Figure 6.5 Identification of positive clones containing pFM707 by plasmid DNA isolation and restriction. Lane 1: restriction with NcoI (6685, 4929 bp); lane 2: restriction with BstBI (6822, 4792 bp); lane 3: restriction with XhoI (linearised plasmid DNA); M: molecular weight marker.

6.4.6 Expression and purification of pFM700 and pFM707

pFM700 and pFM707 were transformed into *E. coli* BL21(DE3) competent cells (method 2) and cells were grown on a large scale (method 11).

ThiH was anaerobically purified in a glove box at 19 ºC using the following buffers: *A* - 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 12.5 % (w/v) glycerol; *B* - buffer A + 2.5 mM DTB; *C* - Buffer A+ 1 mM HABA; *D* - Buffer A+ 1 mM EDTA

To the still frozen cell paste (2 g), was introduced inside the glove box and re-suspended in anaerobic buffer A (5 mL) stirring continuously for ~15 min. The suspension was then lysed, on ice, by sonication (10 min, amplitude: 70, pulser: 2 sec), and cleared by centrifugation in microcentrifuge tubes (20 min). The brown supernatant was applied to the StrepTactin column (4.5 mL) previously equilibrated with anaerobic buffer A (5 column volumes). The column was washed with buffer A (5 column volumes), and proteins were eluted at 1 mL/min with buffer B (4 column volumes), collecting 0.5 mL fractions. Proteins were stored at -80 ºC in this buffer (buffer B). The column was then regenerated with buffer C (5 column volumes, strong orange colour) and washed with buffer D (storage buffer) until the orange colour disappear, generally 5 column volumes. Protein samples (50 μL) were removed from the fractions collected during the wash and elution from the column, and subsequently analysed by SDS-PAGE (method 8).

6.4.7 Anaerobic purifications of ThiGH with Q-Sepharose

pRL1020 was transformed into *E. coli* BL21(DE3) (method 2) and cells were grown on a large scale as described in method 11. ThiGH was anaerobically purified in a glove box at 19 ºC using the following buffers: Lysis buffer - 50 mM Tris-HCl pH 8.0, 10% (w/v) glycerol; *A* - 50 mM Tris-HCl pH 8.0, 10% (w/v) glycerol; *B* - 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 10% (w/v) glycerol.

To the still frozen cell paste (15 g), lysozyme (0.1 mg/mL), benzonase (10 U/mL) and PMSF (100 mM stock solution in *i*-PrOH, 1 mM final concentration) were added under aerobic conditions; the cells were then introduced inside the glove box and re-suspended in anaerobic lysis buffer (50 mL) stirring continuously for ~30 min. The suspension was then lysed, on ice, by sonication (30 min, amplitude: 70, pulser: 2 sec), and after the addition of PEI [0.2 % (v/v); pH 8.0, adjusted with HCl] cleared by centrifugation in gastight centrifuge tubes (JA-14 rotor, 12000 rpm, 30-40 min). The brown supernatant was applied to the Q-Sepharose column (80 mL) previously equilibrated with anaerobic buffer A (4 column volumes). The column was washed with buffer A (4 column volumes), and proteins were eluted at 2 mL/min with an

increasing gradient of buffer B from 0 to 100% over 2 column volumes and isocratically with buffer B over 1 column volume, collecting 5 mL fractions. To avoid protein precipitation, glycerol was added to a final concentration of 20% (w/v). All fractions were stored at -80 ºC. Protein samples (50 μL) were removed from the fractions collected during the wash and elution from column, and subsequently analysed by SDS-PAGE (method 8).

6.4.8 Anaerobic purifications of ThiGH with S-Sepharose

Fraction 48 eluted from Q-Sepharose (method described above) was anaerobically purified in a glove box at 19 ºC using the following buffers: Adjusting buffer - 50mM Bis-Tris-HCl pH 6.2, 10% (w/v) glycerol; *A*: 50mM Bis-Tris-HCl pH 6.7, 10% (w/v) glycerol; *B*: 50mM Bis-Tris-HCl pH 6.2, 0.5 M NaCl, 10% (w/v) glycerol.

Fraction eluted from Q-Sepharose (6 mL), was adjusted to pH 6.7 and applied to a S-Sepharose column (5 mL) previously equilibrated with anaerobic buffer A. In the case of the experiment at pH 6.7, 6 mL of adjusting buffer were added to the protein sample. The column was washed with buffer A (10 column volumes), and proteins were eluted at 1 mL/min with an increasing gradient of buffer B from 0 to 100 % over 2 column volumes and isocratically with buffer B over 1 column volume, collecting 0.5 mL fractions. All fractions were stored at -80 ºC. Protein samples (50 μL) were removed from the fractions collected during the wash and elution from column, and subsequently analysed by SDS-PAGE (method 8).

6.4.9 Effect of pH and additives

A set of purification experiments were carried out using modified buffers, i.e. variation of the pH, addition of dithionite, AdoMet and/or tyrosine. In all experiments ThiGH was expressed from pRL1020/BL21(DE3) (method 2 and 11) and purified using method 13, the modification were applied to the buffers described in this method.

Table 6.8 lists all the modifications made to the purification buffers. AdoMet, tyrosine and dithionite solutions were prepared and added to the buffers anaerobically. pH variations were carried out aerobically and after buffers were degassed and placed inside an anaerobic glove box.

Table 6.8 Buffers used for purification of ThiGH-His from pRL1020/BL21(DE3).

Buffer	pH	Additives	Final concentration/ mM	Solvent
Tris	8.0	-	-	-
		-	-	-
MOPS	7.7	-	-	-
		Dithionite	1.0	Water
		AdoMet	0.5	Water
		AdoMet	0.5	Water
		Tyrosine	1.0	HCl, 100 mM
		Tyrosine	1.0	HCl, 100 mM
	7.4	Tyrosine	2.0	HCl, 100 mM
	7.0	-	-	-
	6.5	-	-	-

6.4.10 Quantification of AdoMet in SAM-e® tablets by HPLC

In order to quantify the amount of AdoMet present in SAM-e® tablets by HPLC a program was developed and optimised by using standard solutions of AdoMet. A range of concentrations of standards (0.3 μM to 30 mM) were used to construct a calibration curve to estimate the concentration of AdoMet present in the samples. Figure 6.5 shows the AdoMet calibration curve.

In order to extract pure AdoMet, SAM-e® tablets (55 mg) were powdered, dissolved in assay buffer (2 mL, 50 mM MOPS pH 7.7, 100 mM NaCl, 10 % glycerol) and cleared by centrifugation (12,000 rpm, 10 min). The samples were then diluted (10 fold dilutions) in assay buffer and analysed by RP-HPLC (UV-visible detection at 254 nm) using a C18

RP column (Phenomenex Prodigy 5µ, 250 x 5.0 mm) pre-equilibrated in solvent A (25 mM acetic acid with 8 mM sodium octanesulfonate, 2 mL/min). An aliquot (40 µL) was injected onto the column eluted with solvent A for 3 min followed by a gradient to 100% solvent B (50% acetonitrile, 50% solvent A) over 15 min, kept at 100% solvent B for 2 min, reversed to 0% solvent B over 2 min and re-equilibrated for 10 min. The retention time (R_t) for AdoMet standards and SAM-e® extracted samples was 26 min. Each tablet contained 35% of pure AdoMet. The extracted samples were added to the purification buffers in order to achieve a final concentration of 0.5 mM.

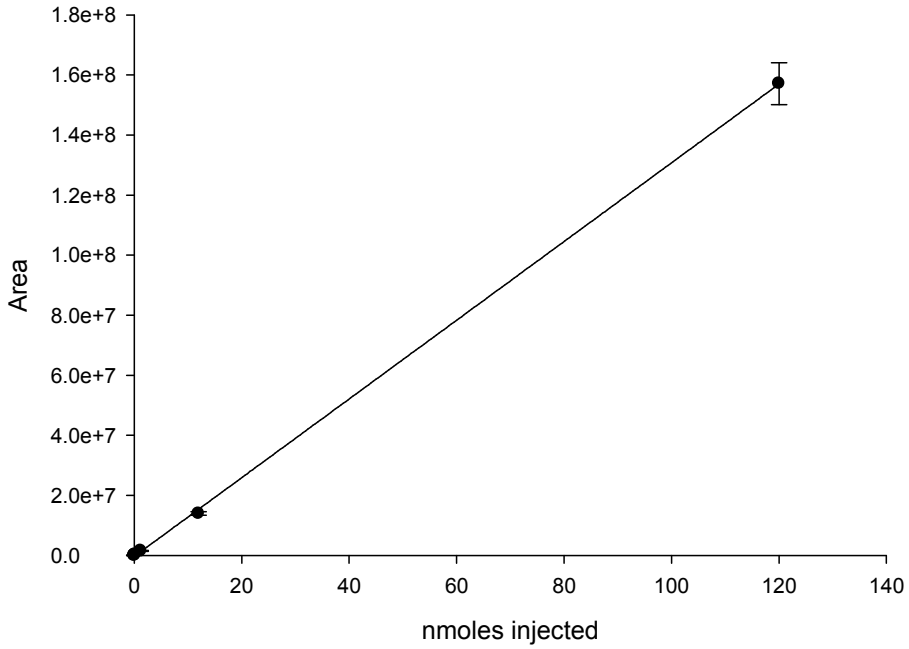


Figure 6.6 Calibration curve for AdoMet standard solutions. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.99, yo=-3.16E+5, a=1.3E+6.

6.4.11 Reconstitution and concentration of ThiGH

DTT solution in buffer D (50 mM MOPS, 100 mM NaCl, 10% (w/v) glycerol; 25 µL; 200 mM) was added to a sample of purified ThiGH complex (1 ml, 83 µM) and gently mixed. After 15 min, 5 mol eq of $FeCl_3$ in water (40 µL, 10 mM) were added carefully in small aliquots (5 µL), and after a further 10 min, 5 mol eq of Na_2S in water were added likewise (total added 40 µL, 10 mM). The protein solution was incubated for another 2 hours, and then the precipitated iron sulfide was removed by applying the solution to

a NAP-10 gel filtration column pre-equilibrated and eluted with buffer D. The iron content was analyzed by the method of Fish (method 14). Reconstituted ThiGH could be concentrated to a maximum concentration of 40 mg/mL using a Biomax 5000 molecular weight cut-off Ultrafree 0.5-mL centrifugal filter.

6.4.12 UV-visible spectra of ThiGH

UV-visible spectra of purified ThiGH were recorder by using an Ocean Optics USB2000 spectrophotometer, connected to a cuvette holder placed inside an anaerobic glove box. The protein samples were defrosted inside the glove box, diluted to 1 mg/mL with anaerobic buffer D and recorded at the range 250-700 nm, against buffer D. For reconstituted protein, the iron excess removed from samples by centrifugation (2 min, 12,000 rpm) and protein spectra recorded.

6.4.13 Analytical gel filtration of ThiGH-His complex

Purified ThiGH-His samples (fractions 4-6 from two different purifications, method 13) were pooled, chemically reconstituted and concentrated to 40 mg/mL (methods 6.4.11). The resulting protein sample was applied to a S-200 column (33 × 750 mm) pre-equilibrated with buffer D [50 mM MOPS pH 7.7, 200 mM NaCl, 5 mM DTT, 10% (w/v) glycerol]. The protein was eluted at 2 mL/min with this buffer, collecting 10 mL fractions. Protein elution was monitored at 280 nm. Eluted proteins were stored at -80 ºC. Aliquots of each fraction (50 μL) were analysed by SDS-PAGE (method 8) and iron content of fractions containing either ThiGH complex or monomeric ThiH was measure (method 14). ThiH-His fractions eluted from the S-200 were concentrated to 60 mg/mL without protein precipitation.

6.4.14 Initial studies on the crystallisation of ThiH

ThiGH complex and monomeric ThiH were expressed and separated using method 13 and 6.4.13. The iron content of the protein samples was measured (table 2.7).

Protein samples from the S200 column were placed inside an anaerobic glove box; fractions 24-26 of ThiGH complex were pooled and concentrated using a microcentrifuge (45 min, 9,000 rpm) to 28.3 mg/mL. Fraction 34 containing monomeric ThiH was concentrated to 60 mg/mL using a microcentrifuge (60 min, 9,000 rpm). All crystallisation experiments were carried out under strict anaerobic environment, in a controlled temperature room (18ºC). All solutions used for the experiment were filter sterilised.

ThiGH-His was expressed from pRL1020/BL21(DE3), purified (method 13), reconstituted, concentrated to 40 mg/mL. An analytical gel filtration chromatography was carried out under strictly anaerobic conditions (method 6.4.13) to separate ThiGH complex from monomeric ThiH. The iron content of the protein samples was measured (table 2.7). Fractions 24-26 of ThiGH complex were pooled and concentrated to 28.3 mg/mL. Fraction 34 of monomeric ThiH was concentrated to 60 mg/mL. These protein samples were then used to set up the crystallisation trial experiments. All crystallisation experiments were carried out under strict anaerobic environment, inside a glove box, in a controlled temperature room.

Screening for optimal crystal growth conditions of ThiGH and monomeric ThiH were prepared in 24 well plates. Precipitant solutions (table 6.9) were prepared and dispensed onto the plate (1000 μL each) using a pippeting robot (Beckman workstation). To each precipitant stock solution glycerol was added to 25% final concentration. The plates were then transferred to the glove box port compartment and degassed four times (four cycles of 3 min purge/15 min rest) and then transferred inside the glove box and allowed to degas for a further 24 hours.

Table 6.9 Precipitant solutions used for crystallisation screening

Tube	Precipitant	Volume added	Final concentration
1	Ammonium sulphate, 3.5 M	5.71	2.0 M
2	Ammonium sulphate, 3.5 M	7.14	2.5 M
3	Ammonium sulphate, 3.5 M	8.57	3.0 M
4	Lithium sulphate, 2.0 M	9.00	1.8 M
5	Lithium sulphate, 2.0 M	7.50	1.5 M
6	Lithium sulphate, 2.0 M	6.00	1.2 M
7	Sodium formate, 4.0 M	8.75	3.5 M
8	Sodium formate, 4.0 M	6.25	2.5 M
9	Sodium formate, 4.0 M	5.00	2.0 M
10	100% PEG 400	1.50	15 %
11	100% PEG 400	2.00	20 %
12	100% PEG 400	3.00	30 %
13	50% w/v PEG 1500	3.00	15 %
14	50% w/v PEG 1500	4.00	20 %
15	50% w/v PEG 1500	6.00	30 %
16	50% w/v PEG 4000	3.00	15 %
17	50% w/v PEG 4000	4.00	20 %
18	50% w/v PEG 4000	6.00	30 %
19	50% w/v PEG 8000	3.00	15 %
20	50% w/v PEG 8000	4.00	20 %
21	50% w/v PEG 8000	6.00	30 %
22	25% w/v PEG 20000	3.00	15 %
23	25% w/v PEG 20000	4.00	20 %
24	25% w/v PEG 20000	6.00	30 %
25	100% Isopropanol	0.20	2 %
26	100% Isopropanol	0.50	5 %
27	100% Isopropanol	1.00	10 %

Stock solutions (10 mM) of additives (tyrosine, AdoMet and Dxp), and buffers (Tris pH 8.8, 100mM and MOPS pH 7.7, 100 mM) were prepared aerobically, placed inside the glove box and allowed to degas.

Protein solutions were completed by the addition of further reagents (AdoMet, tyrosine or Dxp). Drops were prepared by placing drops of protein solution (1.5 μL) onto cover slips (22 mm diameter) and then adding precipitant solution from the well (1.5 μL). The cover slip was then carefully inverted to cover the rim of each well of the plate. Table 6.10 shows the protein, buffer and additives on each tray.

Table 6.10 Protein, additives and buffers used for crystallisation screening

Tray nº	**Protein/ volume added**	**Additives*/ volume added**	**Buffer**
1	ThiGH/30 μL	AdoMet/3.5 μL Dxp/1.75 μL Tyrosine/1.75 μL	MOPS pH 7.7
2	ThiGH/90 μL	AdoMet/9.0 μL Dxp/4.5 μL Tyrosine/4.5 μL	Tris pH 8.8
3	ThiH/45 μL	AdoMet/4.5 μL Dxp/2.25 μL Tyrosine/2.25 μL	MOPS pH 7.7

* Stock solutions of 10 mM.

6.5 Experimental for chapter 3

Flavodoxin (FldA) and Flavodoxin reductase (Fpr) was supplied by Dr. Kriek in our laboratory purification procedure is described in detail in reference[160]. EPR, radiochemical assays, TP course formation, LC-MS and modelling of the cleavage of tyrosyl radical are described in references[135,160,166].

6.5.1 Standard ThiGH Activity Assay

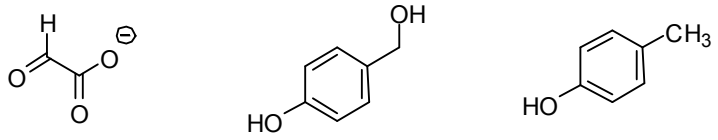
The following proteins and reagents were added, anaerobically, to a reaction mixture in the following order (with final concentrations): reconstituted and concentrated ThiGH (255 μM), tyrosine (0.51 mM), AdoMet (0.74 mM), NADPH (0.67 mM), FldA (62 μM), Fpr (10 μM). Negative control experiments were prepared using same reagents but lacking tyrosine. FldA, Fpr and NADPH were mixed and left for 15 min until solution turned blue. ThiGH, AdoMet and tyrosine solutions were mixed and after the reductant mixture was added. The reactions (7×200 μL aliquots) were incubated at 37 ºC for 5, 15, 30, 60, 90, and 120 min and subsequently stored at -80 ºC.

A standard method was developed for the analysis of these supernatants as follows. Samples were thawed and proteins were immediately precipitated with 20% perchloric acid (12 μL) and then cleared by centrifugation (14,000 rpm, 15 min), and the supernatants were analysed by HPLC using two different methods: AdoMet, tyrosine, and DOA (40 μL) analysed by direct UV-visible detection (λ_{ex}= 223 nm, λ_{em}= 280 nm). and glyoxylate (40 μL) derivatised using 2-oxoacid method[169] and analysed by HPLC using fluorescent detection (λ_{exc}= 340 nm, λ_{em}= 420 nm).

6.5.2 ^{13}C NMR spectra of ThiGH assays

Two complete ThiGH assays (each of 277 μL) were prepared as described previously, except that where possible materials were prepare in sodium phosphate buffer (50 mM, pH 7.5). Thus assays contained (at final concentrations): reconstituted ThiGH (444 μM), L-[U-^{13}C]-tyrosine (357 μM), AdoMet (hydrochloride salt, 655 μM), FldA (50 μM), Fpr (9 μM) and NADPH (540 μM). Two assays lacking the reducing system were prepared as negative controls. Solutions were kept in sealed Eppendorf vials and incubated for 2 hours at 37 ºC. To the each solution was added perchloric acid [15 μL, 20% (v/v)] and after 30 min in air, the precipitated protein was removed by centrifugation (15 min, 13000 rpm). The corresponding supernatants were combined and the volume adjusted to 550 μL with D_2O. ^{13}C NMR spectra (10240 scans) were recorded on a Bruker DXP 300 operating at 75 MHz. For the samples extracted into

organic solvents, the supernatants of ten assays were extracted with CDCl3 (500 µL per assay) and the combined organic phase was analysed by ^{13}CNMR spectrometry on a Bruker DPX 400 operating at 100 MHz.



Glyoxylate *p*-hydroxybenzyl alcohol *p*-cresol

Glyoxylate: ^{13}C NMR (75 MHz; aqueous phosphate buffer) δ 173.7 (COO), 86.8 (COHOH).

4-Hydroxybenzylalcohol: ^{13}C NMR (75 MHz; MeOD) δ 57.4 (COH, C-4), 133.0 (CCH_2, C-1), 129.3 (2CH), 115.6 (2CH), 64.6 (CH_2OH).

p-Cresol: ^{13}C NMR (100 MHz; $CDCl_3$) δ 153.6 (COH), 130.5 (2CH), 130.2 (CCH_3), 115.6 (2CH), 20.9 (CH_3).

Organic extract [U-^{13}C]-p-cresol: ^{13}C NMR (100 MHz; $CDCl_3$) δ 153.2 (t, *J* = 66 Hz, COH), 130.1 (m, 2CH, CCH_3), 115.0 (m, 2CH), 20.4 (m, CH_3).

Spectrum of assays after protein precipitation includes hydrated [U-^{13}C]-glyoxylate: ^{13}C NMR (75 MHz; sodium phosphate buffer, 50 mM, pH 7.5) δ 173.4 (d, J_{cc} = 69 Hz, COO), 86.4 (d, J_{cc} = 69 Hz, $C(OH)_2$).

Spectrum of negative control assays after protein precipitation contains L-[U-13C]-tyrosine: ^{13}C NMR (75 MHz; sodium phosphate buffer, 50 mM, pH 7.5) δ 174.2 (COO, d, *J* = 53 Hz, C1), 155.1 (COH, t, *J* = 64 Hz, C-7), 130.8 (2CH, t, *J* = 54 Hz, C-5 and C-9), 127.1 (CCH_2 (m, *J* = 54 Hz, C-4), 115.9 (2CH, t, *J* = 64, C-6 and C-8), 56.3 (CN, dd, *J*= 53 Hz, *J* = 33 Hz, C-2), 35.6 (CH_2, t, *J* = 38 Hz, C-3).

.

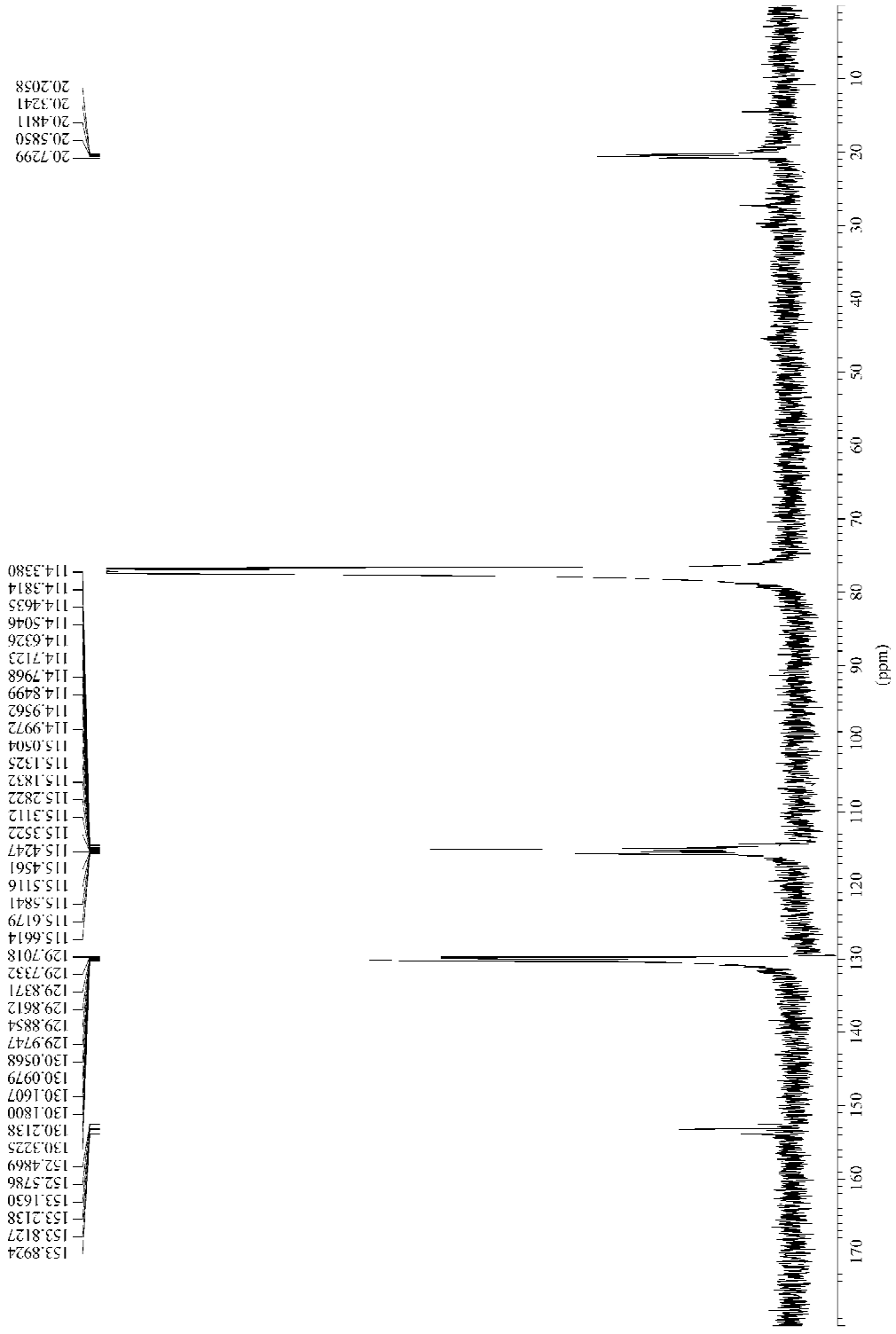


Figure 6.7 NMR spectra of organic extract of a ThiGH assay showing formation of ^{13}C labelled p-cresol

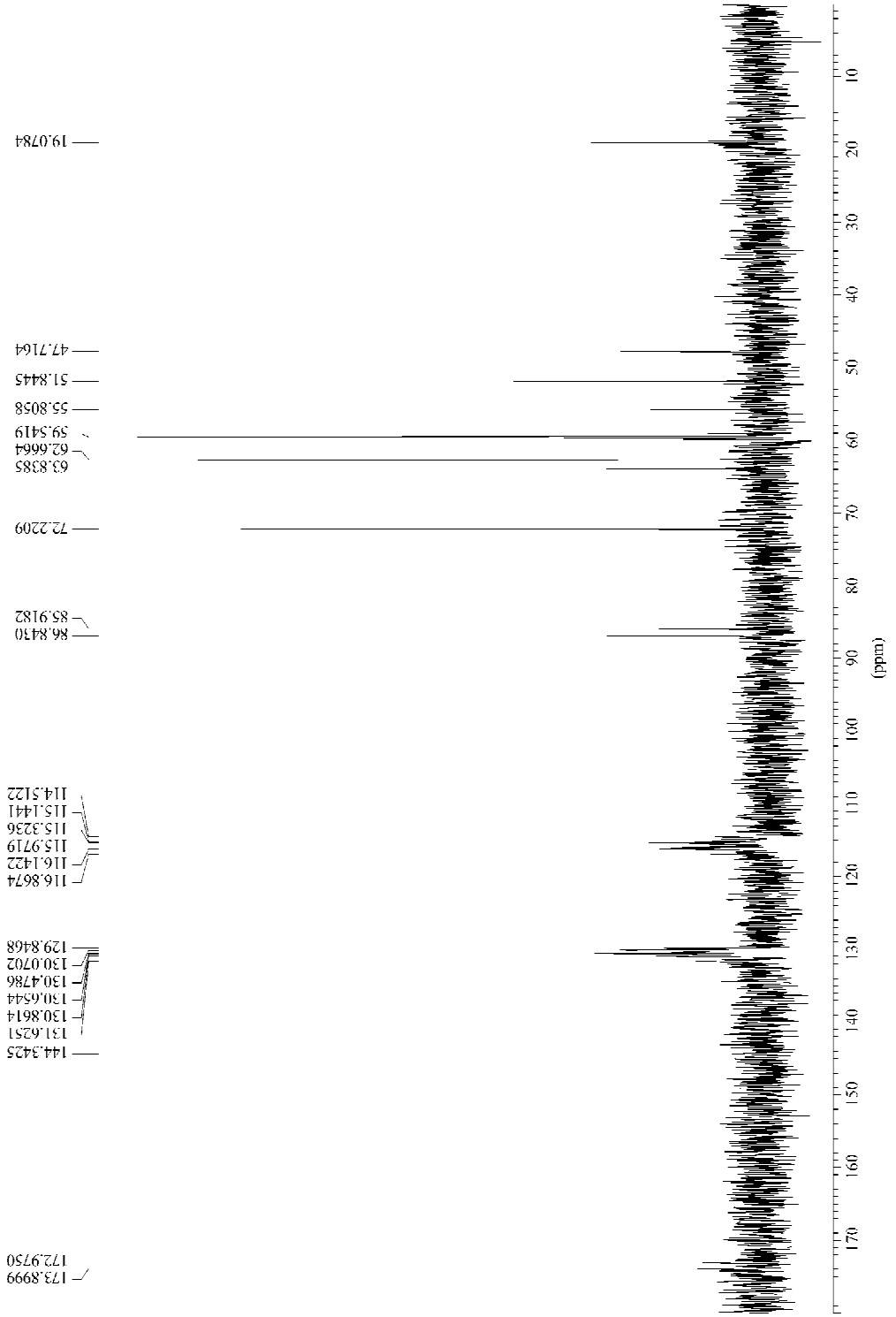


Figure 6.8 NMR spectra of cleared aqueous supernatant from active ThiGH Assay showing the formation of ^{13}C glyoxylate (hydrated form).

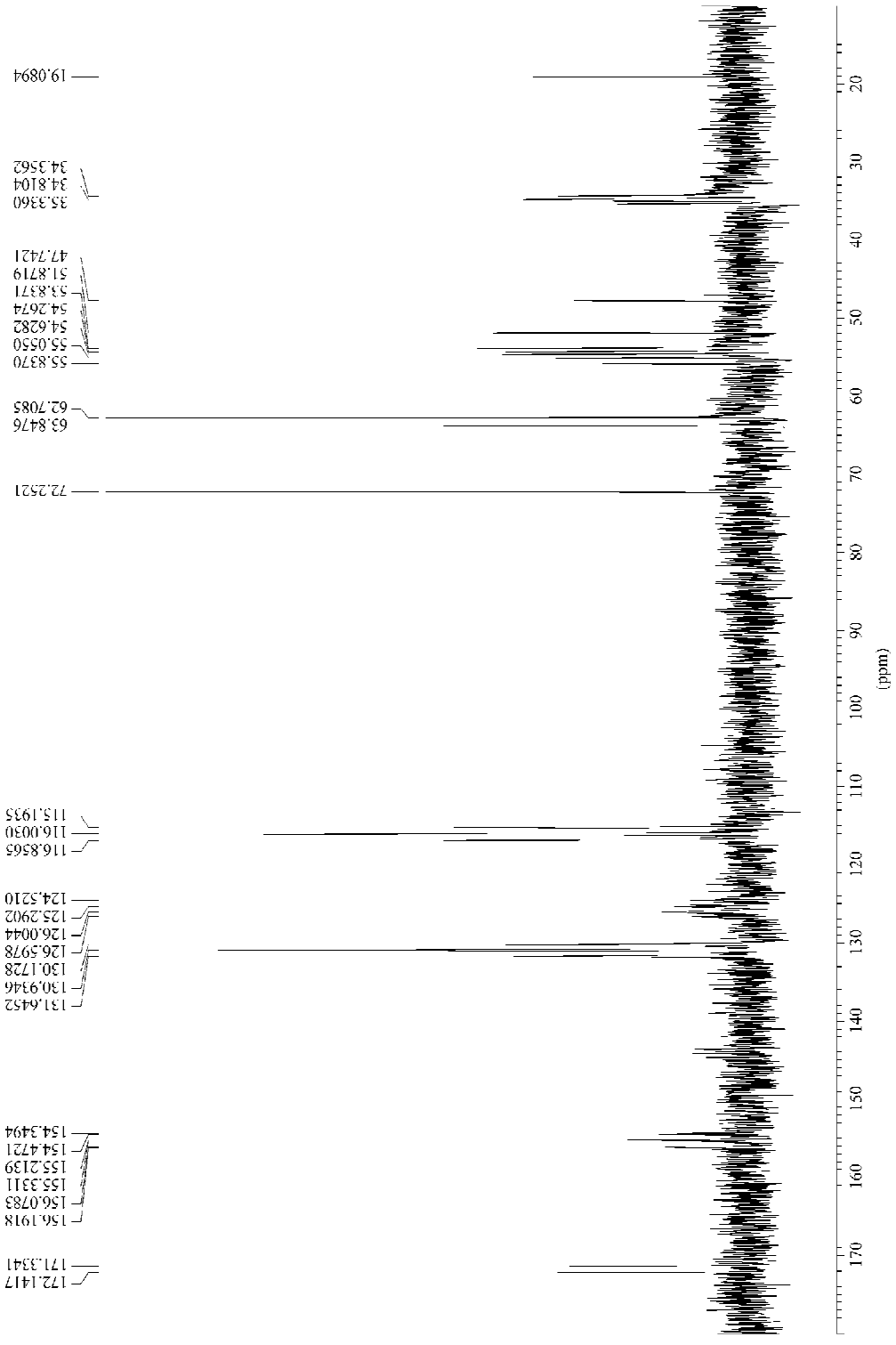


Figure 6.9 NMR spectra of cleared aqueous supernatant from negative control assay.

6.5.3 GC-MS analysis of ThiGH Assays

The organic extract prepared for ^{13}C NMR analysis was also analysed by GC-MS, essentially as described by Luo *et al*.[205].

GC-MS analysis data was obtained on a Thermo TraceMS equipped with an AS800 auto sampler and a PE-Wax (30 m x 0.25 mm x 0.25 μm) column. Helium was used as a carrier gas (flow 1 mL/min) and samples (1.0 μL) were injected with a speed of 10 (JL/s). The initial temperature of the column oven was 40 ºC for 4 min. This temperature was increased to 240 ºC at a speed of 20 ºC / min and maintained at this level for another 6 min. The detector was set at 70 eV EI-MS, the source temperature at 200 ºC with a trap current 150 JA and using the full scan acquisition mode (2 scans/s from 20-500 amu).

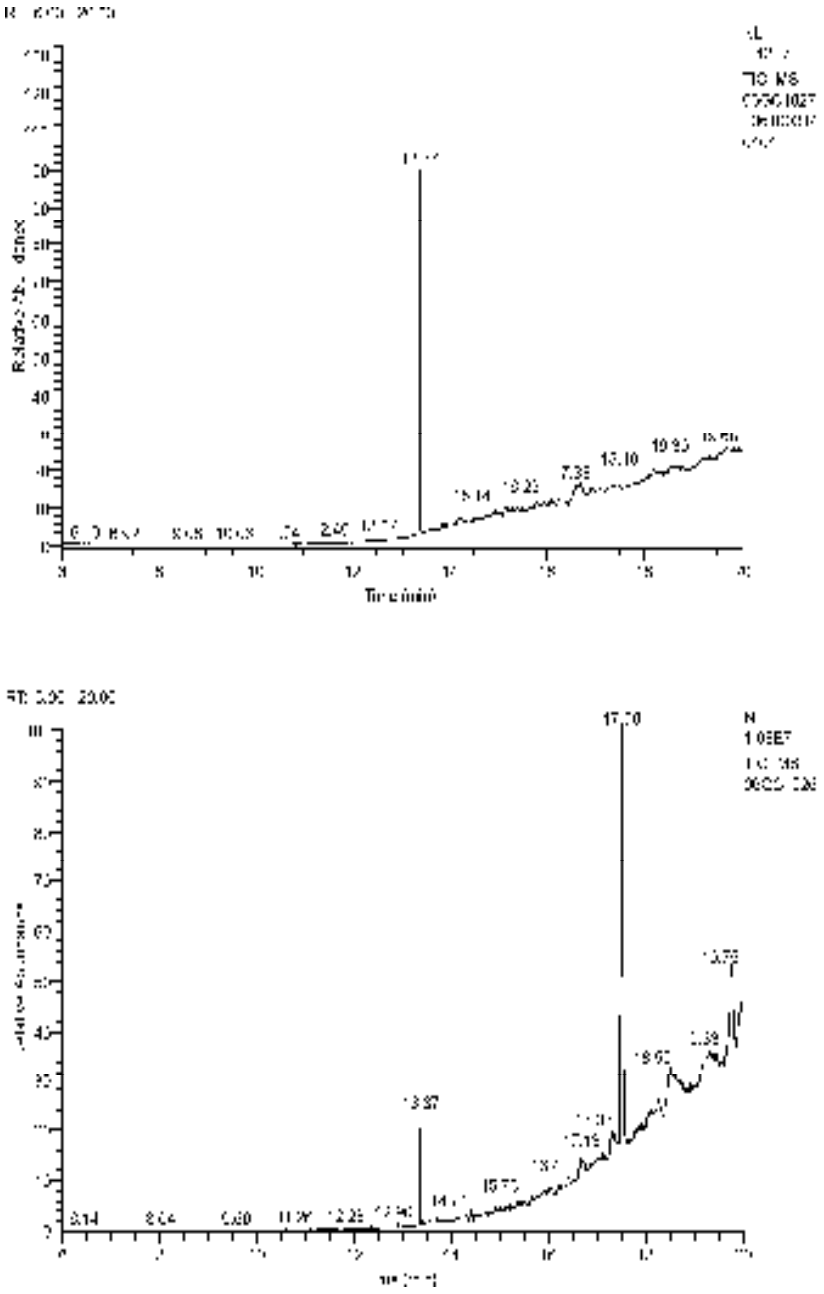


Figure 6.10 GC-MS Traces from p-Cresol analysis. Top: p-cresol standard. Bottom: organic extract of a ThiGH assay.

p-Cresol standards: calculated for C_7H_7OH [M+H]: 108.14; found 108.2. EI m/z (%): 107.1(100), 108.2(72), 77.1(30), 28.2(30), 79.1(26), 43.1(18), 57.1(17), 55.1(16), 51.1(15), 80.2(12), 53.1(12), 49.9(12), 39.1(12), 32.1(11), 109.2 (8.9).

^{13}C p-cresol from assay: calculated for $^{13}C_7H_7OH$ [M+H]: 115.14; found 115.2. EI m/z (%) : 114.1(100), 115.2(55), 83.1(53), 85.2(45), 55.2(45), 57.2(33), 28.2(29), 86.2(28), 54.1(28), 56.2(27), 42.2(23), 43.2(22), 29.2(21), 41.1(19), 84.2(18), 68.2(16), 67.1(12), 69.2(11), 97.2(10), 71.2(10).

6.5.4 Quantification of *p*-cresol, AdoH and tyrosine by HPLC

Method 1

RP-HPLC analysis was carried out using a wavelength detector set to 280 nm and a C18 column (Hypersil, 5μ, 150 x 4.6 mm) equilibrated in solvent A (0.1% acetic acid). The column was eluted at 0.8 ml/min for 6 min with solvent A, followed by a gradient to 100% solvent B (0.1% acetic acid in MeOH) over 18 min, kept at 100% solvent B for 2 min and returned to starting condition in 3 min and then re-equilibrated for 10 min. The following retention times were observed for standards: *p*-cresol, R_t =14.0 min; hydroxybenzyl alcohol, R_t = 11.4 min.

Method 2

RP-HPLC analysis was carried out using a dual wavelength detector set to 230 nm and 280 nm and a C18 column (Phenomenex Prodigy 5μ, ODS-2, 150 x 4.6 mm) equilibrated in solvent A (0.1% trifluoroacetic acid). The column was eluted at 0.8 ml/min for 6 min with solvent A, followed by a gradient to 100% solvent B (0.1% trifluoroacetic acid in Acetonitrile) over 27 min, kept at 100% solvent B for 1 min and then re-equilibrated for 8 min. The following retention times were observed for standards: hydroxybenzyl alcohol, R_t = 21 min, tyrosine, R_t = 16.5 min, AdoH, Rt = 17.4 min.

Method 3

RP-HPLC analysis was carried out using a dual wavelength detector set to 215 and 260 nm using a Synergy Polar-RP column (4μ, 150 x 4.6 mm) pre-equilibrated in solvent A (water with 0.1% formic acid). The flow rate was maintained at 0.8 mL/min using a binary solvent mixture. An aliquot (40 μL) was injected onto the column eluted with solvent A for 6 min followed by a gradient to 100% solvent B (50% acetonitrile, 50% solvent A) over 20 min, followed by a 5-min gradient to 10% solvent A, 5 min isocratic

at 10% solvent A, and a 4-min gradient to 100% solvent A, and re-equilibrated for 15 min isocratic at 100% solvent A. This method did not resolve the reaction products.

Method 4

RP-HPLC analysis was carried out using a dual wavelength detector set to 215 nm and 260 nm and a C18 column (Phenomenex Prodigy 5µ, ODS-3, 250 x 4.6 mm) equilibrated in solvent A (0.1% TFA). The column was eluted at 0.8 ml/min with solvent B (Acetonitrile, 0.1% trifluoroacetic acid), using the same elution profile as method 3. The following retention times were observed for standards: tyrosine, R_t = 17 min, AdoH, R_t = 17.4 min.

Method 5

RP-HPLC (UV-visible detection at 254 nm) using a C18 RP column (Phenomenex Prodigy 5µ, 250 x 5.0 mm) pre-equilibrated in solvent A (25 mM acetic acid with 8 mM sodium octanesulfonate, 2 mL/min). An aliquot (40 µL) was injected onto the column eluted with solvent A for 3 min followed by a gradient to 100% solvent B (50% acetonitrile, 50% solvent A) over 15 min, kept at 100% solvent B for 2 min, reversed to 0% solvent B over 2 min and re-equilibrated for 10 min. The retention time (R_t) for AdoMet was 26 min.

Method 6

An aliquot of supernatant (40 µL) was analysed by HPLC using a dual wavelength detector set to 223 nm and 280 nm and a C18 RP column (Phenomenex Prodigy 5µ, ODS-2, 150 x 4.6 mm) equilibrated in solvent A (0.1% acetic acid). The column was eluted at 0.8 mL/min for 4 min with solvent A, followed by a gradient to 100% solvent B (0.1% acetic acid in MeOH) over 18 min, kept at 100% solvent B for 2 min and returned to starting condition in 3 min and then re-equilibrated for 5 min. The following retention times were observed for standards: tyrosine, R_t = 8.3 min; 5'-deoxyadenosine (AdoH), R_t = 11.8 min; *p*-hydroxybenzyl alcohol, R_t = 12.6 min; *p*-cresol, R_t = 16.7 min.

A range of concentrations of standards (0-80 nmoles in 40 µL) were used to construct calibration curves (figure 6.11, 6.12 and 6.13) for estimating the concentrations of

tyrosine, AdoH and *p*-cresol. Concentrations were determined from experiments carried out in duplicate. The data was fitted to an exponential process and the first order rate constant and final concentrations calculated using Sigmaplot.

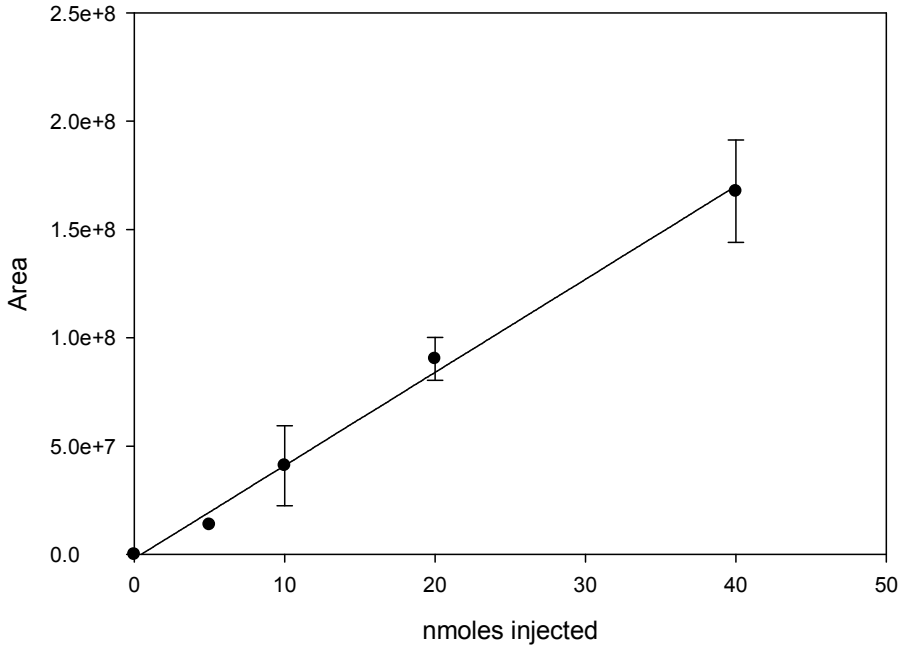


Figure 6.11 Calibration curve for p-cresol standards. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.9782, yo=9.401E+6, a=4.4E+6.

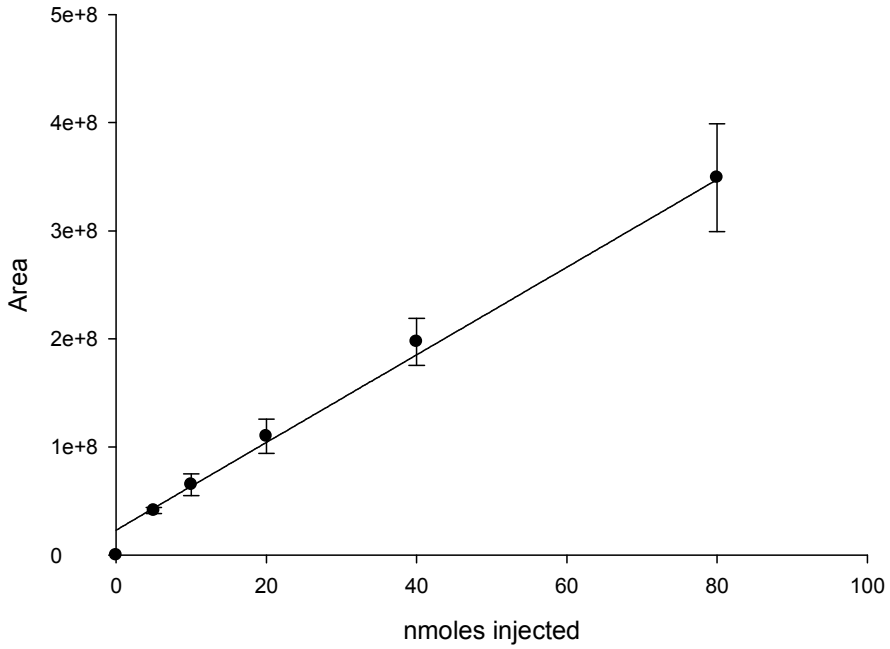


Figure 6.12 Calibration curve for AdoH standards. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.98, y0=2.3E+7, a=4E+6.

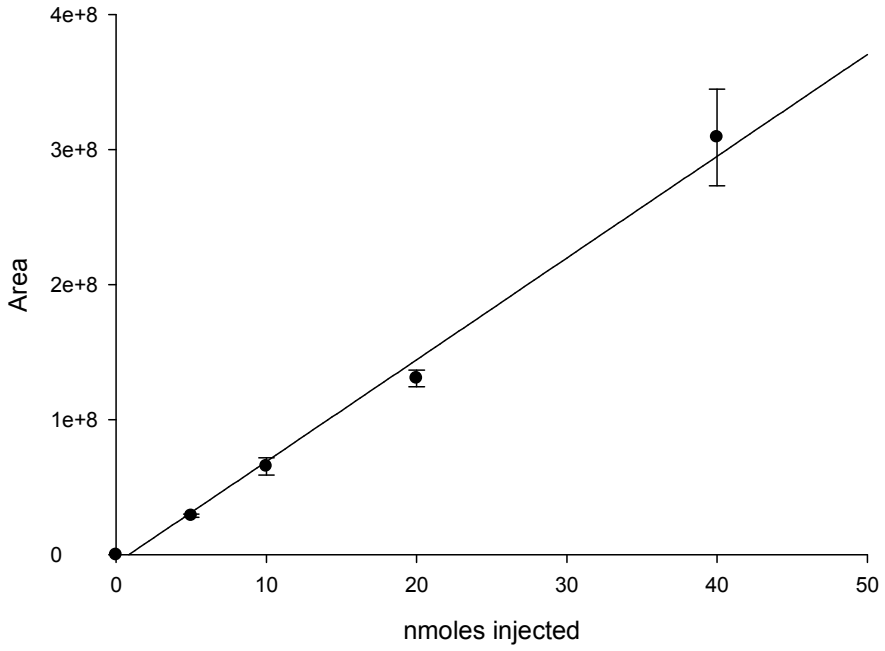


Figure 6.13 Calibration curve for tyrosine standards. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.99, y0=-6.3E+6, a=7.5E+6.

6.5.5 Quantification of glyoxylate by HPLC

The glyoxylic acid content of the samples was estimated using a method devised for the detection of 2-oxoacids[169] in which the 2-oxoacid is converted to the fluorescent 2-quinoxalinol derivative.

An aliquot of cleared supernatant (10 μL) was mixed with HEPES (40 μL, 50 mM, pH 7.5) and this sample was acidified with HCl (100 μL, 0.5 M) and ortho-phenylene diamine (50 μL, 10 mg/mL in 0.5 M HCl) was added. The mixture was heated to 95 ºC for 10 min in a heat-block, left at room temperature for 2 min and NaOH (120 μL, 1.25 M) was added. A calibration was obtained by derivatisation of 0.5 mM glyoxylic acid (0; 2.5; 5; 7.5; 10; 15 and 20 μL) diluted in 50 mM HEPES (pH 7.5) to a constant volume (50 μL) as described for the samples above.

Derivatised samples were analysed by HPLC using a Hypersil BDS C8 (5μ, 150 x 4.6 mm) column pre-equilibrated in solvent C (100 mM ammonium bicarbonate, 1 ml/min). An aliquot (40 μL) was injected onto the column eluted with solvent C for 5 min before a gradient to 80% solvent D (acetonitrile) over 15 min was applied. The gradient was reversed to 0% Solvent D over 5 min, and the column was re-equilibrated for 10 min before the next injection. The 2-quinoxalinol (R_t 11.8 min) was detected using fluorescent detection (λ_{ex} 340 nm, λ_{em} 420 nm).

A range of concentrations of standards (0-40 μM) were used to construct calibration curves for estimating the concentrations of glyoxylate. The standards used were 2-quinoxalinol (figure 6.14) and derivatised glyoxylate (figure 6.15), using method described above. By using this calibration it was possible to determine, for each experiment, the yield of the derivatisation reaction and concentrations were determined from experiments carried out in duplicate. The data were fitted to an exponential process and the first order rate constant and final concentrations calculated using Sigmaplot.

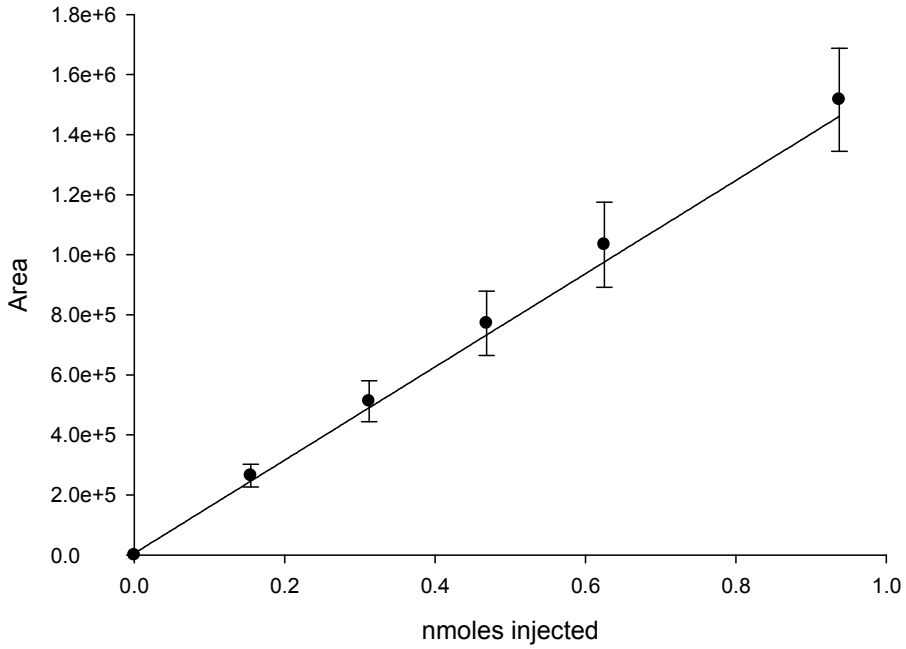


Figure 6.14 Calibration curve for 2-quinoxalinol standards. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.99, y0=5.6E^{+3}, a=1.5E^{+6}.

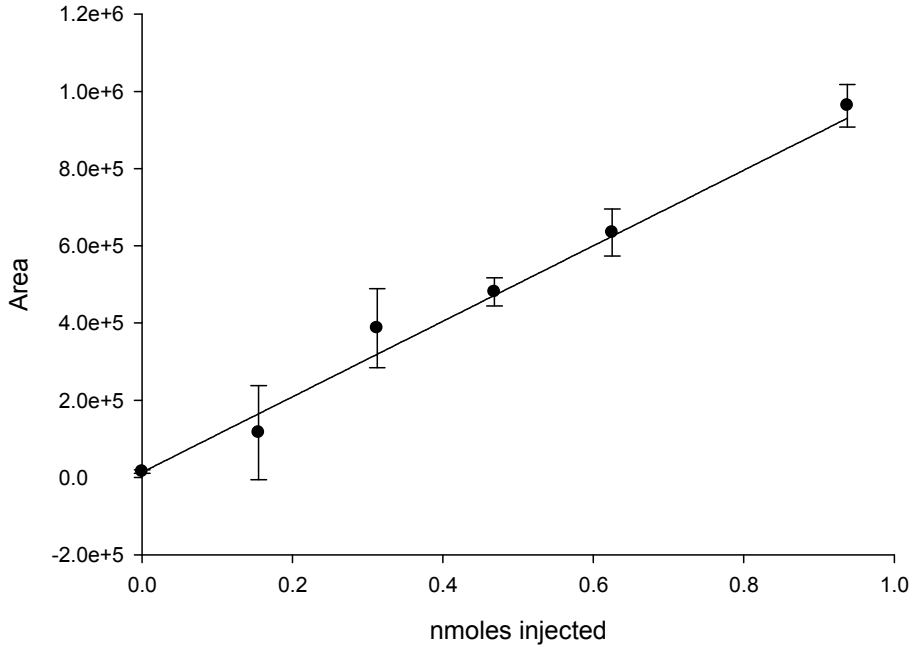


Figure 6.15 Calibration curve for derivatised glyoxylate standards. Data fitted using Sigmaplot to a linear function (y=ax+y0): R=0.97, y0=1.34E^{+4}, a=9.78E^{+5}.

6.6 Experimental for chapter 4

6.6.1 Assembly of pFM001 and pFM101

D. vulgaris thiSGH (2.381 kb) was amplified by PCR (method 1) using genomic DNA from *Desulfovibrio vulgaris* and two different sets of primers: ThiS Fwd/ThiS Rev. The sequences for the oligonucleotides used in this reaction are listed in table 6.11.

The PCR product (*thiSGH*) was restricted with *PciI*/*SacI* and cloned into pRL1020 on a *NcoI*/*SacI* fragment. Both, digested pRL1020 and PCR product were loaded onto a 1% agarose gel (appendix B) and analysed by gel electrophoresis (figure 6.16). The PCR product was then ligated into pRL1020 backbone using T4 ligase (method 7). Ligation reactions were incubated overnight at 4 ºC and then used to transform *E. coli* TOP10 competent cells (method 2). Positive colonies were then identified by analytical restriction digestion using *Xho*I (figure 6.17, method 5). The resultant vector was named pFM001.

Table 6.11 Sequences of oligonucleotides used in polymerase chain reaction.

Primers	Sequence 5′ → 3′
*ThiS Fwd**	CTA GCT AGC AAC ATG TTG ACG ATA GTT GTG AAC GGT
*ThiS Rev**	GGA ATT CGA GCT CTC AAT GGT GGT GAT GGT GGT GTC GGG CCT CGC ATC GGA C

* primers used to assemble plasmid encoding for *ThiSGH-His*

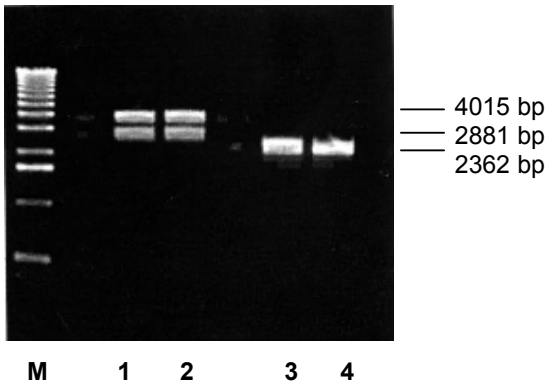


Figure 6.16 Identification of restriction digestion of PCR product and pRL1020. Lane 1: pRL1020 restricted with NcoI/SacI (4015 bp: pRL1020 backbone, 2881 bp: ThiSGH-His section of pRL1020); Lane 2 and 3: PCR of thiSGH-His restricted with PciI/SacI (2362 bp); M: molecular weight marker.

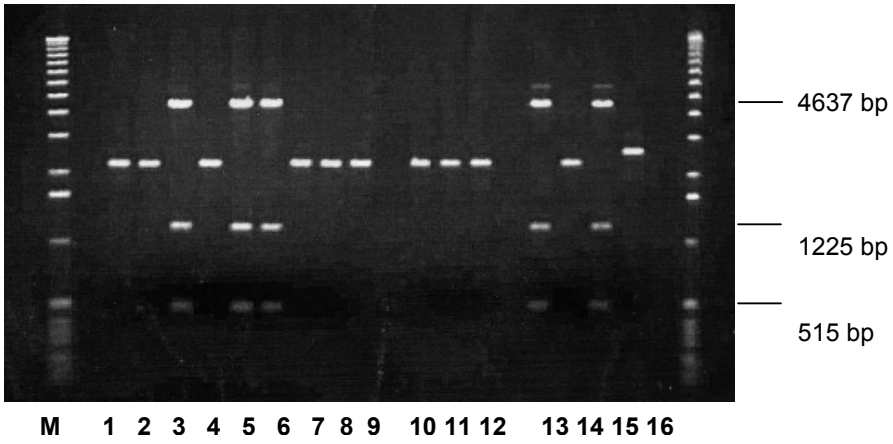


Figure 6.17 Identification of positive clones containing pFM001 by plasmid DNA isolation and restriction. Lanes 1-16: restriction of pFM001 with XhoI (4637, 1225 and 515 bp); M: molecular weight marker.

pRL1021, a plasmid bearing the *iscSUA-hscBA-fdx* operon, was restricted with *EcoRI*/*SacI* to insert the *iscSUA-hscBA-fdx* fragment into equally restricted pFM001 (method 7). The ligation reactions were incubated overnight at 4 ºC and then used to transform TOP10 competent cells (method 2). Positive colonies were identified by

analytical restriction digestion using *PstI*/*NdeI* (figure 6.18, method 5). The new vector was named pFM101.

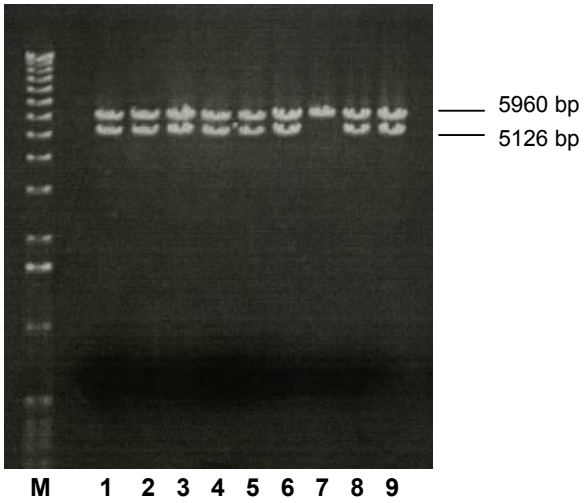


Figure 6.18 Identification of positive clones containing pFM101 by plasmid DNA isolation and restriction. Lanes 1-9: restriction of pFM101 with PstI/NdeI (5960 bp, 5126 bp); M: molecular weight marker.

Table 6.12 contains a list of all the plasmids that were assembled. All the plasmid maps are in appendix A.

Table 6.12 List of assembled plasmids using D. vulgaris genes.

Name	Insert	Parental plasmid
pFM001	*PciI/ SacI, thiSGH*	pRL1020
pFM101	*NotI/ EcoRI, iscSUA-hscBA-fdx*	pFM001

6.6.2 Expression and purification of pFM001 and pFM101

pFM001/BL21(DE3) was cultured aerobically on a large scale (method 11), induced with arabinose and grown at 27 ºC for 4 hours before being harvested. The cell paste was weight and stored at -80 ºC until further use.

ThiH was isolated using optimised purification method (method 13) and resultant fractions analysed by SDS-Page gel analysis (method 8). pFM101 was expressed and purified using the same procedures.

6.6.3 Mutagenesis of pMK024

pMK024 was mutated at the *NcoI* site using QuikChange® Multi Site-Directed Mutagenesis Kit from Stratagene, the resulting plasmid was named pFM024.

Table 6.13 Sequences of oligonucleotides used in the NcoI site mutagenesis.

Primers	**Sequence 5′ → 3′**
Forward	AG ATG AGT GTG GTT GCG GCG
Reverse	GTA ATC CAT CGT TAA TTA CTC CTG TTA CT

6.6.4 Assembly of pFM003 and pFM103

Two plasmids were assembled using genomic DNA from *Thermotoga maritima.*

T. maritima ThiH (1.4 Kb) was amplified by PCR (method 1) using two different sets of primers: ThiH Fwd and ThiH Rev in order to replace the *E. coli* ThiH present in pRL1020. The oligonucleotides sequences are listed in table 6.14.

Table 6.14 Sequences of oligonucleotides used in polymerase chain reaction.

Primers	Sequence 5′ → 3′
*ThiH Fwd**	GGC GGC GCC ATG GGC TGT ATG TAT GTG TTT GTG
*ThiH Rev**	GCC GCC GTC GAC TTA GTG GTG GTG GTG GTG GTG AAA TCT GAC ATC CCT CTC

The PCR product was loaded onto a 1% agarose gel (appendix B) and analysed by gel electrophoresis (figure 6.19). *D. vulgaris thiSGH* PCR amplification was carried out as a positive control.

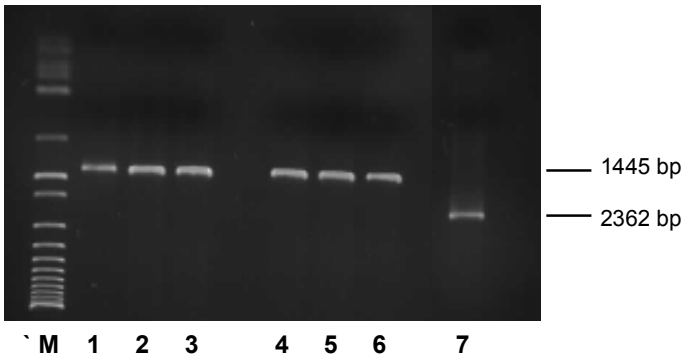


Figure 6.19 Identification of restriction digestion of PCR product and pRL1020. Lane 1-6: T. maritima ThiH (1445 bp); Lane 7: D. vulgaris ThiSGH PCR product (2362 bp); M: molecular weight marker.

The PCR product (ThiH) was restricted with *NcoI*/*SalI*, cloned into pBAD-HisA and pFM024 on a *NcoI*/*XhoI* fragment. Restricted pBAD-HisA, pFM024 and PCR product were analysed by gel electrophoresis (figure 6.20).

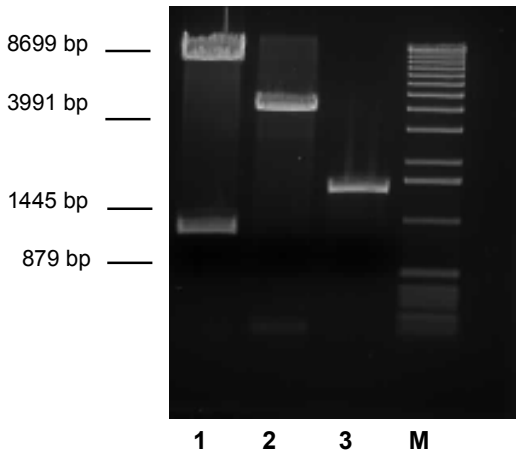


Figure 6.20 Identification of restriction digestion of PCR product, pBAD-HisA and pFM024. Lane 1: pFM024 restricted with NcoI/XhoI (backbone: 8699 bp, LipA section: 879 bp); Lane 2: pBAD-HisA restricted with NcoI/XhoI (backbone: 8699 bp, 113 bp); Lane 3: PCR ThiH restricted with NcoI/SalI (1445 bp); M: molecular weight marker.

The PCR product was ligated into pBAD-HisA and pFM024 using T4 ligase (method 7). Ligation reactions were incubated overnight at 4 ºC and then used to transform *E. coli* TOP10 competent cells (method 2). Positive colonies were then identified by analytical restriction digestion. Potential colonies containing the *thiH* operon were restricted using *PciI*/*EcoRI*, *EcoRI*/*NcoI* or just *EcoRI* (figure 6.21, method 5). To identify colonies containing the *T. maritima thiH-isc* operon colonies were restricted using *EcoRI* and *EagI*/*NcoI* (figure 6.22, method 5). The resultant vectors were named pFM003 and pFM103. Table 6.15 contains a list of all the plasmids that were assembled. All the plasmid maps are in appendix A.

Table 6.15 List of assembled plasmids using T. maritima genes.

Name	Insert	Parental plasmid
pFM003	*NcoI*/*SalI*, *thiH*	pBAD-HisA (*NcoI*/*XhoI*)
pFM103	*NcoI*/*SalI*, *thiH*	pFM024 (*NcoI*/*XhoI*)

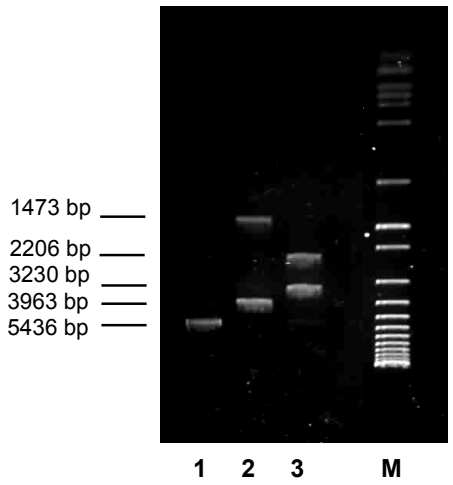


Figure 6.21 Identification of positive colonies containing pFM003. Lane 1: pFM003 restricted with EcoRI (5436 bp); Lane 2: pFM003 restricted with NcoI/EcoRI (1473 bp, 3963 bp); Lane 3: pFM003 restricted with PciI/EcoRI (2206 bp, 3230 bp); M: molecular weight marker.

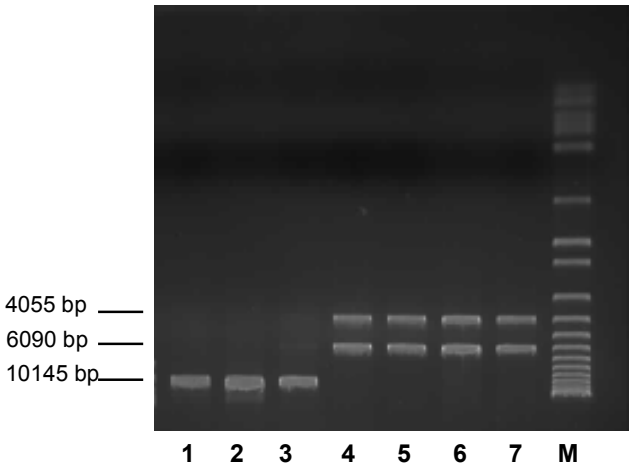


Figure 6.22 Identification of positive colonies containing pFM103. Lane 1-3: pFM103 restricted with EcoRI (10145 bp); Lane 4-7: pFM103 restricted with EagI/NcoI (4055 bp, 6090 bp); M: molecular weight marker.

6.6.5 Expression and purification of pFM003 and pFM103

pFM003/BL21(DE3) was cultured aerobically on a large scale (method 11), induced with arabinose and grown at 27 ºC for 4 hours before being harvested. The cell paste was weight and stored at -80 ºC until further use.

ThiH was isolated using optimised purification method (method 13) and resultant fractions analysed by SDS-Page gel analysis (method 8). pFM103 was expressed and purified using the same procedures.

REFERENCES

(1) Singleton, C. K.; Martin, P. R. *Cur. Mol. Med.* 2001, *1*, 197-207.

(2) Bender, D. A. *Proc. Nutr. Soc.* 1999, *58*, 427-33.

(3) Harper, C. *Eur. J. Neurol.* 2006, *13*, 1078-82.

(4) Jansen, B. C. P. *The Vitamins Chemistry, Physiology, Pathology, Methods*; 2nd ed.; Academic Press: New York, 1972; Vol. 5.

(5) Williams, R. R., Spies, R. D. *Vitamin B1 (Thiamin) and its Use in Medicine* MacMillan: New York, 1938.

(6) Spenser, I. D.; White, R. L. *Angew. Chem. Int. Ed. Engl.* 1997, *36*, 1033-1046.

(7) Nelson, D., Cox, M. *Lehninger Principles of Biochemistry*; 3 ed.; Worth Publishers: United States of America, 2000.

(8) Jordan, F. *Nat. Prod. Rep.* 2003, *20*, 184-201.

(9) Washabaugh, M. W.; Jencks, W. P. *Biochemistry* 1988, *27*, 5044-53.

(10) Page, M.; Williams, A. In *Organic and Bio-organic Mechanisms*; Addison Wesley Longman: Harlow, 1997, p 169-171.

(11) Stryer, L. In *Biochimica*; 4th ed.; Zanichelli, Ed. Bologna, 1996, p 598-606.

(12) Breslow, R. *J. Am. Chem. Soc.* 1958, *80*, 3719-3726.

(13) Astashkin, A. V.; Seravalli, J.; Mansoorabadi, S. O.; Reed, G. H.; Ragsdale, S. W. *J. Am. Chem. Soc.* 2006, *128*, 3888-9.

(14) Furdui, C.; Ragsdale, S. W. *Biochemistry* 2002, *41*, 9921-37.

(15) White, R. L.; Spenser, D. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*; Neidhardt, F. C., Curtiss, R. R., Ingraham, J. L., Lin, E. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E., Eds.; American Society for Microbiology: Washington, D.C., 1996; Vol. 1, p 1175-1209.

(16) Estramareix, B.; David, S. *Biochim. Biophys. Acta* 1990, *1035*, 154-160.

(17) Estramareix, B.; Therisod, M. *J. Am. Chem. Soc.* 1984, *106*, 3857-3860.

(18) Newell, P. C.; Tucker, R. G. *Biochemical Journal* 1968, *106*, 271-277.

(19) Himmeldirk, K.; Sayer, B. G.; Spenser, I. D. *J. Am. Chem. Soc.* 1998, *120*, 3581-3589.

(20) Yamada, K.; Morisaki, M.; Kumaoka, H. *Biochim. Biophys. Acta* 1983, *756*, 41-8.

(21) Lawhorn, B. G.; Mehl, R. A.; Begley, T. P. *Org. Biomol. Chem.* 2004, *2*, 2538-46.

(22) Zeidler, J.; Sayer, B. G.; Spenser, I. D. *J. Am. Chem. Soc.* 2003, *125*, 13094-105.

(23) Gralnick, J.; Webb, E.; Beck, B.; Downs, D., *J. Bacteriol.* 2000, *182*, 5180-7.

(24) Leonardi, R.; Fairhurst, S. A.; Kriek, M.; Lowe, D. J.; Roach, P. L. *FEBS Lett.* 2003, *539*, 95-9.

(25) Begley, T. *Faseb J.* 1999, *13*, A1426-A1426.

(26) Palenchar, P. M.; Buck, C. J.; Cheng, H.; Larson, T. J.; Mueller, E. G. *J. Biol. Chem.* 2000, *275*, 8283-6.

(27) Lehmann, C.; Begley, T. P.; Ealick, S. E. *Biochemistry* 2006, *45*, 11-9.

(28) Lauhon, C. T.; Kambampati, R. *J. Biol. Chem.* 2000, *275*, 20096-103.

(29) Xi, J.; Ge, Y.; Kinsland, C.; McLafferty, F. W.; Begley, T. P. *Proc. Natl. Acad. Sci. U S A* 2001, *98*, 8513-8.

(30) Taylor, S. V.; Kelleher, N. L.; Kinsland, C.; Chiu, H. J.; Costello, C. A.; Backstrom, A. D.; McLafferty, F. W.; Begley, T. P. *J. Biol. Chem.* 1998, *273*, 16555-60.

(31) Settembre, E. C.; Dorrestein, P. C.; Zhai, H.; Chatterjee, A.; McLafferty, F. W.; Begley, T. P.; Ealick, S. E. *Biochemistry* 2004, *43*, 11647-57.

(32) Wang, C.; Xi, J.; Begley, T. P.; Nicholson, L. K. *Nat. Struct. Biol.* 2001, *8*, 47-51.

(33) Appleyard, M. V.; Sloan, J.; Kana'n, G. J.; Heck, I. S.; Kinghorn, J. R.; Unkles, S. E. *J. Biol. Chem.* 1998, *273*, 14869-76.

(34) Hochstrasser, M. *Nat. Cell Biol.* 2000, *2*, E153-7.

(35) Lake, M. W.; Wuebbens, M. M.; Rajagopalan, K. V.; Schindelin, H. *Nature* 2001, *414*, 325-9.

(36) Rudolph, M. J.; Wuebbens, M. M.; Rajagopalan, K. V.; Schindelin, H. *Nat. Struct. Biol.* 2001, *8*, 42-6.

(37) Pickart, C. M. *Annu. Rev. Biochem*. 2001, *70*, 503-33.

(38) Zheng, N.; Wang, P.; Jeffrey, P. D.; Pavletich, N. P. *Cell* 2000, *102*, 533-9.

(39) Chiu, H. J.; Reddick, J. J.; Begley, T. P.; Ealick, S. E. *Biochemistry* 1999, *38*, 6460-6470.

(40) Dorrestein, P. C.; Zhai, H.; McLafferty, F. W.; Begley, T. P. *Chem. Biol.* 2004, *11*, 1373-81.

(41) Dorrestein, P. C.; Huili Zhai, H.; Taylor, S. V.; McLafferty, F. W.; Begley, T. P. *J. Am. Chem. Soc.* 2004, *126*, 3091-6.

(42) Frey, P. A.; Hegeman, A. D.; Ruzicka, F. J. *Crit. Rev. Biochem. Mol. Biol.* 2008, *43*, 63-88.

(43) Park, J. H., Dorrestein, P. C., Zhai, H., Kinsland, C., McLafferty,F. W., Begley, T. P. *Biochemistry* 2003, *42*, 12430-12438.

(44) Settembre, E. C., Dorrestein, P. C., Zhai, H., Chatterjee, A., McLafferty, F. W., Begley, T. P., Ealick, S.E. *Biochemistry* 2004, *43*, 11647-11657.

(45) Chatterjee, A.; Jurgenson, C. T.; Schroeder, F. C.; Ealick, S. E.; Begley, T. P. *J. Am. Chem. Soc.* 2006, *128*, 7158-9.

(46) Chatterjee, A.; Jurgenson, C. T.; Schroeder, F. C.; Ealick, S. E.; Begley, T. P. *J. Am. Chem. Soc.* 2007, *129*, 2914-22.

(47) Godoi, P. H.; Galhardo, R. S.; Luche, D. D.; Van Sluys, M. A.; Menck, C. F.; Oliva, G. *J. Biol. Chem.* 2006, *281*, 30957-66.

(48) Jurgenson, C. T.; Chatterjee, A.; Begley, T. P.; Ealick, S. E. *Biochemistry* 2006, *45*, 11061-70.

(49) Begley, T. P.; Chatterjee, A.; Hanes, J. W.; Hazra, A.; Ealick, S. E. *Curr. Opin. Chem. Biol.* 2008, *12*, 118-25.

(50) Webb, M. E.; Marquet, A.; Mendel, R. R.; Rebeille, F.; Smith, A. G. *Nat. Prod. Rep.* 2007, *24*, 988-1008.

(51) Leonardi, R.; Roach, P. L. *J. Biol. Chem.* 2004, *279*, 17054-62.

(52) Martinez-Gomez, N. C.; Robers, M.; Downs, D. M. *J. Biol. Chem.* 2004, *279*, 40505-10.

(53) Sprenger, G. A.; Schorken, U.; Wiegert, T.; Grolle, S.; de Graaf, A. A.; Taylor, S. V.; Begley, T. P.; Bringer-Meyer, S.; Sahm, H. *Proc. Natl. Acad. Sci. U S A* 1997, *94*, 12857-62.

(54) Vander Horn, P. B.; Backstrom, A. D.; Stewart, V.; Begley, T. P. *J. Bacteriol.* 1993, *175*, 982-92.

(55) Wright, C. M.; Christman, G. D.; Snellinger, A. M.; Johnston, M. V.; Mueller, E. G. *Chem. Commun. (Camb)* 2006, 3104-6.

(56) Skovran, E.; Downs, D. M. *J. Bacteriol.* 2000, *182*, 3896-903.

(57) Skovran, E.; Downs, D. M. *J. Bacteriol.* 2003, *185*, 98-106.

(58) Leonardi, R. PhD Thesis, University of Southampton, 2004.

(59) Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. *Nucleic Acids Res.* 2001, *29*, 1097-106.

(60) Miller, J. R.; Busby, R. W.; Jordan, S. W.; Cheek, J.; Henshaw, T. F.; Ashley, G. W.; Broderick, J. B.; Cronan, J. E.; Marletta, M. A. *Biochemistry* 2001, *40*, 151.

(61) Frey, M.; Rothe, M.; Wagner, A. F.; Knappe, J. *J. Biol. Chem.* 1994, *269*, 12432-7.

(62) Kulzer, R.; Pils, T.; Kappl, R.; Huttermann, J.; Knappe, J. *J. Biol. Chem.* 1998, *273*, 4897-903.

(63) Fontecave, M. *Abstracts of Papers of the Am. Chem. Soc.* 2000, *220*, 163-INOR.

(64) McIver, L.; Baxter, R. L.; Campopiano, D. J. *J. Biol. Chem.* 2000, *275*, 13888-94.

(65) Ollagnier-de-Choudens, S.; Sanakis, Y.; Hewitson, K. S.; Roach, P. L.; Baldwin, J. E.; Munck, E.; Fontecave, M. *J. Biol. Inorg. Chem.* 2000, *39*, 4165-73.

(66) Miller, J. R.; Busby, R. W.; Jordan, S. W.; Cheek, J.; Henshaw, T. F.; Ashley, G. W.; Broderick, J. B.; Cronan, J. E.; Marletta, M. A. *Biochemistry* 2000, *39*, 15166-15178.

(67) Henshaw, T. F.; Cheek, J.; Broderick, J. B. *J. Am. Chem. Soc.* 2000, *122*, 8331-8332.

(68) Ollagnier, S.; Mulliez, E.; Schmidt, P. P.; Eliasson, R.; Gaillard, J.; Deronzier, C.; Bergman, T.; Graslund, A.; Reichard, P.; Fontecave, M. *J. Biol. Chem.* 1997, *272*, 24216-23.

(69) Lieder, K. W.; Booker, S.; Ruzicka, F. J.; Beinert, H.; Reed, G. H.; Frey, P. A. *Biochemistry* 1998, *37*, 2578-85.

(70) Marquet, A.; Bui, B. T.; Smith, A. G.; Warren, M. J. *Nat. Prod. Rep.* 2007, *24*, 1027-40.

(71) Mejean, A.; Bui, B. T.; Florentin, D.; Ploux, O.; Izumi, Y.; Marquet, A. *Biochem. Biophys. Res. Commun.* 1995, *217*, 1231-7.

(72) Bianchi, V.; Eliasson, R.; Fontecave, M.; Mulliez, E.; Hoover, D. M.; Matthews, R. G.; Reichard, P. *Biochem. Biophys. Res. Commun.* 1993, *197*, 792-7.

(73) Ifuku, O.; Koga, N.; Haze, S.; Kishimoto, J.; Wachi, Y. *Eur. J. Biochem.* 1994, *224*, 173-8.

(74) Mulliez, E.; Padovani, D.; Atta, M.; Alcouffe, C.; Fontecave, M. *Biochemistry* 2001, *40*, 3730-6.

(75) Magnusson, O. T.; Reed, G. H.; Frey, P. A. *Biochemistry* 2001, *40*, 7773-82.

(76) Wu, W.; Booker, S.; Lieder, K. W.; Bandarian, V.; Reed, G. H.; Frey, P. A. *Biochemistry* 2000, *39*, 9561-70.

(77) Becker, A.; Fritz-Wolf, K.; Kabsch, W.; Knappe, J.; Schultz, S.; Volker Wagner, A. F. *Nat. Struct. Biol.* 1999, *6*, 969-75.

(78) Logan, D. T.; Andersson, J.; Sjoberg, B. M.; Nordlund, P. *Science* 1999, *283*, 1499-504.

(79) Krieger, C. J.; Roseboom, W.; Albracht, S. P.; Spormann, A. M. *J. Biol. Chem.* 2001, *276*, 12924-7.

(80) Layer, G.; Pierik, A. J.; Trost, M.; Rigby, S. E.; Leech, H. K.; Grage, K.; Breckau, D.; Astner, I.; Jansch, L.; Heathcote, P.; Warren, M. J.; Heinz, D. W.; Jahn, D. *J. Biol. Chem.* 2006, *281*, 15727-34.

(81) Walsby, C. J.; Ortillo, D.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M. *J. Am. Chem. Soc.* 2002, *124*, 11270-1.

(82) Cosper, N. J.; Booker, S. J.; Ruzicka, F.; Frey, P. A.; Scott, R. A. *Biochemistry* 2000, *39*, 15668-15673.

(83) Frey, P. A. *Annu. Rev. Biochem.* 2001, *70*, 121-48.

(84) Broach, R. B.; Jarrett, J. T. *Biochemistry* 2006, *45*, 14166-74.

(85) Marquet, A.; Florentin, D.; Ploux, O.; Bui, B. T. S. *J. Phys. Org. Chem.* 1998, *11*, 529-535.

(86) Cosper, M. M.; Jameson, G. N.; Hernandez, H. L.; Krebs, C.; Huynh, B. H.; Johnson, M. K. *Biochemistry* 2004, *43*, 2007-21.

(87) Tse Sum Bui, B.; Benda, R.; Schunemann, V.; Florentin, D.; Trautwein, A. X.; Marquet, A. *Biochemistry* 2003, *42*, 8791-8.

(88) Ugulava, N. B.; Gibney, B. R.; Jarrett, J. T. *Biochemistry* 2001, *40*, 8343-8351.

(89) Ugulava, N. B.; Surerus, K. K.; Jarrett, J. T. *J. Am. Chem. Soc.* 2002, *124*, 9050-1.

(90) Berkovitch, F.; Nicolet, Y.; Wan, J. T.; Jarrett, J. T.; Drennan, C. L. *Science* 2004, *303*, 76-9.

(91) Ugulava, N. B.; Sacanell, C. J.; Jarrett, J. T. *Biochemistry* 2001, *40*, 8352-8.

(92) Ollagnier-de-Choudens, S.; Mulliez, E.; Fontecave, M. *FEBS Lett.* 2002, *532*, 465-8.

(93) Lotierzo, M.; Tse Sum Bui, B.; Florentin, D.; Escalettes, F.; Marquet, A. *Biochem. Soc. Trans.* 2005, *33*, 820-3.

(94) Lepore, B. W.; Ruzicka, F. J.; Frey, P. A.; Ringe, D. *Proc. Natl. Acad. Sci. U S A* 2005, *102*, 13819-24.

(95) Marsh, E. N.; Patwardhan, A.; Huhta, M. S. *Bioorg. Chem*. 2004, *32*, 326-40.

(96) Frey, P. A.; Ballinger, M. D.; Reed, G. H. *Biochem. Soc. Trans.* 1998, *26*, 304-10.

(97) Frey, P. A.; Magnusson, O. T. *Chem. Rev.* 2003, *103*, 2129-48.

(98) Baraniak, J.; Moss, M. L.; Frey, P. A. *J. Biol. Chem.* 1989, *264*, 1357-60.

(99) Moss, M. L.; Frey, P. A. *J. Biol. Chem.* 1990, *265*, 18112-5.

(100) Ballinger, M. D.; Frey, P. A.; Reed, G. H. *Biochemistry* 1992, *31*, 10782-9.

(101) Frey, P. A.; Chang, C. H.; Ballinger, M. D.; Reed, G. H. *Methods Enzymol.* 2002, *354*, 426-35.

(102) Wang, S. C.; Frey, P. A. *Trends Biochem. Sci.* 2007, *32*, 101-10.

(103) Ruzicka, F. J.; Lieder, K. W.; Frey, P. A. *J. Bacteriol.* 2000, *182*, 469-76.

(104) Petrovich, R. M.; Ruzicka, F. J.; Reed, G. H.; Frey, P. A. *Biochemistry* 1992, *31*, 10774-8.

(105) Cosper, N. J.; Booker, S. J.; Ruzicka, F.; Frey, P. A.; Scott, R. A. *Biochemistry* 2000, *39*, 15668-73.

(106) Cheek, J.; Broderick, J. B. *J. Biol. Inorg. Chem.* 2001, *6*, 209-26.

(107) Fontecave, M.; Mulliez, E.; Ollagnier-de-Choudens, S. *Curr. Opin. Chem. Biol.* 2001, *5*, 506-11.

(108) Jarrett, J. T. *Curr. Opin. Chem. Biol.* 2003, *7*, 174-82.

(109) Cosper, M. M.; Cosper, N. J.; Hong, W.; Shokes, J. E.; Broderick, W. E.; Broderick, J. B.; Johnson, M. K.; Scott, R. A. *Protein Sci.* 2003, *12*, 1573-7.

(110) Cosper, M. M.; Jameson, G. N.; Davydov, R.; Eidsness, M. K.; Hoffman, B. M.; Huynh, B. H.; Johnson, M. K. *J. Am. Chem. Soc.* 2002, *124*, 14006-7.

(111) Krebs, C.; Broderick, W. E.; Henshaw, T. F.; Broderick, J. B.; Huynh, B. H. *J. Am. Chem. Soc.* 2002, *124*, 912-3.

(112) Walsby, C. J.; Hong, W.; Broderick, W. E.; Cheek, J.; Ortillo, D.; Broderick, J. B.; Hoffman, B. M. *J. Am. Chem. Soc.* 2002, *124*, 3143-51.

(113) White, R. H. *Biochim. Biophys. Acta* 1979, *583*, 55-62.

(114) Ollagnier-de-Choudens, S.; Fontecave, M. *FEBS Lett.* 1999, *453*, 25-28.

(115) Pierrel, F.; Bjork, G. R.; Fontecave, M.; Atta, M. *J. Biol. Chem.* 2002, *277*, 13367-13370.

(116) Wong, K. K., Murray, B. W., Lewisch, S. A., Baxter, M. K., Ridky, T. W., Ulissi-DeMario, L., and Kozarich, J. W. *Biochemistry* 1993, *32*, 14102-14110.

(117) Djaman, O.; Outten, F. W.; Imlay, J. A. *J. Biol. Chem.* 2004, *279*, 44590-9.

(118) Frazzon, J.; Dean, D. R. *Proc. Natl. Acad. Sci. U S A* 2001, *98*, 14751-3.

(119) Zheng, L. M.; Cash, V. L.; Flint, D. H.; Dean, D. R. *J. Biol. Chem.* 1998, *273*, 13264-13272.

(120) Kriek, M.; Peters, L.; Takahashi, Y.; Roach, P. L. *Protein Expr. Purif.* 2003, *28*, 241-5.

(121) Nakamura, M.; Saeki, K.; Takahashi, Y. *J. Biochem. (Tokyo)* 1999, *126*, 10-8.

(122) Takahashi, Y.; Nakamura, M. *J. Biochem. (Tokyo)* 1999, *126*, 917-26.

(123) Mihara, H.; Esaki, N. *App Microbiol. Biotech.* 2002, *60*, 12-23.

(124) Marquet, A. *Curr. Opin. Chem. Biol.* 2001, *5*, 541-549.

(125) Beinert, H. *Anal. Biochem.*1983, *131*, 373-378.

(126) Begley, T. P.; Xi, J.; Kinsland, C.; Taylor, S.; McLafferty, F. *Curr. Opin. Chem. Biol.* 1999, *3*, 623-629.

(127) Zheng, L.; White, R. H.; Cash, V. L.; Jack, R. F.; Dean, D. R. *Proc. Natl. Acad. Sci. U S A* 1993, *90*, 2754-8.

(128) Agar, J. N.; Krebs, C.; Frazzon, J.; Huynh, B. H.; Dean, D. R.; Johnson, M. K. *Biochemistry* 2000, *39*, 7856-7862.

(129) Smith, A. D.; Agar, J. N.; Johnson, K. A.; Frazzon, J.; Amster, I. J.; Dean, D. R.; Johnson, M. K. *J. Am. Chem. Soc.* 2001, *123*, 11103-4.

(130) Ollagnier-de-Choudens, S.; Mattioli, T.; Tagahashi, Y.; Fontecave, M. *J. Biol. Chem.* 2001, *276*, 22604-22607.

(131) Schwartz, C. J.; Giel, J. L.; Patschkowski, T.; Luther, C.; Ruzicka, F. J.; Beinert, H.; Kiley, P. J. *Proc. Natl. Acad. Sci. U S A* 2001, *98*, 14895-900.

(132) Hoff, K. G.; Silberg, J. J.; Vickery, L. E. *Proc. Natl. Acad. Sci. U S A* 2000, *97*, 7790-7795.

(133) Silberg, J. J.; Hoff, K. G.; Tapley, T. L.; Vickery, L. E. *J. Biol. Chem.* 2001, *276*, 1696-1700.

(134) Tokumoto, U.; Takahashi, Y. *J. Biochem.* 2001, *130*, 63-71.

(135) Kakuta, Y.; Horio, T.; Takahashi, Y.; Fukuyama, K. *Biochemistry* 2001, *40*, 11007-12.

(136) Fontecave, M.; Ollagnier-de-Choudens, S. *Arch. Biochem. Biophys.* 2007.

(137) Fish, W. W. *Methods Enzymol.* 1988, *158*, 357-364.

(138) Unden, G.; Becker, S.; Bongaerts, J.; Holighaus, G.; Schirawski, J.; Six, S. *Arch. Microbiol.* 1995, *164*, 81-90.

(139) Unden, G.; Becker, S.; Bongaerts, J.; Schirawski, J.; Six, S. *Antonie Van Leeuwenhoek* 1994, *66*, 3-22.

(140) Flint, D. H.; Tuminello, J. F.; Emptage, M. H. *J. Biol. Chem.* 1993, *268*, 22369-76.

(141) Rebeil, R.; Sun, Y. B.; Chooback, L.; Pedraza-Reyes, M.; Kinsland, C.; Begley, T. P.; Nicholson, W. L. *J. Bacteriol.* 1998, *180*, 4879-4885.

(142) Schmidt, T. G. M., Koepke, J., Frank, R., Skerra, A. *J Mol Biol* 1996, *255*, 753-766.

(143) Voss, S., Skerra, A. *Protein Eng.* 1997, *10*, 975-982.

(144) Weber, P. C.; Pantoliano, M. W.; Thompson, L. D. *Biochemistry* 1992, *31*, 9350-4.

(145) AcacloneSoftware In *http://www.acaclone.com/*; 1.0 ed.

(146) Livnah, O.; Bayer, E. A.; Wilchek, M.; Sussman, J. L. *Proc. Natl. Acad. Sci. U S A* 1993, *90*, 5076-80.

(147) ExPASy; http://www.expasy.ch/: 1993.

(148) Johnson, M. K., Spiro, T. G., and Mortenson, L. E. *J. Biol. Chem.* 1982, *257*, 2447-2452.

(149) Broderick, J. B.; Henshaw, T. F.; Cheek, J.; Wojtuszewski, K.; Smith, S. R.; Trojan, M. R.; McGhan, R. M.; Kopf, A.; Kibbey, M.; Broderick, W. E. *Biochem. Biophys. Res. Commun.* 2000, *269*, 451-456.

(150) Mihara, H.; Kurihara, T.; Yoshimura, T.; Soda, K.; Esaki, N. *J. Biol. Chem.* 1997, *272*, 22417-24.

(151) Ugulava, N. B.; Gibney, B. R.; Jarrett, J. T. *Biochemistry* 2000, *39*, 5206-5214.

(152) Vollmer, S. J., Switzer, R., and Debrunner, P. G. *J. Biol. Chem.* 1983, *258*, 14284-14293.

(153) Bui, B. T. S.; Florentin, D.; Fournier, F.; Ploux, O.; Mejean, A.; Marquet, A. *FEBS Lett.* 1998, *440*, 226-230.

(154) Ollagnier, S.; Meier, C.; Mulliez, E.; Gaillard, J.; Schuenemann, V.; Trautwein, A.; Mattioli, T.; Lutz, M.; Fontecave, M. *J. Am. Chem. Soc.* 1999, *121*, 6344-6350.

(155) Cudney, R.; Patel, S.; Weisgraber, K.; Newhouse, Y.; McPherson, A. *Acta Crystallogr. D Biol. Crystallogr.* 1994, *50*, 414-23.

(156) Gosavi, R. A.; Mueser, T. C.; Schall, C. A. *Acta Crystallogr. D Biol. Crystallogr*. 2008, *64*, 506-14.

(157) Roach, P. L.; Clifton, I. J.; Fulop, V.; Harlos, K.; Barton, G. J.; Hajdu, J.; Andersson, I.; Schofield, C. J.; Baldwin, J. E. *Nature* 1995, *375*, 700-4.

(158) Settembre, E. C.; Dorrestein, P. C.; Park, J. H.; Augustine, A. M.; Begley, T. P.; Ealick, S. E. *Biochemistry* 2003, *42*, 2971-2981.

(159) Chatterjee, A.; Han, X.; McLafferty, F. W.; Begley, T. P. *Angew. Chem. Int. Ed. Engl.* 2006, *45*, 3507-10.

(160) Kriek, M.; Martins, F.; Leonardi, R.; Fairhurst, S. A.; Lowe, D. J.; Roach, P. L. *J. Biol. Chem.* 2007, *282*, 17413-23.

(161) White, R. H.; Rudolph, F. B. *Biochim. Biophys. Acta* 1978, *542*, 340-347.

(162) Estramareix, B.; Therisod, M. *Biochim. Biophys. Act* 1972, *192*, 375-380.

(163) Liu, A.; Graslund, A. *J. Biol. Chem.* 2000, *275*, 12367-12373.

(164) Hinckley, G. T.; Frey, P. A. *Biochemistry* 2006, *45*, 3219-25.

(165) Gerrits, J.; Eidhof, H.; Brunnekreeft, J. W.; Hessels, J. *Methods Enzymol.* 1997, *279*, 74-82.

(166) Kriek, M.; Martins, F.; Challand, M. R.; Croft, A.; Roach, P. L. *Angew. Chem. Int. Ed. Engl.* 2007, *46*, 9223-6.

(167) Guianvarch, D.; Florentin, D.; Bui, B. T. S.; Nunzi, F.; Marquet, A. *Biochem. Biophys. Res. Commun.* 1997, *236*, 402-406.

(168) Bryant, P., PhD Thesis, University of Southampton, 2008.

(169) McNeill, L. A.; Bethge, L.; Hewitson, K. S.; Schofield, C. J. *Anal. Biochem.* 2005, *336*, 125-31.

(170) Andrei, P. I.; Pierik, A. J.; Zauner, S.; Andrei-Selmer, L. C.; Selmer, T. *Eur. J. Biochem.* 2004, *271*, 2225-30.

(171) Selmer, T.; Andrei, P. I. *Eur. J. Biochem.* 2001, *268*, 1363-72.

(172) Yu, L.; Blaser, M.; Andrei, P. I.; Pierik, A. J.; Selmer, T. *Biochemistry* 2006, *45*, 9584-92.

(173) Smith, M. D.; Rosenow, M. A.; Wang, M.; Allen, J. P.; Szostak, J. W.; Chaput, J. C. *PLoS ONE* 2007, *2*, e467.

(174) Price, M. N.; Arkin, A. P.; Alm, E. J. *PLoS Genet* 2006, *2*, e96.

(175) Trivedi, S.; Gehlot, H. S.; Rao, S. R. *Genet. Mol. Res.* 2006, *5*, 816-27.

(176) Vieille, C.; Burdette, D. S.; Zeikus, J. G. *Biotechnol. Annu. Rev.* 1996, *2*, 1-83.

(177) Vieille, C.; Zeikus, G. J. *Microbiol. Mol.Biol. Rev.* 2001, *65*, 1-43.

(178) Drake, H. L.; Daniel, S. L. *Res. Microbiol.* 2004, *155*, 869-83.

(179) Le Fourn, C.; Fardeau, M. L.; Ollivier, B.; Lojou, E.; Dolla, A. *Environ. Microbiol.* 2008, *10*, 1877-87.

(180) Witzmann, S.; Bisswanger, H. *Biochim. Biophys. Acta* 1998, *1385*, 341-52.

(181) Thoma, R.; Obmolova, G.; Lang, D. A.; Schwander, M.; Jeno, P.; Sterner, R.; Wilmanns, M. *FEBS Lett.* 1999, *454*, 1-6.

(182) Nicolet, Y.; Rubach, J. K.; Posewitz, M. C.; Amara, P.; Mathevon, C.; Atta, M.; Fontecave, M.; Fontecilla-Camps, J. C. *J. Biol. Chem.* 2008.

(183) Shimomura, Y.; Kamikubo, H.; Nishi, Y.; Masako, T.; Kataoka, M.; Kobayashi, Y.; Fukuyama, K.; Takahashi, Y. *J. Biochem.* 2007, *142*, 577-86.

(184) Henrissat, B.; Teeri, T. T.; Warren, R. A. *FEBS Lett.* 1998, *425*, 352-4.

(185) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* 1997, *25*, 3389-402.

(186) European Bioinformatics Institute, E.-E.; http://www.ebi.ac.uk/: 2007

(187) Postgate, J. R.; Campbell, L. L. *Bacteriol. Rev.* 1966, *30*, 732-8.

(188) Saida, F.; Uzan, M.; Odaert, B.; Bontems, F. *Curr. Protein. Pept. Sci.* 2006, *7*, 47-56.

(189) Baick, J. W.; Yoon, J. H.; Namgoong, S.; Soll, D.; Kim, S. I.; Eom, S. H.; Hong, K. W. *J Microbiol.* 2004, *42*, 111-6.

(190) Voordouw, G.; Hagen, W. R.; Kruse-Wolters, K. M.; van Berkel-Arts, A.; Veeger, C. *Eur. J. Biochem.* 1987, *162*, 31-6.

(191) Novy, R.; Drott, D.; Yaeger, K.; Mierendorf, R. *Newsletter of Novagen, Inc.* 2001, *12*, 1-3.

(192) Dong, H.; Nilsson, L.; Kurland, C. G. *J. Mol. Biol.* 1996, *260*, 649-63.

(193) Kitamura, M.; Kojima, S.; Ogasawara, K.; Nakaya, T.; Sagara, T.; Niki, K.; Miura, K.; Akutsu, H.; Kumagai, I. *J. Biol. Chem.* 1994, *269*, 5566-73.

(194) Mansure, J. J.; Hallenbeck, P. C. *Biotechnol. Lett.* 2008.

(195) National Centre for Biotechnology Information; http://www.ncbi.nlm.nih.gov/: 2008.

(196) Rubach, J. K.; Brazzolotto, X.; Gaillard, J.; Fontecave, M. *FEBS Lett.* 2005, *579*, 5055-60.

(197) Vignais, P. M.; Billoud, B.; Meyer, J. *FEMS Microbiol. Rev.* 2001, *25*, 455-501.

(198) Nicolet, Y.; Lemon, B. J.; Fontecilla-Camps, J. C.; Peters, J. W. *Trends. Biochem. Sci.* 2000, *25*, 138-43.

(199) Fontecilla-Camps, J. C.; Volbeda, A.; Cavazza, C.; Nicolet, Y. *Chem. Rev.* 2007, *107*, 4273-303.

(200) Corpet, F. *Nucleic Acids Res.* 1988, *16*, 10881-90.

(201) Hewitson, K. S., PhD Thesis, Oxford University, 2000.

(202) Sambrook, J.; Fritsch, E. E.; Maniatis, T. *Molecular Cloning: A laboratory manual*; Cold Spring Harbor Laboratory NY, 1989.

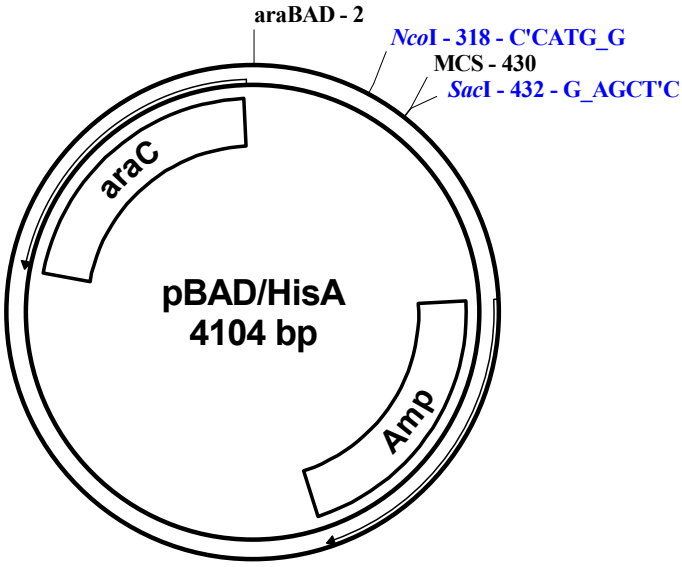
(203) Laemmli, U. K. *Nature* 1970, *227*, 680-5.

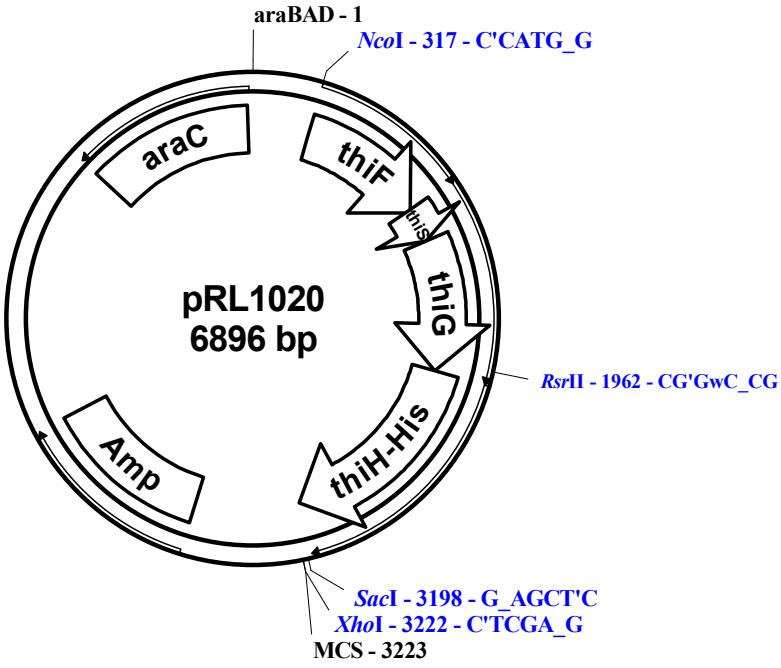
(204) Bradford, M. M. *Anal. Biochem.* 1976, *72*, 248-54.

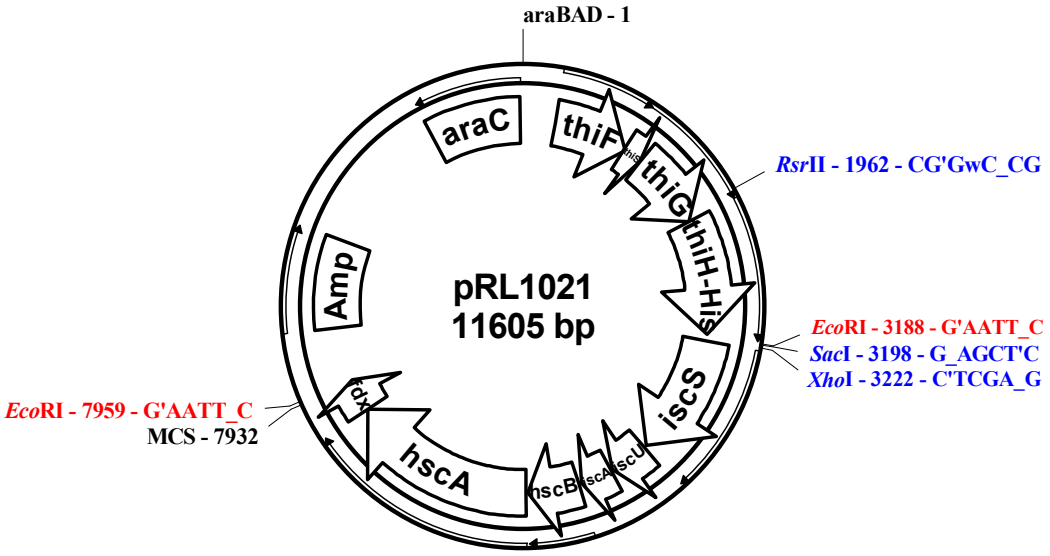
(205) Luo, W.; Tsai, A.; Nimmagadda, S.; Bodepudi, V.; Loor, R. *Clin. Chem.* 2005, *51*, A146-a146

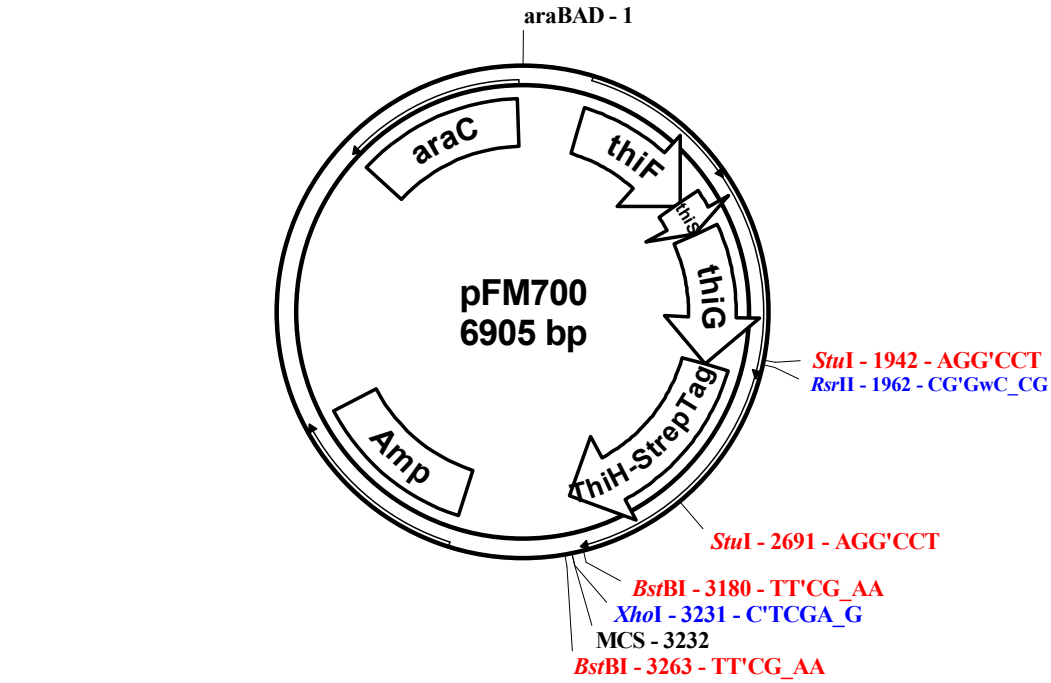
Appendix A

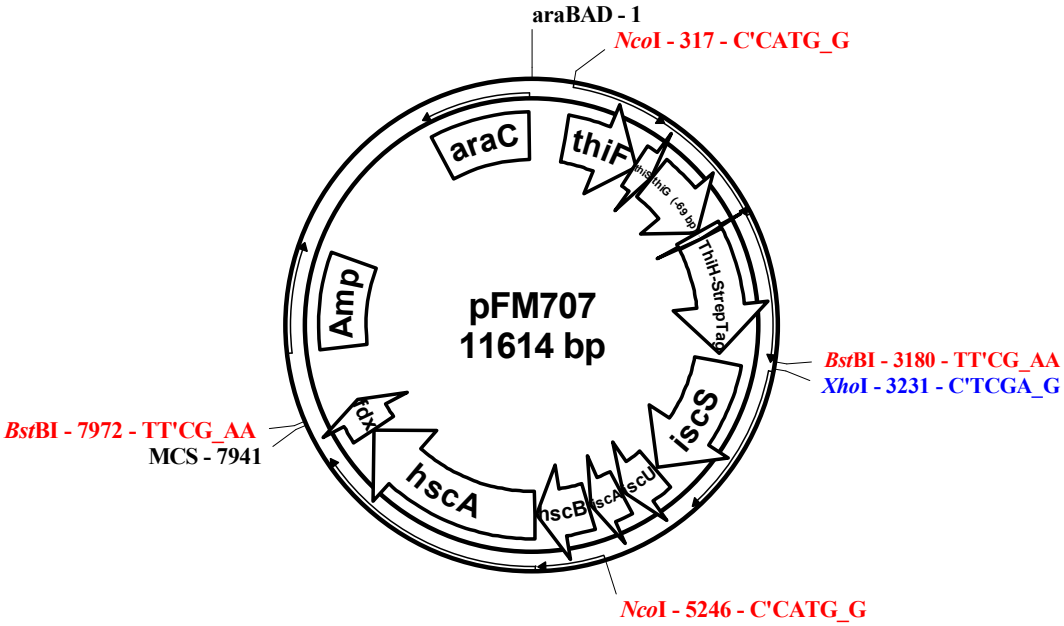
Maps of plasmids used throughout this project. Restriction sites of interest are indicated in blue for unique restriction sites or red for multiple restriction sites.

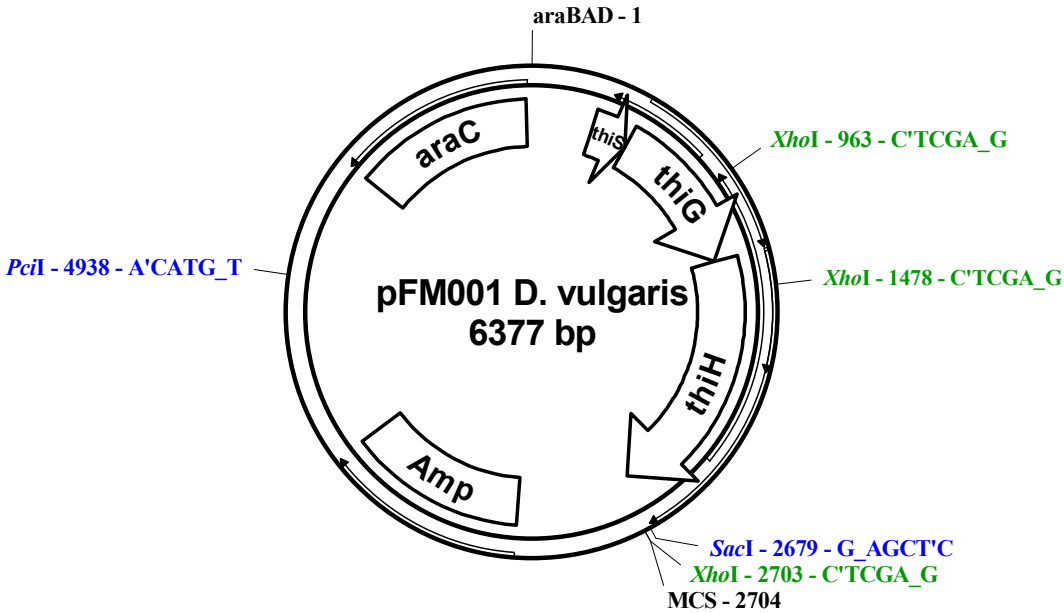


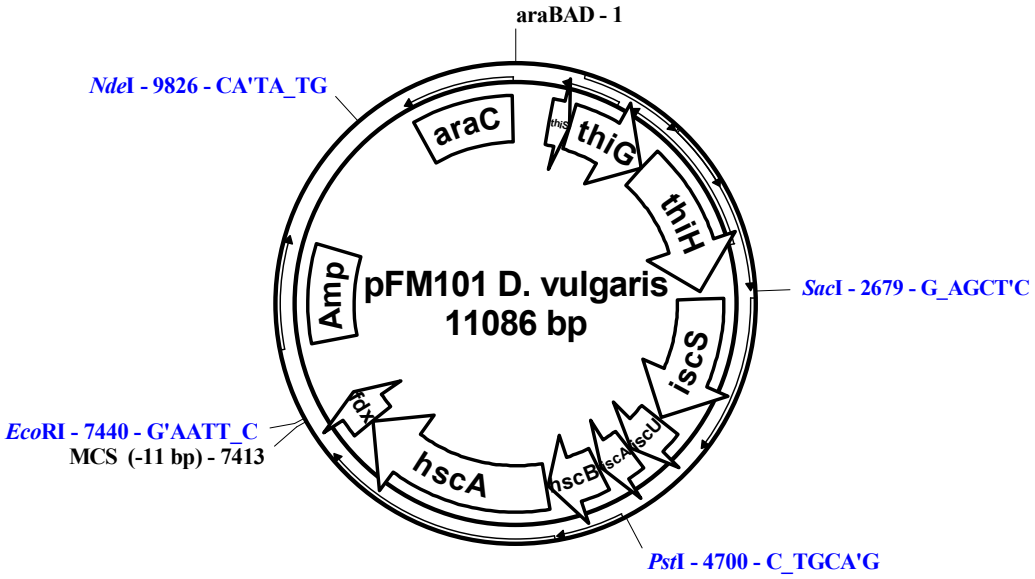


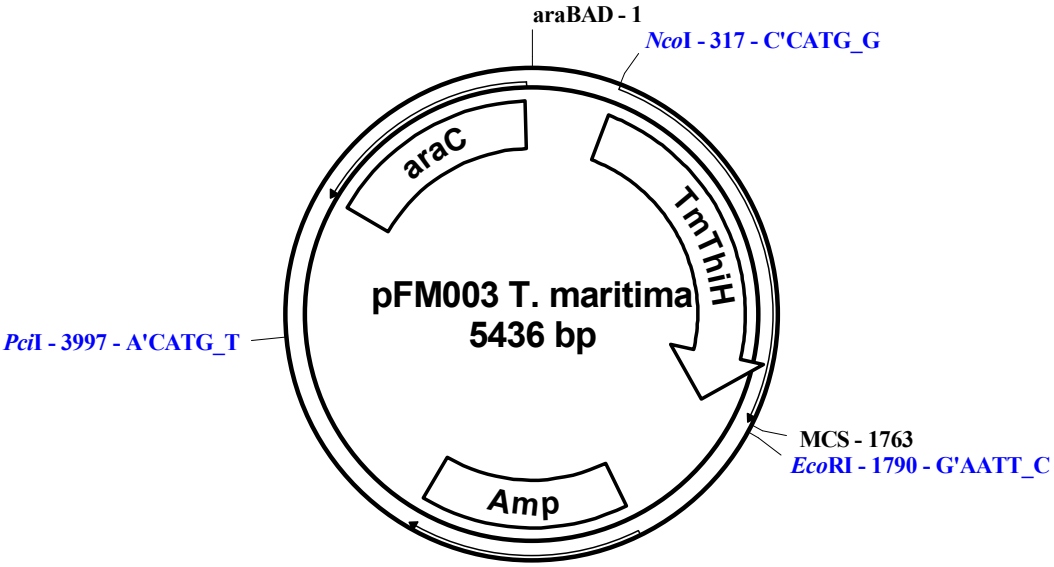

araBAD - 1
araC
thiF
thiG
ThiH-His
Amp
pRL1021
11605 bp
iscS
hscA
fdx
RsrII - 1962 - CG'GwC_CG
EcoRI - 3188 - G'AATT_C
SacI - 3198 - G_AGCT'C
XhoI - 3222 - C'TCGA_G
EcoRI - 7959 - G'AATT_C
MCS - 7932

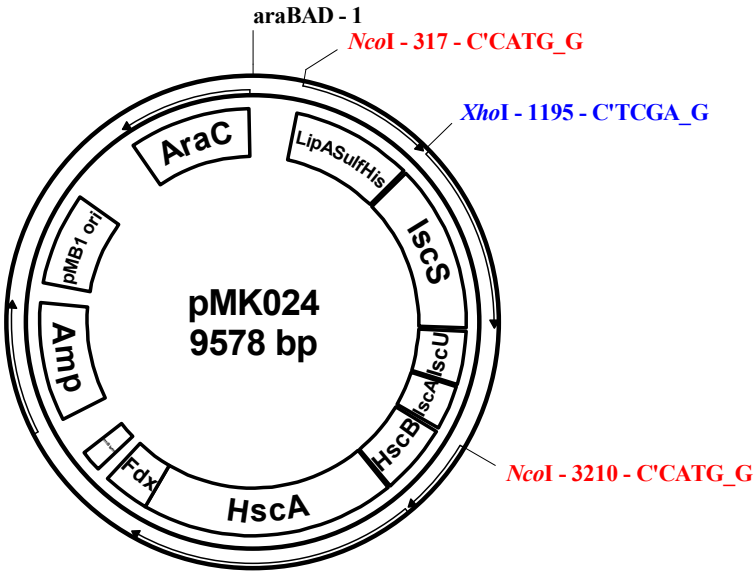

araBAD - 1
araC
thiF
thiG
thiG
ThiH-StrepTag
Amp
pFM700
6905 bp
StuI - 1942 - AGG'CCT
RsrII - 1962 - CG'GwC_CG
StuI - 2691 - AGG'CCT
BstBI - 3180 - TT'CG_AA
XhoI - 3231 - C'TCGA_G
MCS - 3232
BstBI - 3263 - TT'CG_AA

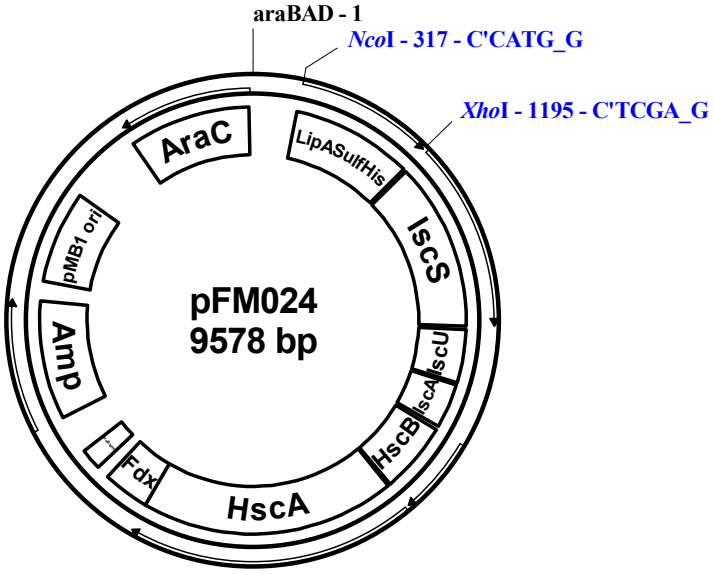

araBAD - 1
NcoI - 317 - C'CATG_G
araC
thiF
thiS
thiG (-59 bp)
ThiH-StrepTag
pFM707
11614 bp
Amp
BstBI - 3180 - TT'CG_AA
XhoI - 3231 - C'TCGA_G
iscS
BstBI - 7972 - TT'CG_AA
MCS - 7941
hscA
hscB
hscA
iscU
NcoI - 5246 - C'CATG_G

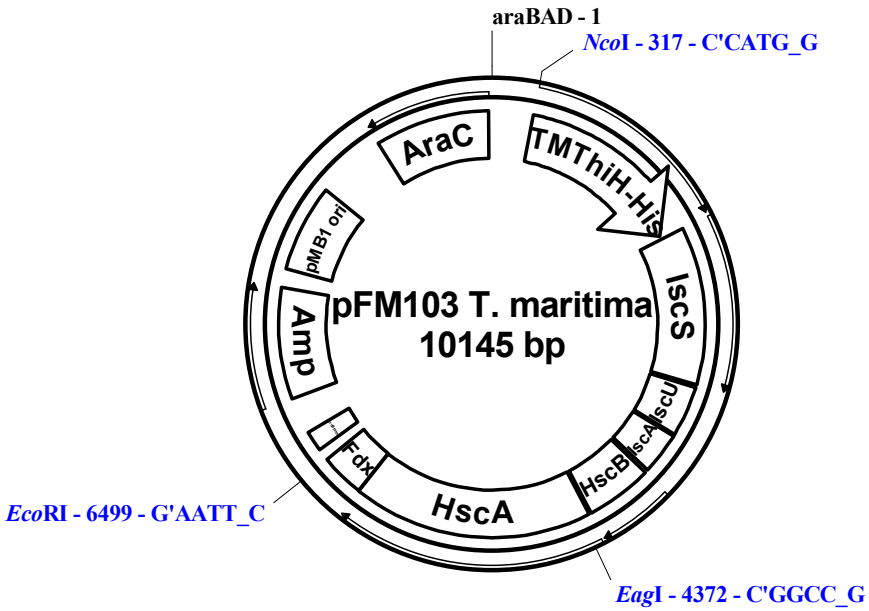

araBAD - 1
araC
thiS
thiG
XhoI - 963 - C'TCGA_G
PciI - 4938 - A'CATG_T
pFM001 D. vulgaris
6377 bp
XhoI - 1478 - C'TCGA_G
thiH
Amp
SacI - 2679 - G_AGCT'C
XhoI - 2703 - C'TCGA_G
MCS - 2704


araBAD - 1
NdeI - 9826 - CA'TA_TG
araC
thiS
thiG
thiH
Amp
pFM101 D. vulgaris
11086 bp
SacI - 2679 - G_AGCT'C
iscS
fdx
EcoRI - 7440 - G'AATT_C
MCS (-11 bp) - 7413
hscA
hscB
iscA
iscU
PstI - 4700 - C_TGCA'G


araBAD - 1
NcoI - 317 - C'CATG_G
araC
TmThiH
pFM003 T. maritima
5436 bp
PciI - 3997 - A'CATG_T
MCS - 1763
EcoRI - 1790 - G'AATT_C
Amp


araBAD - 1
NcoI - 317 - C'CATG_G
XhoI - 1195 - C'TCGA_G
NcoI - 3210 - C'CATG_G
AraC
LipASulfHis
IscS
IscU
IscA
HscB
HscA
Fdx
Amp
pMB1 ori
pMK024
9578 bp


araBAD - 1
NcoI - 317 - C'CATG_G
XhoI - 1195 - C'TCGA_G
AraC
LipASulfHis
IscS
IscU
IscA
HscB
HscA
Fdx
Amp
pMB1 ori
pFM024
9578 bp


araBAD - 1
NcoI - 317 - C'CATG_G
AraC
TMThiH-His
IscS
IscU
IscA
HscB
HscA
FdX
pMB1 ori
Amp
pFM103 T. maritima
10145 bp
EcoRI - 6499 - G'AATT_C
EagI - 4372 - C'GGCC_G

Appendix B

1. Agarose gel electrophoresis

50 X Running buffer solution, 50 X TAE (per litre):

Tris base	242 g
Glacial acetic acid	57.1 mL
0.5 M EDTA	100 mL
Adjusted to pH 8.0 with HCl.	

1 % Agarose gel (per gel):

Agarose	0.4 g
TAE buffer, 1 X	40 mL

The suspension was heated at top power rating in a microwave until the agarose had dissolved, poured and allowed to solidify at 4 ºC.

6 X Gel electrophoresis loading buffer:

Glycerol	30 % (v/v)
Bromophenol blue	0.25 % (w/v)

Electrophoresis gel stain:

Ethidium bromide	1 μg/mL

2. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

5 X Running buffer (per litre):

Tris base	15 g
Glycine	94 g
SDS	5 g

SDS-PAGE gel

15 % resolving gel (5 mL per gel):

H_2O	1.1 mL
30 % (w/v) Acrylamide/bis-acrylamide (37.5:1)	2.5 mL
1.5 M Tris-HCl, pH 8.8	1.3 mL
10 % (w/v) SDS	50 μL
10 % (w/v) Ammonium persulphate,	50 μL
TEMED	2 μL

3 % stacking gel (2 mL per gel):

H_2O	1.4 mL
30 % (w/v) Acrylamide/bis-acrylamide (37.5:1)	0.33 mL
1.0 M Tris, pH 6.8	0.25 mL
10 % (w/v) SDS	20 μL
10 % (w/v) Ammonium persulphate	20 μL
TEMED	2 μL

Freshly prepared ammonium persulfate and TEMED were added just prior to pouring the gels.

2 X SDS-PAGE loading buffer

Tris-HCl	50 mM, pH 6.5
Bromophenol Blue	0.1 % (w/v)
SDS	2 % (w/v)
Glycerol	10 % (w/v)
DTT	100 mM

SDS-PAGE gel stain (per litre):

Acetic acid	100 mL
Methanol	450 mL
H_2O	450 mL
Coomassie Brilliant Blue	2.5 g

SDS-PAGE gel destain (per litre):

Acetic acid	100 mL
Methanol	300 mL
H_2O	600 mL

3. Bradford Assay

Bradford Reagent (per litre):

88 % Phosphoric acid	100 mL
Ethanol	50 mL
Coomassie Brilliant Blue	100 mg
H_2O	850 mL

Coomassie Brilliant Blue was first dissolved in ethanol, then phosphoric acid and water were added and the solution filtered.

4. Growth Media

2 X Tryptone/Yeast extract (2YT) (per litre):

Bacto tryptone	16 g
Yeast extract	10 g
NaCl	5 g

Glycerol minimal medium (GMM) (per litre):

5X M9 salts solution	200 mL

The medium (800 mL) was autoclaved, allowed to cool down and the following sterile solutions added:

1 M $MgSO_4$	2 mL
1 M $CaCl_2$	0.1 mL
Glycerol (20 %)	20 mL

SOC medium (per litre):

Bacto tryptone	20 g
Yeast extract	5 g
NaCl	0.6 g

The medium (970 mL) was autoclaved, allowed to cool down and the following sterile solutions added:

1M $MgCl_2$	10 mL
1M $MgSO_4$	10 mL
2M Glucose	10 mL

$MgCl_2$ and $MgSO_4$ were filter sterilised.

The medium was autoclaved and stored at room temperature until used. Prior to use, sterilised 20 % (w/v) glucose was added to a final concentration of 0.2 % (w/v).

2YT agar plates (per litre):

Bacto tryptone	16 g
Yeast extract	10 g
NaCl	5 g
Agar	15 g

The solution was autoclaved and allowed to cool down to 50 ºC before adding ampicillin to a final concentration of 100 μg/mL. This medium was poured into petri dishes, allowed to harden and stored at 4 ºC

Appendix C

Sequences of genes used for sequence alignments

Escherichia coli thiH

MKTFSDRWRQLDWDDIRLRINGKTAADVERALNASQLTRDDMMALLSPAASGYLEQLAQRAQRLTRQRFG
NTVSFYVPLYLSNLCANDCTYCGFSMSNRIKRKTLDEADIARESAAIREMGFEHLLLVTGEHQAKVGMDY
FRRHLPALREQFSSLQMEVQPLAETEYAELKQLGLDGVMVYQETYHEATYARHHLKGKKQDFFWRLETPD
RLGRAGIDKIGLGALIGLSDNWRVDSYMVAEHLLWLQQHYWQSRYSVSFPRLRPCTGGIEPASIMDERQL
VQTICAFRLLAPEIELSLSTRESPWFRDRVIPLAINNVSAFSKTQPGGYADNHPELEQFSPHDDRRPEAV
AAALTAQGLQPVWKDWDSYLGRASQRL

Desulfovibrio vulgaris thiH

MSFYDELARWPHETLDALIASSTADDVLRALAATRPGPADLAALLSPAAMPYLEDMAQRAHELTVRHFGR
TIQLFTPLYLANHCTNQCRYCGFNARNHIRRDQLDAERIMAEGRAIASTGLRQLLLLTGDAPRISTVSYI
AEAAHRLRPLFPSIGVEVYAMQVEEYAELVAGGVESLTMFQETYNPGLYAWLHPAGPKRDFRFRLDAPER
GCLGGMRSVGLGALLGLDDWRRDAFYTAMHGAWLQRYYPATEVSFSVPRMRPHTGSFEPQHPVSDRELVQ
ILTAYRIFMPTAGITVSSREAAAFRDNLIPLGVTRMSAGVSTAVGGHAAGGDGNVASTEGAAAAIAASTA
ISSGDATGDLSGRSTKVDMTSGMAHAVTEDLQQGADAGPSQFDISDDRSVEEMVSAITARGYQPVFKDWE
PPQDNVYACGAAGHADGTVRCEAR

Thermotoga maritima thiH

MCMYVFVKERVESRSFIPEEKIFELLEKTKNPDPARVREIIQKSLDKNRLEPEETATLLNVEDPELLEEI
FEAARTLKERIYGNRIVLFAPLYIGNDCINDCVYCGFRVSNKVVERRTLTEEQLKEEVKALVSQGHKRLI
VVYGEHPNYSPEFIARTIDIVYNTKYGNGEIRRVNVNAAPQTIEGYKIIKSVGIGTFQIFQETYHRETYL
KLHPRGPKSNYNWRLYGLDRAMMAGIDDVGIGALFGLYDWKFEVMGLLYHTIHLEERFGVGPHTISFPRI
KPAINTPYSQKPEHVVSDEDFKKLVAIIRLSVPYTGMILTAREPAKLRDEVIKLGVSQIDAGSRIGIGAY
SHKEDDEDRKRQFTLEDPRPLDQVMRSLLKEGFVPSFCTACYRAGRTGEHFMEFAIPGFVKNFCTPNALF
TLQEYLCDYATEETRKVGEEVIERELQKMNPKIRERVREGLEKIKRGERDVRF

Desulfovibrio vulgaris hydG

MAEMKATWLDDAALEATLERNAQEDAVKAREVIAKARLLGGLDLDDVATLIALRDPELVEEMFQTARDVK
EEIYGNRLVLFAPLYISNLCSNECLYCAFRRSNTELDRKALDMDAIADETRLIVQQGHKRILLVAGESYP
REGFDYVLRAIDAVYSVHEGTGEIRRLNVNVAPLTVEQFRNLKARDIGTYQLFQETYHRGTYAKVHLAGP
KKDFDWRATAMDRAMEAGIDDVGIGPLFGLYDWRFEVLATLRHAQHLEEAFGVGCHTISVPRLEPACGSD
MASNPPRPVSNADFMRLVAILRLAVPYTGIIMSTRESAEMRTQTLALGVSQISAGSRTNPGGYAENEREE
AAQFQLGDHRSLSEVIADVGSMGYIPSFCTACYRMGRTGHDFMDLAKPGLIKQKCGPNALATFKEYLLDY
GTPEARAVGESVIAADLGKLDEKTRRVAERLIARVDEGRRDVFV

Moorella thermoacetica thiH

MQHGYRADFINHEEIEGYLEEAKRATRDVAVRIIEKAREAKGLEPYEVAVLLQNDDADVRRRMFTAAREI
KEKIYGQRIVLFAPLYFSDYCINNCRYCGYRRENKFERRRLEPEELEREVRILESLGHKRLALEAGEDPV
HCPLEYTLDVINRIYRITEANGSIRRVNVNIAATTVDAYRQLKAAGIGTYVLFQETYHRPTYAYMHPGGP
KADYDWHTTAMDRAMEGGIDDVGLGVLFGLYDYKFEVMGLLYHARHLEETFGVGPHTISVPRLRPAYNIT
LEKFPYLVDDEDFKKLVAIIRLAVPYTGMIISTRETAELRAELLELGVSQISAGSCTGVGGYGRHYADQE
DDIPQFEIGDHRHPDEVIGDLCRRGYLPSYCTACYRRGRTGDRFMSLAKTGEIQHCCQPNAILTFKEYLL
DYARPATREVGETTIREHLARIPSPAIRAETERRLERIAAGERDLYF

Vibrio vulnificus thiH

MSFVEHFNQLNWDDIRLSIYSKTALDVERALSKNKLDLEDFKALISPAAEHYLEAMAQRSYALTRKRFGN
TLSLYIPLYLSNLCANACTYCGFSMENRLKRRTLTVQEIDAEMAAIKAMKFDSVLLVTGEHEGKVGMNYF
RQVLPQIKRAFNYLAMEVQPLSQEHYQELKLLGLDAVMVYQETYHPATYAQHHLRGNKTNFNYRLETPDR
LARAGIDKIGIGALIGLEEWRTDCFYVAAHLDYLERTYWQTRYSISFPRLRPCAGALQPKSVMSDKQLVQ
LICAYRLFNPEVELSLSTRESALFRDNVLPLGITSMSAASKTQPGGYAMPDVELEQFEISDERSAAEVEQ
MIRSKGFDPVWRDWHAAYSGARFS

Chlorobium tepidum thiH

MIALPAWLTDERLSEDIEPLLRQTDNESLERLAAEAQAVTLRRFGRVISLYTPLYLSNFCSSGCVYCGFA
SDRRSPRRKLDTDEIEKELLAMKALGVSDVLLLTGERTNSVGFDYLRRAVDIAARHMPRVAVEAFPMSVA
EYRGLAECGCTGLTIYQETYDPDHYRELHRWGPKQDFLERLETPERAITGGIRSVGIGALLGLSEPVGEA
LAVLRHARYLCKTYWKAGVTVSFPRIRPQEGGFQPSFTVSDRFLARMIFAFRIGMPDVDLVLSTRESSNF
RDGMAGLGITRMSIASRTTVGGYVEKETAGASQFEVSDNRSVEAFCAALRAKDLEPVFKNWDAAYNNPLP
AEECT

Campylobacter jejuni thiH

MQDYMQHLPHMQEIKSEILNKVLTQVQSYDESQFSAKDVKNALNQTHLSIEHLKALLSSAAEDFIEELAF
KSAKVKQKYFGNSISLFTPLYLSNYCNSKCVYCGFQKGNKIARAKLNEAEIHEEMQAIAKSGLEEILMLT
GEGREFASVEYIANACKIAREYFKVVGVEIYPMNEDEYKILHEKGCDYVTVFQETYNPLKYSKIHLAGEK
RIFPYRFNAQERALKAGMRGVAFAALLGIDDFRKDALATALHAHFLQQAYSHAEISISVPRLRPIINNAK
IHPKDVSEKRLLQVLCAYRLFLPFAGITISSRERIGFRDEVIKLGATKMSAGVSVGIGEHKGEKKGDEQF
EISDDRSVDEILAMLKRSNLQAVMSDSIYVG

Codon Bias

D. vulgaris ThiH

UUU	8.8(4)	UCU	19.7(9)	UAU	2.2(1)	UGU	19.7(9)
UUC	8.8(4)	UCC	19.7(9)	UAC	2.2(1)	UGC	37.2(17)
UUA	2.2(1)	UCA	28.4(13)	UAA	0.0(0)	UGA	15.3(7)
UUG	6.6(3)	UCG	37.2(17)	UAG	2.2(1)	UGG	35.0(16)
CUU	2.2(1)	CCU	10.9(5)	CAU	6.6(3)	CGU	15.3(7)
CUC	8.8(4)	CCC	19.7(9)	CAC	6.6(3)	CGC	28.4(13)
CUA	4.4(2)	CCA	39.4(18)	CAA	4.4(2)	CGA	10.9(5)
CUG	10.9(5)	CCG	50.3(23)	CAG	26.3(12)	CGG	52.5(24)
AUU	8.8(4)	ACU	8.8(4)	AAU	2.2(1)	AGU	10.9(5)
AUC	8.8(4)	ACC	43.8(20)	AAC	10.9(5)	AGC	24.1(11)
AUA	4.4(2)	ACA	17.5(8)	AAA	0.0(0)	AGA	10.9(5)
AUG	28.4(13)	ACG	32.8(15)	AAG	8.8(4)	AGG	28.4(13)
GUU	10.9(5)	GCU	8.8(4)	GAU	2.2(1)	GGU	6.6(3)
GUC	6.6(3)	GCC	15.3(7)	GAC	6.6(3)	GGC	35.0(16)
GUA	4.4(2)	GCA	28.4(13)	GAA	2.2(1)	GGA	4.4(2)
GUG	17.5(8)	GCG	32.8(15)	GAG	6.6(3)	GGG	19.7(9)

T. maritima (TM1267)

UUU	14.8(7)	UCU	6.3(3)	UAU	6.3(3)	UGU	6.3(3)
UUC	33.8(16)	UCC	10.5(5)	UAC	38.0(18)	UGC	10.5(5)
UUA	0.0(0)	UCA	6.3(3)	UAA	2.1(1)	UGA	0.0(0)
UUG	4.2(2)	UCG	6.3(3)	UAG	0.0(0)	UGG	4.2(2)
CUU	19.0(9)	CCU	6.3(3)	CAU	0.0(0)	CGU	2.1(1)
CUC	38.0(18)	CCC	10.5(5)	CAC	21.1(10)	CGC	6.3(3)
CUA	2.1(1)	CCA	12.7(6)	CAA	2.1(1)	CGA	0.0(0)
CUG	16.9(8)	CCG	21.1(10)	CAG	23.2(11)	CGG	0.0(0)
AUU	6.3(3)	ACU	6.3(3)	AAU	6.3(3)	AGU	0.0(0)
AUC	40.1(19)	ACC	10.5(5)	AAC	31.6(15)	AGC	6.3(3)
AUA	29.5(14)	ACA	8.4(4)	AAA	29.5(14)	AGA	52.7(25)
AUG	19.0(9)	ACG	25.3(12)	AAG	38.0(18)	AGG	21.1(10)
GUU	16.9(8)	GCU	4.2(2)	GAU	12.7(6)	GGU	19.0(9)
GUC	16.9(8)	GCC	10.5(5)	GAC	31.6(15)	GGC	12.7(6)
GUA	6.3(3)	GCA	12.7(6)	GAA	54.9(26)	GGA	31.6(15)
GUG	31.6(15)	GCG	21.1(10)	GAG	52.7(25)	GGG	2.1(1)

E. Coli ThiH

UUU	15.7(6)	UCU	5.2(2)	UAU	21.0(8)	UGU	5.2(2)
UUC	21.0(8)	UCC	15.7(6)	UAC	15.7(6)	UGC	7.9(3)
UUA	13.1(5)	UCA	10.5(4)	UAA	0.0(0)	UGA	2.6(1)
UUG	7.9(3)	UCG	13.1(5)	UAG	0.0(0)	UGG	23.6(9)
CUU	18.4(7)	CCU	7.9(3)	CAU	21.0(8)	CGU	21.0(8)
CUC	10.5(4)	CCC	2.6(1)	CAC	7.9(3)	CGC	36.7(14)
CUA	13.1(5)	CCA	2.6(1)	CAA	26.2(10)	CGA	7.9(3)
CUG	52.5(20)	CCG	28.9(11)	CAG	34.1(13)	CGG	15.7(6)
AUU	18.4(7)	ACU	7.9(3)	AAU	13.1(5)	AGU	10.5(4)
AUC	13.1(5)	ACC	15.7(6)	AAC	13.1(5)	AGC	13.1(5)
AUA	5.2(2)	ACA	5.2(2)	AAA	23.6(9)	AGA	5.2(2)
AUG	26.2(10)	ACG	15.7(6)	AAG	7.9(3)	AGG	2.6(1)
GUU	15.7(6)	GCU	15.7(6)	GAU	28.9(11)	GGU	10.5(4)
GUC	10.5(4)	GCC	39.4(15)	GAC	34.1(13)	GGC	31.5(12)
GUA	5.2(2)	GCA	2.6(1)	GAA	39.4(15)	GGA	7.9(3)
GUG	10.5(4)	GCG	34.1(13)	GAG	23.6(9)	GGG	7.9(3)