EXPLORING PROKARYOTIC THIAMIN BIOSYNTHESIS: MECHANISTIC STUDIES ON THIAMIN THIAZOLE SYNTHASE AND PYRIMIDINE SYNTHASE

by Amrita Brajagopal Hazra

EXPLORING PROKARYOTIC THIAMIN BIOSYNTHESIS: MECHANISTIC STUDIES ON THIAMIN THIAZOLE SYNTHASE AND PYRIMIDINE SYNTHASE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Amrita Brajagopal Hazra

August 2010

EXPLORING PROKARYOTIC THIAMIN BIOSYNTHESIS: MECHANISTIC STUDIES ON THIAMIN THIAZOLE SYNTHASE AND PYRIMIDINE SYNTHASE

Amrita Brajagopal Hazra, Ph.D.

Cornell University 2010

Thiamin (Vitamin B1) is made from a coupling a thiazole and a pyrimidine unit, which are assembled separately. Studies have shown that the biosyntheses of thiazole and pyrimidine are different in prokaryotes versus eukaryotes. Understanding of thiamin biosynthesis is still incomplete and a lot of new discoveries relating to the enzymes in its biosynthesis have been explored in depth revealing new mechanisms and enzymology.

In prokaryotes, five different enzymes are known to be directly involved in thiamin thiazole biosynthesis. The *in vitro* reconstitution of this enzymatic pathway has been achieved and detailed insights have been obtained, however, the very small quantity of product produced in this in vitro reconstitution prevented direct characterization of its structure. We were able to study the last few steps on the prokaryotic thiamin thiazole pathway in greater detail, and elucidate the structure of the final product of the thiazole synthase to be the thiazole tautomer phosphate. We were also able to assign function to a gene involved in aromatization of the unstable thiazole tautomer phosphate to the thiazole carboxylate phosphate.

We also knew that a single gene product ThiC converts amino-imidazole

ribonucleotide, an intermediate in the purine nucleotide biosynthesis, to HMP, using a complex rearrangement reaction. This enzyme had been very difficult to isolate and study biochemically because it was air-sensitive and its cofactors were unknown. We recently were able to show that it was a [4Fe-4S] cluster containing enzyme, and belonged to the radical SAM family. The 4Fe-4S cluster binding motif (CX2-CX4-C) of ThiC is different from the motif (CX3-CX2-C) conventionally used by the other established members of this family. With the pure protein with a well-reconstituted Fe-S cluster, we were able to achieve remarkable enhancement in activity *in vitro*, in a defined biochemical system. True products of this reaction were thus identified to be HMP-phosphate and 5'-deoxyadenosine. We were also able to establish the fate of all the C atoms of the substrate, and have other insights into the mechanism of this complex enzyme with regard to the unprecedented rearrangement it brings about. Further mechanistic characterization of the remarkable rearrangement reaction catalyzed by ThiC is in progress.

BIOGRAPHICAL SKETCH

Amrita was born and brought up in the city of Pune, India, also referred to as the 'Oxford of the East'. She was the younger of two children that her parents, Sulekha Hazra and Brajagopal Hazra. Her brother Anirban and she grew up in an idyllic neighborhood in N.C.L (National Chemical Laboratory) colony, where they spent a lot of time playing, reading, cycling and climbing trees with their neighbors' children. Her mother was a plant biotechnologist and father was a synthetic organic chemist, and hence dinner table discussions were often informally scientific and quite delightful. Interestingly enough, when she had to decide what field of study to pursue for higher studies, she vacillated between medicine and chemistry, before finally deciding to study chemistry, with an intention to enter biology later. It seemed logical to probe complex biology problems from a chemical perspective, as it better suited her intrinsic approach of studying a problem from a very fundamental level. She did her undergraduate studies in Chemistry at Jadavpur University in Calcutta (2000-2003) where she was taught by some great teachers and it strengthened her urge to study further in this subject. She subsequently joined the Indian Institute of Technology, Kharagpur in 2003 for a Masters Degree in Chemistry, and then went on to do a Ph.D. in Chemical Biology at Cornell University under Tadhg Begley's guidance. In her five years, she learnt biology from a chemist's perspective, made lots of good friends and enjoyed beautiful Ithaca. She also got the chance to spend some part of her first two years at New York City as a part of the Tri-Institutional Program in Chemical Biology and her last year at College Station, TX at Texas A&M University, adding a nice cultural experience to her graduate study days. She has accepted a position in Michiko Taga's group at the Plant and Microbial Biology Department at UC, Berkeley as a postdoctoral researcher. She plans to pursue an academic career in the future, broadly in the field of biochemistry of complex interactions between organisms in nature.

To my parents, my brother and my grandmother, who brought me up to be what I am…

ACKNOWLEDGMENTS

This is definitely the toughest part of the five wonderful years that comprise my Ph.D. – to acknowledge, and more importantly, to acknowledge adequately the various people who make me what I am scientifically and otherwise.

Foremost in mind comes Tadhg Begley, the enthusiastic and energetic individual with a never-ending zest for interesting biological and mechanistic chemistry, good food, strong coffee and witty conversations who welcomed me into the lab 5 years back and encouraged me in as many ways possible to become the excited researcher I am today. Not only did Tadhg teach me the value of doing very high quality science, but also passed on his enthusiasm for science and the scientific community at every opportunity. This experience with have never been the same without his guidance and support.

The other person who played a very strong role in my scientific upbringing is Yueming Li. A scientist (at Sloan Kettering Institute, New York) of very high caliber and a human being *par excellence*, he infused in me his excitement and the confidence that anything was possible. Our lively discussions during my semester project at his lab, and his active participation in teaching me labwork and encouraging me to follow my dreams will stay forever with me.

I would like to also thank Hening Lin with whom I have had very good interactions and whose support has been invaluable for me during my Ph.D.

Additionally, I'd like to thank Frank Schroeder's for his support with my work and help with NMR interpretation that has always been of great use to me, and Yohannes Rezenom and Vanessa Santiago for their help with mass spectrometry

According to the old adage, "Charity begins at home", but for me "Chemistry begins at home"! To be brought up in an atmosphere of chemistry and plant biology and to

see my parents live very high quality lives centered around science was definitely my first inspiration to study chemistry and join grad school. My father's immense dedication and hardwork in his field of synthetic organic chemistry and my mother's dynamism and sincerity in the field of plant biotechnology and their support, love and hardwork goes a long way to make me what I am. My brother Anirban, who is also a (theoretical) chemist, led the way since childhood, teaching me at every step to do what I liked best, to strive towards perfection and to never give up. My wonderful sister-in-law Sonali has been a great support to me through all my Ph.D. years, and life would have never been the same without our regular conversations and lively discussions! My world would have never been complete without my dear grandmother who lived with us while I grew up. Even now, many of my stories tend to start off with, "My grandmother used to say…". I miss her a lot. Even now.

My heartfelt thanks to all my teachers along the way – Mrs. Correa, Ms. Martin, Mrs. Koshy and Mrs. Vinaya Joshi for teaching so well in school, Soumitra Sengupta, Swadesh Raichoudhuri, Kaushik Das, Kajal Rajak, Rupendranath Banerjee, Iva Bose and Narayan Bandhyopadhyay for being fantastic instructors in my undergraduate days, and Amit Basak, Srabani Taraphder, Sujit Roy, Swagata Dasgupta and Ganesh Pandey for giving me a real flavor of research through their teaching and work.

The next stage of learning was when I got into lab at Cornell – and who better than Abhishek Chatterjee to help me learn the ropes. Abhishek was an intrinsic part of all my fundamental queries, scientific debate, training and support of all kinds in the lab and beyond. Also, a special thanks to Sameh Abdelwahed and David Hilmey, very talented synthetic organic chemists, who always made the most complex synthetic reactions look like a piece of cake and to Sean, my classmate, with whom I went through the major part of grad school, and who taught me NMR and to drink strong coffee. Thanks to my former labmates Jennie, Amy G, Amy J, Kristin, Jeremiah,

Dinuka, Tomoshige, Kalyan, Tathagata, Jackie, Matt, Debashree, Udani, Mitch and Grant at Ithaca and Angad, Kayla, Nilkamal, Rung-yi and Yiquan, BJ, Lisa, Dinuka, Sameh, Kalyan and Tomoshige at College Station for making the Begley group my second home and teaching me chemistry and biology and how to have fun on the way. My special thanks to Angad and Rung-yi for being as excited about my projects as I am and taking them on. Thanks goes to all the staff at Cornell's Chemistry Department - Sharon Calhoun, Patricia Hine, Yvonne Ellis, Linda Fields, June Smithers and at Texas A&M's Chemistry Department – Judy Ludwig, Angie Medina, Angie Stickley, Christine Ruggles and our wonderful building manager Ed Janousek for making administrative life simple and enjoyable.

Enough can't be said for all the friends who became closer than family over time – Snehal, Arpita, Rakesh, Abhishek, Sujata, Avisek, Basu, Chandranidi and to my family at Ithaca and the US, who kept me well fed and happy – Romitadi, Souvikda, Bappaida, Trina, Arijit, Shirshendu, Abhimanyu, Rik, Krupa, Saba, Krishna, Vicki, Jackie, Dominika, Jacopo, Jaya, Debamita, Deepti, Sumana, Ruchi, Parthoda and Anando - I've got much more than I can ever give.

And finally, thanks to Biki for all the love, support, trust and faith he had in me as we went through the ups and downs of grad school and life – his love made all the difference.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Exploring prokaryotic thiamin biosynthesis: mechanistic studies on thiamin thiazole synthase and pyrimidine synthase.

1.1 Introduction:

Thiamin-pyrophosphate is a very important coenzyme in living systems with a fundamental role in cellular metabolism[1]. Prokaryotes and some eukaryotes can biosynthesize thiamin, while humans cannot and have a required dietary allowance of 1.4 mg/ day. It serves as an indispensable coenzyme in enzymatic cleavages of carbon-carbon bonds in α,β-dicarbonyl compounds and α-hydroxycarbonyl compounds and is involved in the functioning of many enzymes involved in basic metabolism like pyruvate decarboxylase, α-ketoacid dehydrogenase complexes, transketolase, pyruvate oxidase, acetolactase synthase and pyruvate oxidoreductases[2]. A deficiency of thiamin in the human diet causes the disease states known as Beri-Beri or Werninke-Korsakoff syndrome, both of which can be fatal[3] and is also a frequent complication of alcoholism. Thiamin was the first vitamin to be identified in the late nineteenth century[4]. The chemical mechanism of action of thiamin pyrophosphate was discovered in the mid-twentieth century by R. Breslow when he observed the rapid exchange of the C2-hydrogen of the 3,4-dimethylthiazolium salt by deuterium in neutral D_2O[5]. The formation of a ylide-like species on the thiazole ring of thiamin pyrophoshate facilitates the stabilization of a negative charge on the molecule with which the C2-carbanion forms an adduct, thus playing the role of an electron sink.

Considering its fundamental cellular functions, it was surprising that the mechanistic understanding of thiamin biosynthesis was at its infancy even a decade earlier.

However, recent advances in the mechanistic understanding of the biosynthesis of thiamin reveal a spectacular collection of novel and unique biochemical strategies employed by Nature in the production of this seemingly simple molecule[6], which explains the decades of effort that was involved in the process.

1.2 Background

1.2.1 Proteins and pathways of prokaryotic thiamin biosynthesis: Thiamin is made from a thiazole ring and a pyrimidine ring which are coupled in the final step (Figure 1.1). Each of the two components are made by separate pathways in prokaryotes and eukaryotes, thus vastly increasing the diversity of chemical biology involved in the biosynthesis of thiamin in living systems.

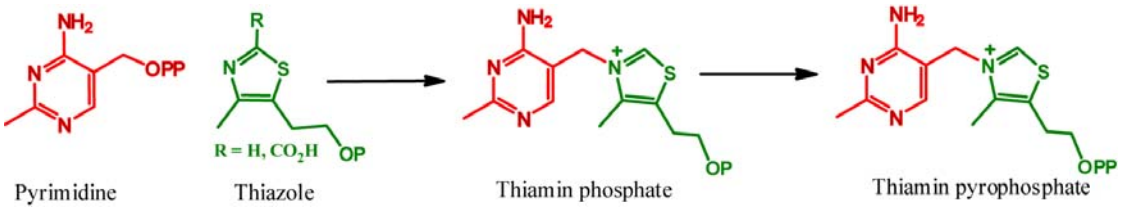


Figure 1.1: The pyrimidine and thiazole rings of thiamin are biosynthesized separately and then coupled together enzymatically to form thiamin phosphate, and further phosphorylated to form thiamin pyrophosphate, the active form of the cofactor.

The two rings are coupled in a penultimate step in the biosynthesis of the cofactor to form **3** by thiamin phosphate synthase (ThiE[7] in prokaryotes, THI6p[8] in eukaryotes), which is then pyrophosphorylated in the final step to form the active form of thiamin in vivo – thiamin pyrophosphate **4**.

The bacterial thiamin biosynthesis and metabolism pathway involves many enzymes which have interesting roles and novel chemistry. The genes involved in thiamin

biosynthesis, salvage, and transport are scattered throughout the chromosome. Thus far, genes that have been identified in *E. coli* and in *S. typhimurium* comprise three operons and four single gene loci. The components of the pathway by which thiamine pyrophosphate is biosynthesized in *B. subtilis* and *E. coli* are shown in Figure 1.2. The two routes are mostly the same, except glycine is used in *B. subtilis* and tyrosine is used in *E. Coli*. for making the dehydroglycine.

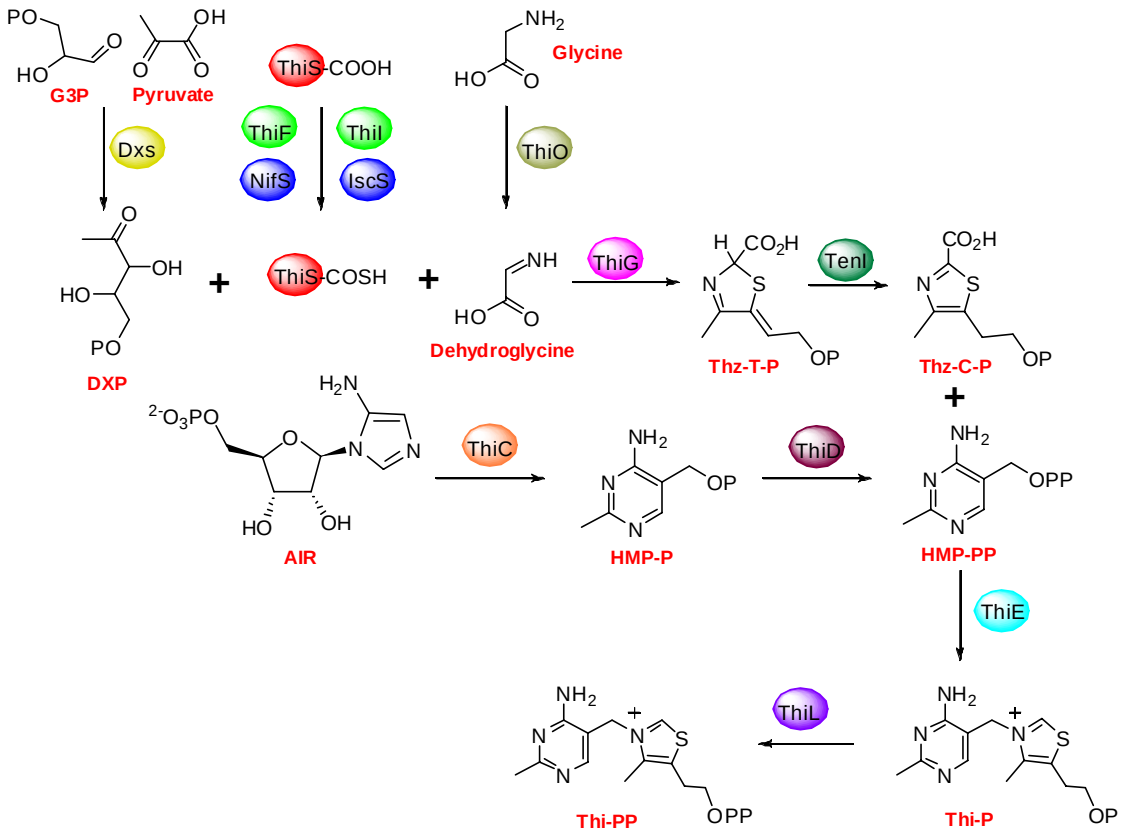


Figure 1.2: Thiamine biosynthetic proteins and their precursors in *B. subtilis* and *E. Coli*.

1.2.2 Prokaryotic thiamin thiazole biosynthesis: The biosynthetic pathway of the thiazole in prokaryotes utilizes 1-deoxy-D-xylulose-5-phosphate (DXP), glycine and

cysteine as its precursors. DXP synthase (Dxs) catalyzes the condensation of glyceraldehyde 3-phosphate (G3P) and pyruvate to give DXP and interestingly utilizes thiamin pyrophosphate as a cofactor[9,10]. DXP or DX is also required for the biosynthesis of pyridoxal[11] and for the biosynthesis of terpenes by the nonmevalonate pathway[12,13].

Thiazole biosynthesis requires five essential protein[14,15] – a small sulfur carrier protein ThiS[16] , an acyl adenylation protein ThiF, a cysteine desulfurase NifS [17], a glycine oxidase ThiO[18] and a thiazole synthase ThiG which brings together all the components to make the 5-membered ring.

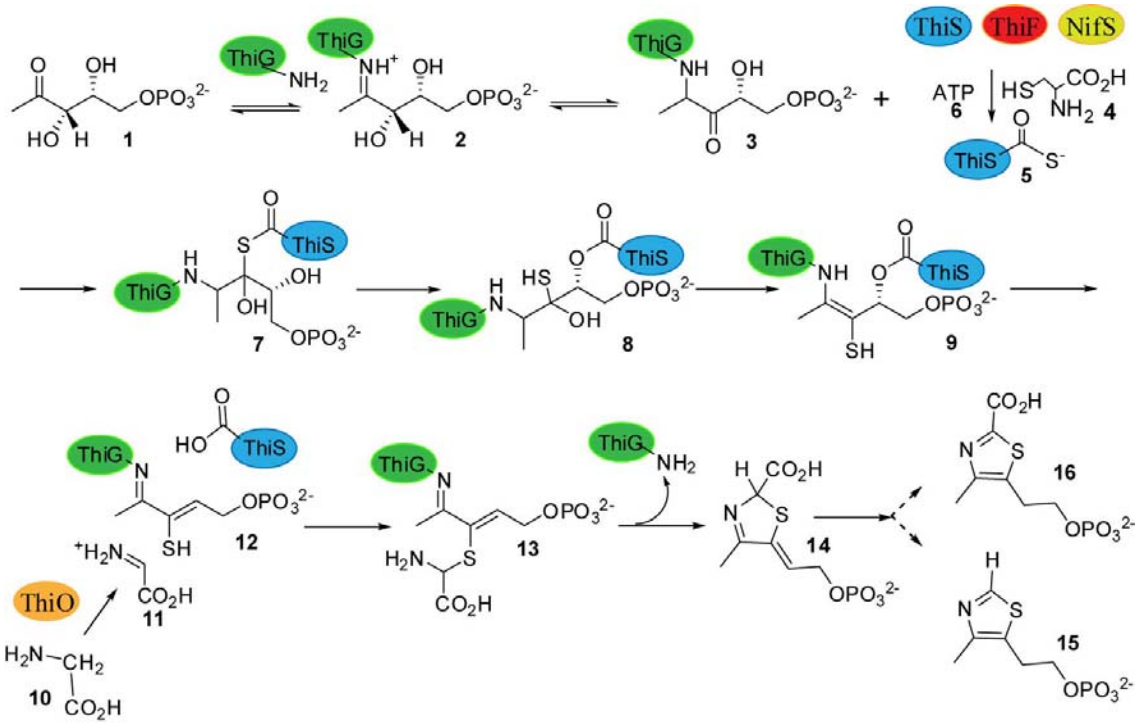


Figure 1.3: Mechanism of thiazole biosynthesis in *B. subtilis*

As shown in Figure 3, DXP **1** forms an imine with lysine 96 of the thiazole synthase. This imine then tautomerizes to aminoketone **3**. Addition of ThiS-thiocarboxylate **6**,

formed separately by reactions catalyzed by ThiF and NifS, to the ketone of **3** gives **7**, which undergoes an S/O acyl shift to **8** followed by loss of water to give **9**. Elimination of ThiS gives **12**. Addition of the thiol of **12** to the glycine imine, formed by ThiO-catalyzed oxidation of glycine, gives **13**. Cyclization via a transimination gives **14**, which could then aromatize by protonation/deprotonation to give **16** or by decarboxylation to give **15**.

1.2.3 Scope of research in thiazole biosynthesis: The late steps (**12** to product) in the biosynthesis have not yet been experimentally characterized. This is mainly because the amount of thiazole produced in this complex reaction is very small. Reliable characterization has to be done by converting it to thiamin phosphate and then by oxidation to the fluorescent compound thiochrome phosphate which can then be easily detected. We attempt to address these steps in prokaryotic thiazole biosynthesis – to characterize the final product of the thiazole synthase ThiG and investigate whether there are other other steps that we have not been able to probe because our method of analysis of thiazole was indirect and relied on derivatization to thiochome phosphate.

1.2.4 Prokaryotic thiamin pyrimidine biosynthesis: The biosynthesis of the pyrimidine moiety of thiamin occurs by a complex enzymatically catalyzed rearrangement of aminoimidazole ribotide. This remarkable reaction had been observed and has existed in literature for quite a while[19,20,21], but the mechanistic analysis had not been explored thus far in great detail. Only a single gene has been identified by mutagenesis studies (ThiC). The reaction was reconstituted in cell free extract at low levels. Yield of this reaction was enhanced by SAM, reduced nicotinamide and by the addition of cell free extract from wild type *E. coli*, suggesting additional proteins may be involved[19]. The

results of a comprehensive carbon and hydrogen labeling studies, carried out using this system, are shown in Figure 1.4.

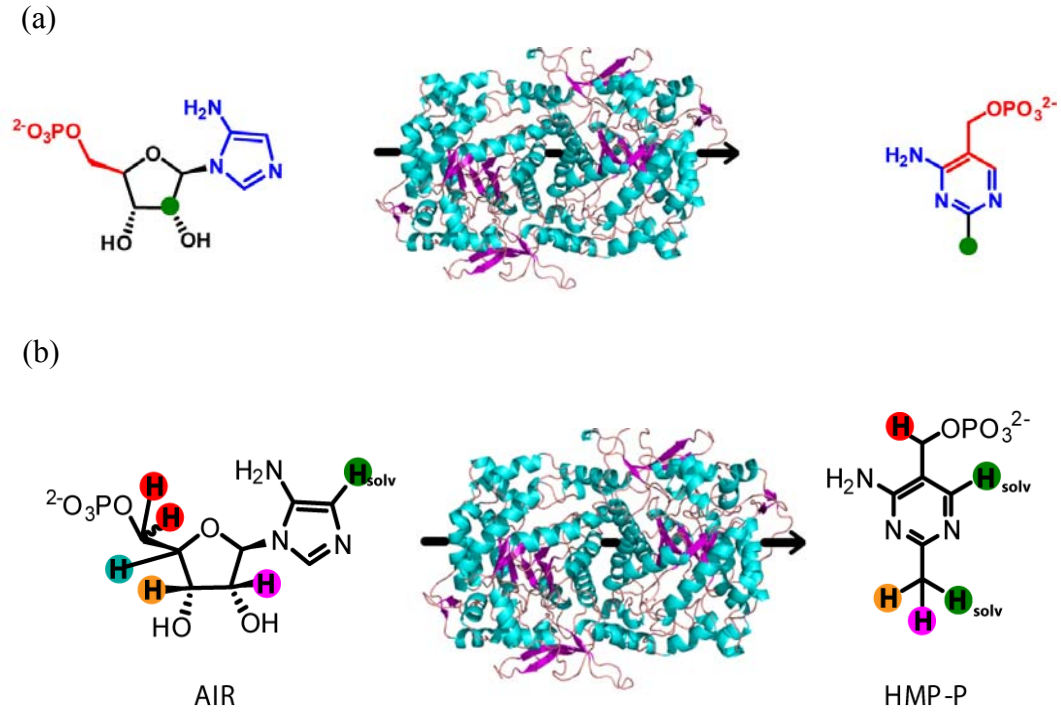


Figure 1.4: Summary of the labeling studies on the biosynthesis of thiamin pyrimidine, catalyzed by ThiC. (a) C-labeling pattern of HMP-P derived from AIR (b) H-labeling pattern of HMP-P derived from AIR and solvent

Although these results clearly established this reaction as the most complex unsolved rearrangement in primary metabolism, its mechanistic details remained unclear. Recently, successful reconstitution of this reaction using purified ThiC enzyme has been carried out in a defined biochemical system and its cofactor requirements and reaction products have been established with clarity. It was further structurally characterized by solving a crystal structure for the protein with imidazole ribose, a substrate analog bound in the active site. ThiC is a radical SAM enzyme has a conserved tri-cysteine motif (CX_2-CX_4-C), requires anaerobic conditions for robust

activity and utilizes SAM as a substrate rather than a cofactor, producing 5'deoxyadenosine in a 1:1 ratio with the product HMP-P[22].

1.2.5 Scope of research in pyrimidine biosynthesis: Though we have established a robust reconstitution system for HMP-P biosynthesis using ThiC, we have yet to delve into detailed mechanistic questions. The radical-induced rearrangement is unprecedented and many questions remain to be answered. The fate of C1 and C3 of the ribose ring has to be established. The catalytic role of the 5'deoxyadenosyl radical produced by SAM and the [4Fe-4S] cluster in the reaction that initiates the reaction by H-abstraction needs to be established. Whether this is from the protein or substrate is yet to be determined. With the protein active site being determined by the crystal structure, identification of conserved active site residues has enabled us to do mutational analysis. Different active site mutants may be made to probe the reaction being carried out to varying extents. Studying of these mutants to trap intermediates on the biosynthetic pathway and identify the mechanistic pathway of ThiC remains to be carried out.

Interestingly, biosynthesis of the pyrimidine moiety is different in *S. cerevisiae* and other fungi and is greatly unexplored yet. The gene product THI5p that has been identified to be responsible for making the thiamine pyrimidine is found to exist in one to five copies in organisms, each gene varying very slightly from the other in primary sequence. Preliminary *in vivo* research indicates histidine and pyridoxal phosphate (PLP) are used as substrates by the gene product THI5p (Figure 1.5)[23]. The activity of this protein has not been reconstituted *in vitro* and the mechanism of the reaction is not understood.

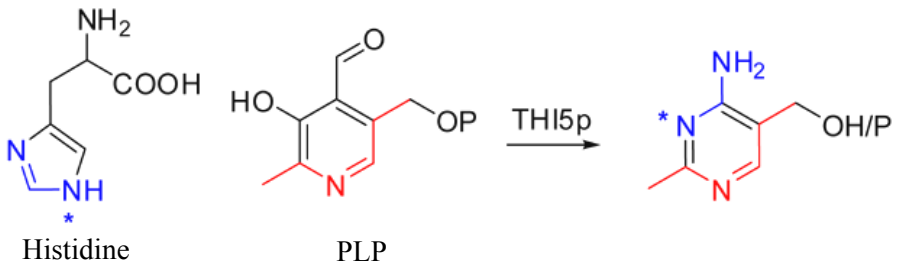


Figure 1.5: HMP biosynthesis in *S. cerevisiae* from PLP and histidine.

Some of these directions of thiamin biosynthesis have been addressed in the following chapters of this thesis.

REFERENCES

1) Jordan, F. *Natural Product Reports* **2003**, 20, (2), 184-201.

2) Perry Frey *Thiamine: Catalytic Mechanisms in Normal and Disease States* **2004** Marcel Dekker, Inc, Part I: Introduction

3) Buttersworth, R.F. *Nutrition Research reviews*, **2003**, 16, 277-283.

4) Verhoef, J.; Snippe, H.; Nottet, Hans, S.L.M. *FEMS Immunology and Medical Microbiology* **1999**, 26(3-4), 185-187.

5) R. Breslow. *J.Am.Chem.Soc.*, **1957**, 79: 1762-1763

6) Begley, T.P.; Chatterjee, A.; Hanes, J.W.; Hazra, A.; Ealick, S.E. *Curr Opin Chem Biol.* **2008**, 12, (2), 118-25.

7) Chatterjee, A.; Han, X.; McLafferty, F.W.; Begley, T.P. *Angew Chem Int Ed* **2006**, 45, 3507–3510.

8) Dorrestein, P.C.; Zhai, H.; McLafferty, F.W.; Begley, T.P. *Chem Biol* **2004**, 11, 1373–1381.

9) Sprenger et al. *Proc Nat Acad Sci* **1997**, 94, 12857–12862

10) Lois L.M.; Campos N; Putra S.R.; Danielsen K; Rohmer M; Boronat A *Proc Nat Acad Sci* **1998**, 95,2105–2110

11) Himmeldirk K.; Kennedy I.A.; Hill R.E.; Sayer B.G.; Spenser I.D. *J Chem Soc Chem Commun* **1996**, 1187–1188

12) Arigoni D.; Sagner S.; Latzel C.; Eisenreich W.; Bacher A.; Zenk M.H. *Proc Nat Acad Sci* **1997** 94, 10600–10605

13) Kuzuyama T.; Takahashi S.; Watanabe H.; Seto H. *Tetrahedron Lett.* **1998** 39 :4509–4512

14) Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999)

15) Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002). *J Biol Chem 277*, 48949-59.

16) Taylor SV, Kelleher NL, Kinsland C, Chiu HJ, Costello CA, Backstrom AD, McLafferty FW, Begley TP *J Biol Chem.* **1998** 273,16555-60.

17) Zheng, Limin; White, Robert H.; Cash, Valerie L.; Dean, Dennis R. *Biochemistry* **1994** *33*, 4714-20

18) Settembre, E. C., Dorrestein, P. C., Park, J. H., Augustine, A. M., Begley, T. P., and Ealick, S. E. *Biochemistry* **2003** 42, 2971-81.

19) Lawhorn, B.G.; Mehl, R.A.; Begley, T.P. *Org. Biomol. Chem.* **2004**, 2, 2538-2546.

20) K. Yamada and H. Kumaoka, *Biochem. Int.*, 1982, **5**, 771–776.

21) B. Estramareix and M. Therisod, *J. Am. Chem. Soc.*, 1984, **106**, 3857–3860.

22) Chatterjee, A., Li, Y., Zhang, Y., Grove, T.L., Lee, M., Krebs, C., Booker, S.J., Begley, T.P. & Ealick, S.E. *Nat Chem Biol.* **2008**, 4, 758-65.

23) Zeidler, J.; Sayer, B.G.; Spenser, I.D. *J. Am. Chem. Soc.* **2003**, 125, 13094-13105.

CHAPTER 2

Biosynthesis of the thiamin thiazole in *Bacillus subtilis*: Identification of the product of the thiazole synthase-catalyzed reaction*

2.1 Introduction

Thiamin pyrophosphate (TPP) is an essential cofactor in all living systems[1,2]. Most prokaryotes and eukaryotes biosynthesize TPP, but humans cannot and require it (1.4mg/day) from dietary sources[3]. TPP consists of a thiazole ring attached to a pyrimidine ring. The biosynthesis of TPP involves separate enzymatic routes for producing each of these heterocycles. Furthermore, these enzymatic routes for production of the thiazole ring and the pyrimidine ring are different in prokaryotes and eukaryotes. The early steps in the biosynthesis of the thiamin thiazole in *B. subtilis* have been studied extensively and the mechanism outlined in Figure 1 now has substantial experimental support.[4-11] In this mechanism, DXP **1** forms an imine with lysine 96 of the thiazole synthase. This imine then tautomerizes to aminoketone **3**. Addition of ThiS-thiocarboxylate **6**, formed separately by reactions catalyzed by ThiF and NifS, to the ketone of **3** gives **7**, which undergoes an S/O acyl shift to **8** followed by loss of water to give **9**. Elimination of ThiS gives **12**. Addition of the thiol of **12** to the glycine imine, formed by ThiO-catalyzed oxidation of glycine, gives **13**. Cyclization via a transimination gives **14**, which could then aromatize by protonation/deprotonation to give **15** or by decarboxylation to give **16**. The late steps (**12** to product) in the biosynthesis have not yet been experimentally characterized.

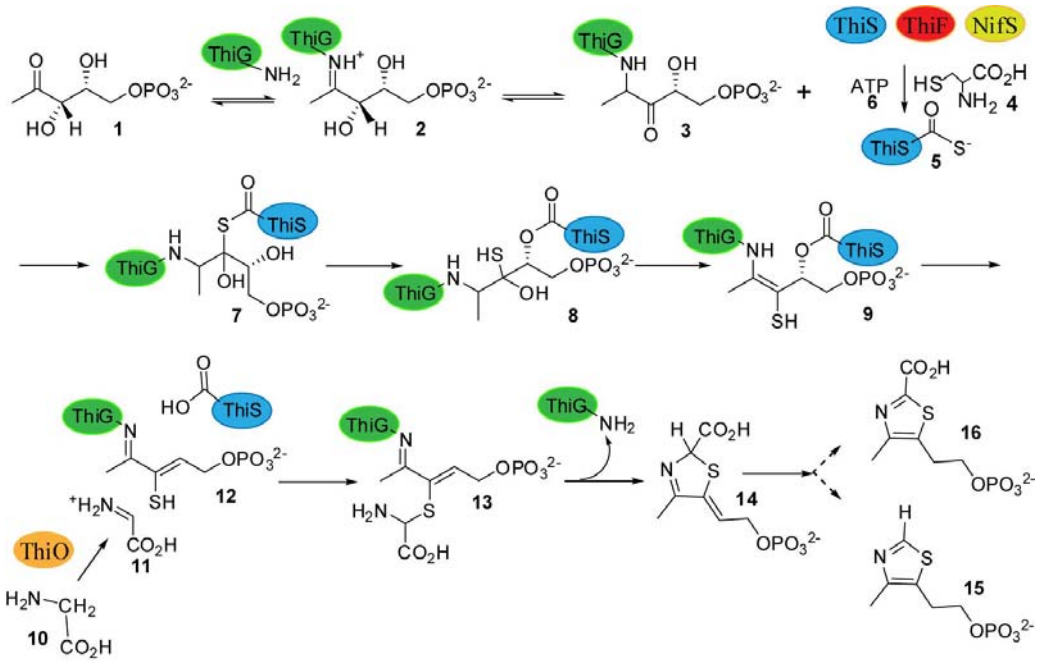


Figure 2.1: Currently proposed mechanism for the formation of the thiamin thiazole in *B. subtilis.* The steps in the mechanism have been validated by MS and NMR analysis till intermediate 12. The later steps of the reconstitution (13 to thiazole product) have not been validated.

It was not possible previously to directly characterize the final product of thiazole biosynthesis because our initial reconstitution yielded very low levels of thiazole and required a highly sensitive but indirect assay for product detection. This assay involved alkylation of the thiazole product with 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (HMP-PP) **17**, followed by the oxidation of the resulting thiamin phosphate **18** to the highly fluorescent thiochrome phosphate **19** (Figure 2.2).

Since thiazole phosphate **15** was, at that time, the only identified substrate for the well-characterized thiamin phosphate synthase, it seemed reasonable to assume that

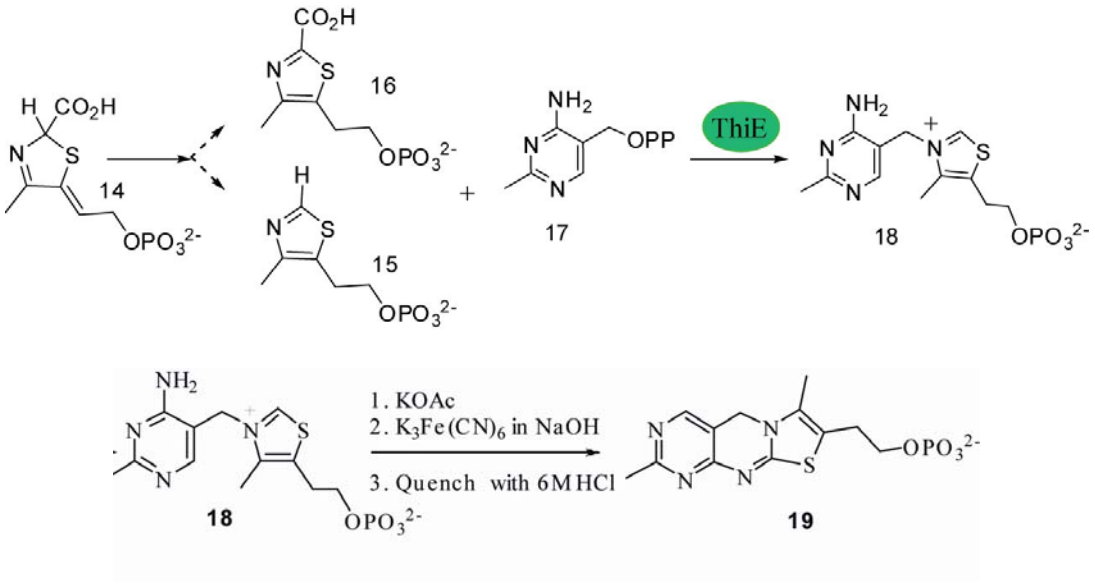


Figure 2.2: Alkylation of the thiazole product with pyrimidine **17,** followed by the thiochrome assay to form thiochrome phosphate.

thiochrome phosphate formation was a reliable way to measure the formation of this thiazole.[7] The amount of thiazole formed in a reconstitution reaction was measured by quenching the reaction at increasing timepoints and converting the thiazole formed to thiochrome phosphate (Figure 2.3)

However, the possibility remained that **14** or **16** could also be substrates for thiamin phosphate synthase, thus leaving unresolved the true identity of the reaction product of the bacterial thiazole synthase. Here we describe an improved reconstitution procedure which enables us to directly characterize the product of the bacterial thiazole synthase as the thiazole tautomer **14**.

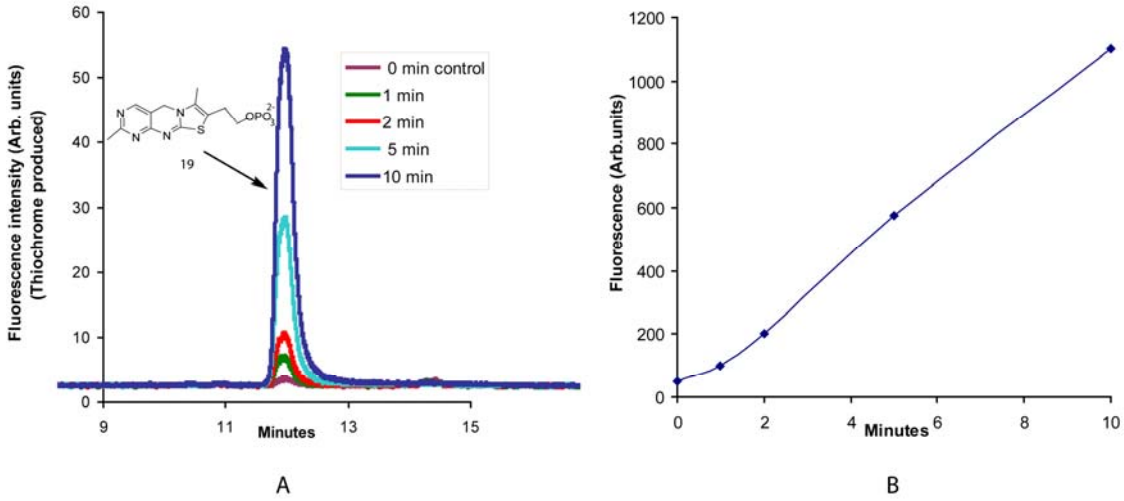


Figure 2.3: A) HPLC chromatogram showing the time-course for the thiazole reconstitution reaction using the thiochrome assay. (Reaction times = 0min, 1 min, 2 min, 10 min and 30min). B) Plot of product formation (as measured by peak area from the chromatogram) versus time.

The unexpected stability of **14** permits its characterization by 1-D and 2-D NMR studies and clarifies the later steps of the thiazole biosynthetic pathway in *B. subtilis*.

2.2 Results/ Discussion

2.2.1 The product of thiazole biosynthesis is not thiazole phosphate.

The previously reported reconstitution procedure was optimized and scaled up to produce larger quantities of the thiazole product. His-tagged proteins ThiF, NifS, ThiO and ThiSG were overexpressed in *E. coli* BL21(DE3). ThiS-COOH (in complex with ThiG), NifS and ThiF were incubated with L-cysteine **5** in the presence of dithiothreitol and ATP to form ThiS-COSH **6**. This was then added to DXP **1** and glycine **10** in the presence of ThiO and ThiG to produce the product of the thiazole

synthase-catalyzed reaction. The resulting reaction mixture was heat denatured, filtered and analyzed by reverse phase HPLC (Figure 2.4).

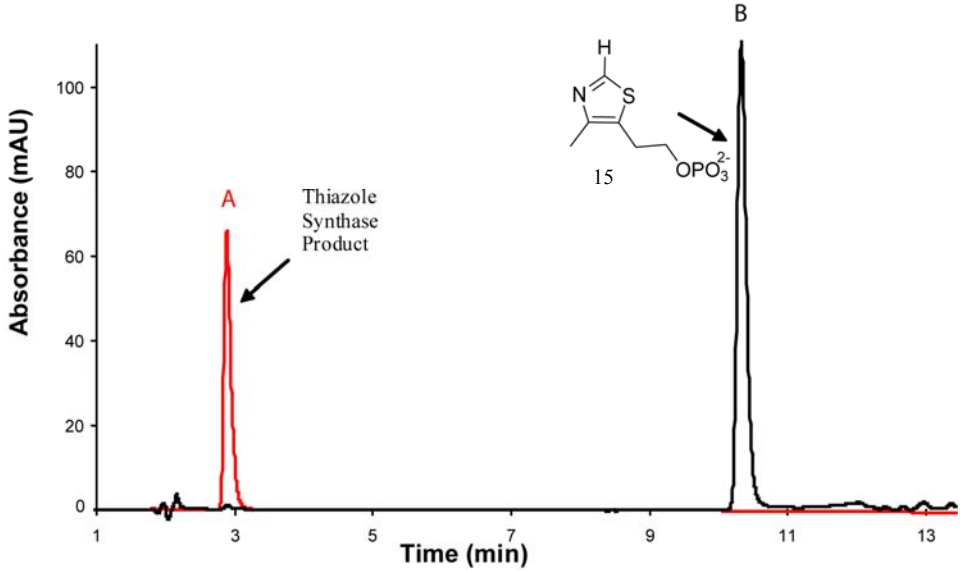


Figure 2.4: HPLC analysis of the product of the bacterial thiazole synthase reaction mixture. A - The enzymatic reaction mixture, B - Thiazole phosphate 15, the previously assumed reaction product.

The product of the reconstitution (peak A) was readily detected and did not comigrate with an authentic sample of thiazole phosphate **15** (peak B). The peak A compound had a UV absorption maximum at 300 nm and when treated with the pyrimidine **17** in the presence of thiamin phosphate synthase followed by thiochrome derivitization (Figure 2.2) produced a fluorescent product, which comigrated with an authentic sample of thichrome phosphate **19** (Figure 2.5). This experiment clearly demonstrates that the product of the thiazole reconstitution is not the anticipated thiazole phosphate **16**.

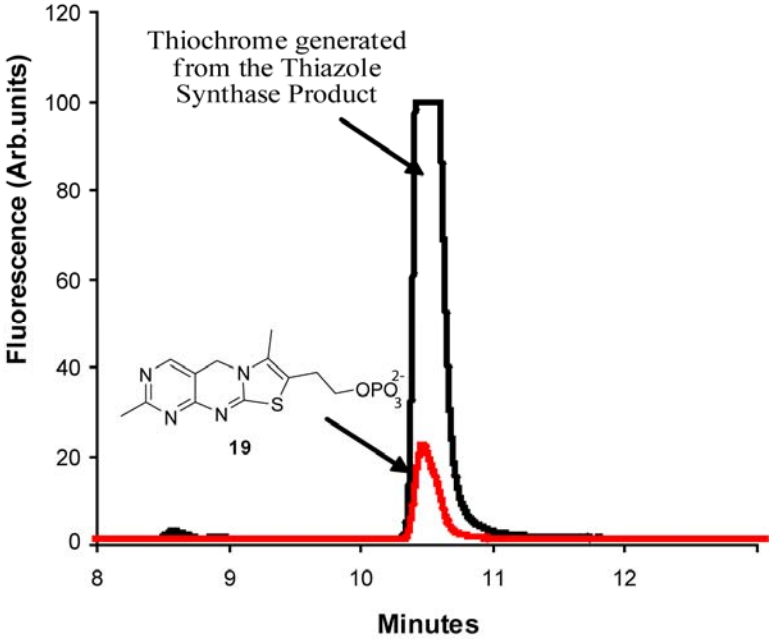


Figure 2.5: HPLC analysis showing the conversion of the peak A compound to thiochrome phosphate (red trace).

2.2.2 Isolation and characterization of the Peak A compound

The most direct way to identify the Peak A compound was to compare its chromatographic behavior with that of authentic samples of thiazoles **14** and **16**, the two most likely alternative products of the bacterial thiazole synthase-catalyzed reaction. Access to these compounds was greatly facilitated by our recent demonstration that species **24** and **25** copurify with the *Saccharomyces cerevisiae* thiazole synthase, an enzyme that catalyzes very different chemistry.[12,13] Release of these metabolites from the *S. cerevisiae* thiazole synthase by heat denaturation, followed by purification by reverse-phase HPLC and treatment of these metabolites with nucleotide pyrophosphatase, generated authentic samples of the required reference compounds (Figure 2.6).

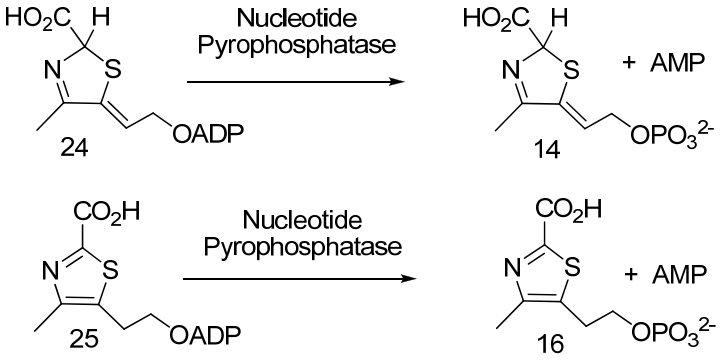


Figure 2.6: Procedures for the production of reference compounds **14** and **16**

HPLC analysis, by strong anion-exchange, clearly demonstrated that the Peak A compound comigrated with **14** (Figure 2.7 a).

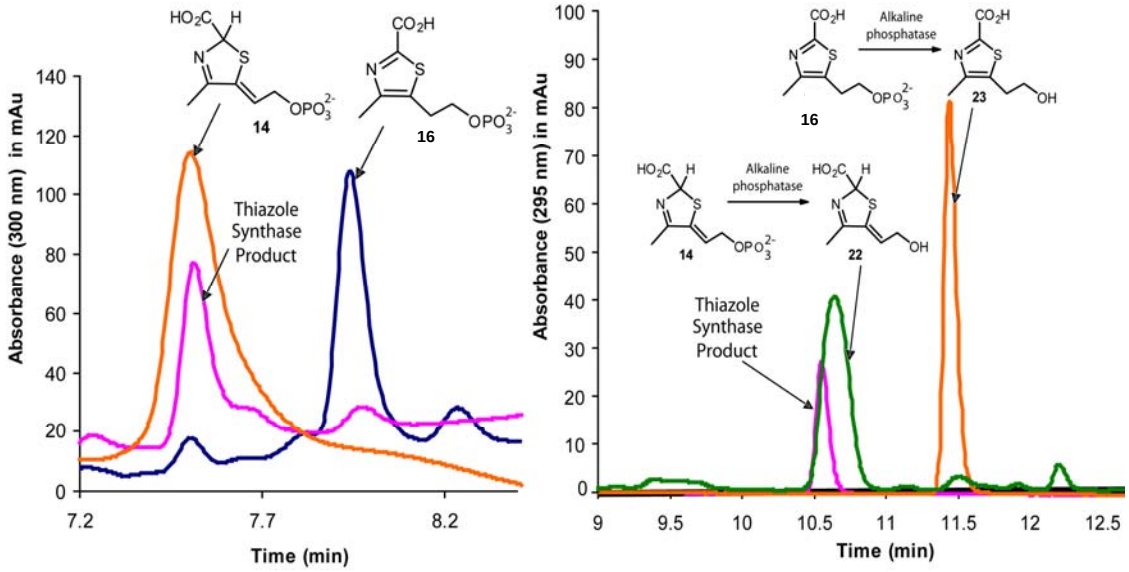


Figure 2.7: (a) - HPLC analysis of the thiazole synthase product (pink) and reference compounds **14** (orange) and **16** (blue). (b) - HPLC analysis of the dephosphorylated product of the thiazole synthase (pink) catalyzed reaction and reference compounds **22** (green) and **23** (orange).

To further confirm this identity, **14, 16** and the Peak A compound were dephosphorylated by treatment with alkaline phosphatase and the resulting alcohols were reanalyzed by reverse phase HPLC. Again, the dephosphorylated Peak A compound comigrated with **22**, the dephosphorylated product of **14** (Figure 2.7 b).

The thiazole tautomer **14** (Peak A compound) is difficult to isolate in quantities suitable for NMR analysis because it has to be purified out by HPLC, buffer exchanged into a volatile buffer, and then the buffer has to be lyophilized to obtain sufficient quantities of **14**. However, **14** decomposes extensively during the later stages of lyophilization presumably due to pH changes that occur during the lyophilization process. No cryoprotectants [18] could be used during lyophilization as added components would interfere with the NMR signals. The dilute samples of 14 used for the HPLC analysis however did not show this decomposition. Hence, alcohol **22** which is relatively stable during lyophilization, could be isolated in sufficient quantities and was used for spectroscopic analyses. 1D ^{1}H- and 2D-^{1}H-dqf-COSY spectra were collected, which are fully consistent with structure **22,** Figure 2.8.

Optimization of the complex reaction catalyzed by the bacterial thiazole synthase allowed for the unequivocal identification of the reaction product as the thiazole tautomer **14** rather than the thiazole **16**. This identification underscores the problem of identifying trace metabolites by enzyme-catalyzed derivatization, even when the derivatizing enzyme has been very well-studied. The stability of compounds **14** and **22** is surprising, as thiazole tautomers should readily aromatize. However, **22** is stable in the experimental time scale of purification and buffer- exchange by HPLC over 10 hours and lyophilization over 48 hours. This stability suggests that an as yet unidentified enzyme may be involved in the catalysis of this aromatization reaction.

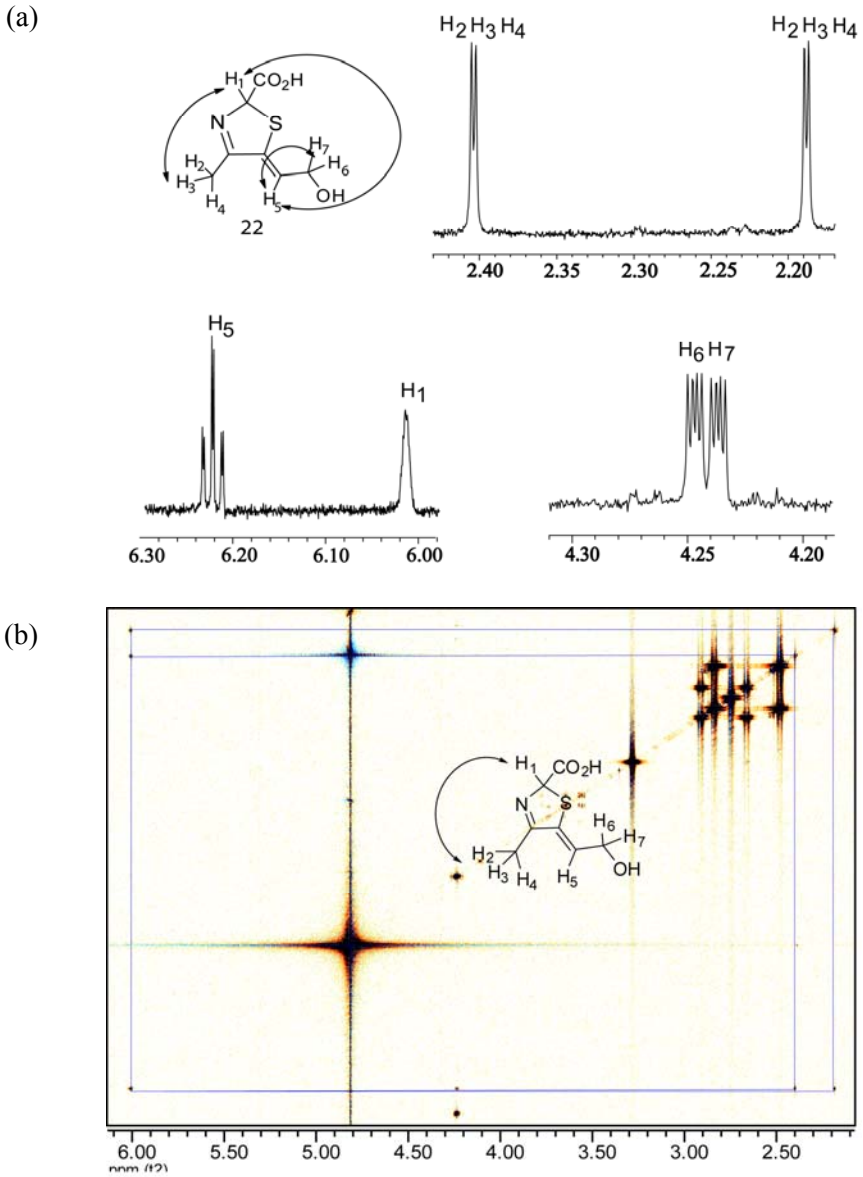
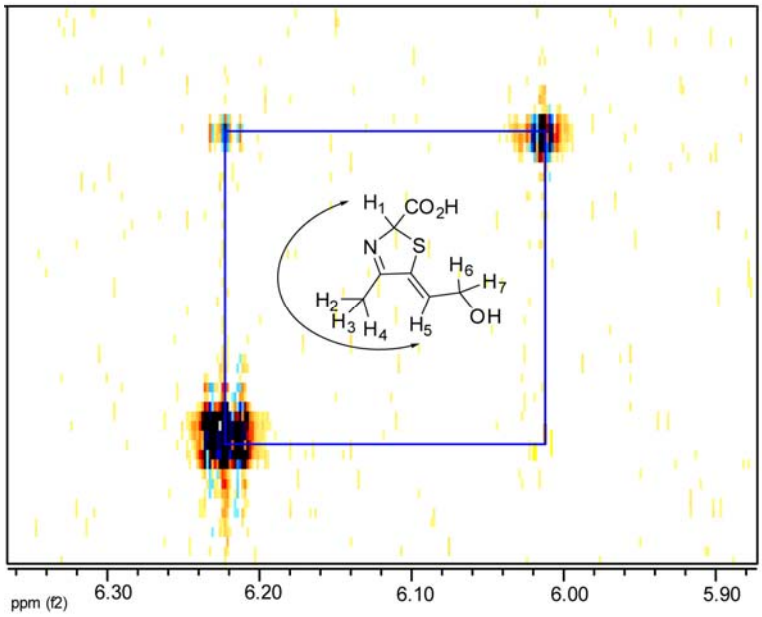


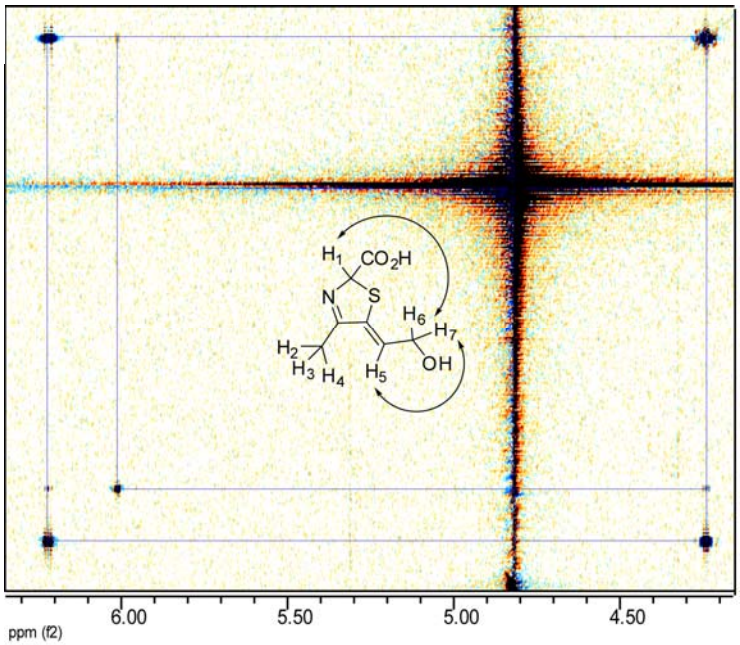
Figure 2.8: (a) ^{1}H-NMR analysis of the dephosphorylated Peak A compound 22. The DXP sample used was labeled with ^{13}C on the methyl group (unrelated reasons) hence the additional splitting of the $H_2/H_3/H_4$ protons. (b) Cross-peak for H1 (6.02 ppm) and H2/H3/H4 (2.2 ppm and 2.4 ppm) observed in the 2-D dqf-COSY experiment. (c) Cross-peak for H1 (6.02 ppm) and H5 (6.22 ppm) observed in the 2-D dqf-COSY experiment. (d) Cross-peak for H1 (6.02 ppm) and H5 (4.24 ppm) observed in the 2-D dqf-COSY experiment.

Figure 2.8 (Continued):

(c)



(d)



2.3 Conclusions

The structure of the product of the bacterial thiazole synthase had remained elusive to us because of its indirect detection by thiochrome derivatization. We were able to obtain conditions for the direct detection of the product of the thiazole biosynthesis reaction by UV-Vis while separating it out by HPLC and showed that the product of bacterial thiazole synthase was not the thiazole phosphate. Further characterization of the product revealed a very unusual molecule, the tautomer form of the thiazole carboxylate phosphate. This molecule is surprisingly stable, as it would be expected to easily tautomerize by loss of a proton or the carboxy group to form the thiazole carboxylate phosphate or the thiazole phosphate respectively. But it survives on the purification and characterization timescale. This opens up questions about the stability and aromaticity of heterocycles, and scope for the synthesis of this novel molecule. It also reiterates the importance of direct detection of biosynthetic metabolites to identify each step on the pathway. The stability of the **14** suggests the presence of another enzyme for its aromatization to either **15** or **16**

2.4 Experimental Methods

2.4.1 Source of Chemicals:

All chemicals and snake venom nucleotide pyrophosphatase were purchased from Sigma-Aldrich Corporation (USA) unless otherwise mentioned. Calf intestinal phosphatase was obtained from New England Biolabs. LB medium was obtained from EMD Biosciences. Kanamycin, ampicillin and IPTG were purchased from LabScientific Inc. NTA resin was the NTA superflow by Qiagen. The microcon membrane filters were from Millipore. Analytical HPLC (Agilent 1100 instrument) was carried out using a Phenomenex Gemini C18 110A (150x4.6 mm, 5 μm ID)

reverse phase column and a Phenosphere Strong Anion-Exchange (SAX) 80A (250x4.6 mm, 5 μm ID) column. HPLC purifications were carried out using a semi-prep Supelco LC-18-T (250x10 mm, 5 μm ID) column. HPLC grade solvents were obtained from Fisher Scientific. Previously synthesized stock of [1-^{13}C]-DXP [14] was used as the substrate of the thiazole reconstitution reactions.

2.4.2 Overexpression and purification of enzymes:

ThiSG, ThiF, NifS, ThiO and ThiE: *E. coli* BL21(DE3) containing the ThiSG overexpression plasmid (ThiG is co-purified with ThiS for stability) in pET16b was grown in LB medium containing ampicillin (40 μg/mL) with shaking at 37 °C until the OD_{600} reached 0.6. At this point, protein overexpression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration = 2 mM) and cell growth was continued at 15 °C for 16 h. The cells were harvested by centrifugation and the resulting cell pellets were stored at -80 °C. To purify the protein, the cell pellets from 1L of culture were resuspended in 25 mL lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, pH 8) and lysed by sonication (Heat systems Ultrasonics model W-385 sonicator, 2 s cycle, 50% duty). The resulting cell lysate was clarified by centrifugation and the ThiSG protein was purified on Ni-NTA resin following the manufacturer's instructions. After elution, the protein was desalted using a 10-DG column (BioRad) pre-equilibrated with 50 mM Tris-HCl buffer, pH 7.8. The remaining proteins ThiF (pET22), NifS (pET16), ThiO (pET22) ThiE (pQE32 and pREP4) were overexpressed and purified in a similar manner.[15,16] NifS, ThiO and ThiE were stored in aliquots at -80 °C in 20% glycerol. ThiSG and ThiF were purified immediately before use.

2.4.3 Reconstitution of the thiazole synthase catalyzed reaction on an analytical scale:

All solutions were made with 50 mM tris buffer, pH 8. Final concentrations of the reactants are given in parentheses. Cysteine (0.35 mM), DTT (0.70 mM), ATP (0.60 mM) and MgCl2 (3.5 mM) were incubated with purified ThiSG (1.25 μM), ThiF (1.24 μM) and 70 μL NifS (1.38 μM) for 1.5 hours. Total volume of this solution was 425 μL. Glycine (6.50 mM), DXP (0.33 mM), $MgCl_2$ (3.5 mM) and ThiO (6.8 μM) were then added to this reaction mixture and the final volume of the reconstitution mixture now was 610μL. This mixture was incubated for an additional 2 hours. The reaction mixture was then analyzed for product formation using the thiochrome assay (see below). In this reconstitution, 16% of the DXP was converted to product. This is a 3-fold improvement over our previously reported reconstitution, and corresponds to about 12 turnovers by the thiazole synthase.

2.4.4 Thiochrome Assay:

The thiochrome assay involves conversion of the thiazole product of the reconstitution to thiamin phosphate (**18**) and further to thiochrome phosphate. The product of the thiazole reconstitution is reacted with HMP-PP (**17**) (0.5 mM) in the presence of thiamin phosphate synthase (ThiE) (1.00 μM). The reaction is allowed to stand at room temperature for 2 hours and then quenched with an equal volume of 10% TCA. Potassium acetate (50 μL of 4M) is added to 100 μL of the quenched reaction followed by oxidative cyclization to thiochrome phosphate (**10**) using 50 μL of a saturated solution of $K_3Fe(CN)_6$ in 7M NaOH. The oxidation reaction is neutralized after 1 minute with 6M HCl and analyzed by reverse phase HPLC with fluorescence detection (excitation at 365 nm, emission at 450 nm). The following linear gradient, at a flow rate of 1 mL/min, was used. Solvent A is water, solvent B is 100 mM K2HPO4, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 10% A, 90%B; 10 min: 25%

A, 15% B, 60% C; 12 min: 25% A, 15% B, 60%; 15 min: 100% B; 17 min: 100%B. A time-course for the thiazole reconstitution is shown in supplementary Figure 2.3.

2.4.5 Reconstitution of the thiazole synthase-catalyzed reaction on a preparative scale:

All solutions were made with 50 mM tris buffer, pH 8. Cysteine (0.35 mM), DTT (0.70 mM), ATP (0.60 mM) and MgCl2 (3.5 mM) were incubated with purified ThiSG (1.25 μM), ThiF (1.24 μM) and NifS (1.38 μM) for 1.5 hours. Total volume of this solution was 1.3 mL. Glycine (6.50 mM), DXP (0.33 mM), MgCl2 (3.5 mM) and ThiO (6.8 μM) were added to this reaction mixture and the reconstitution solution now had a final volume of 1.8mL. This mixture was incubated for an additional 2 hours. The reaction mixture was then analyzed for product formation by reverse phase HPLC analysis, with UV detection. The following linear gradient, at a flow rate of 3 mL/min, was used: Solvent A is water, solvent B is 100 mM KPi, pH 6.6, solvent C is methanol. 0 min: 100% B; 5 min: 10% A, 90% B; 12 min: 25% A, 15% B, 60% C; 18 min: 25% A, 15% B, 60% C; 22 min: 100% B; 25 min: 100%B. A product, eluting at 2.8 min was observed. This product did not comigrate with thiazole phosphate (**16**) (Figure 2.4). The compound eluting at 2.8 min was collected and buffer exchanged into a low concentration of volatile ammonium acetate buffer by HPLC. The following linear gradient was used at a flow rate of 3 mL/min: Solvent A is water, solvent B is 25 mM NH_4OAc, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 10% A, 90%B; 6 min: 15% A, 20% B, 65% C; 8 min: 15% A, 20% B, 65%; 11 min: 100% B; 14 min: 100%B. The collected fractions were then pooled and lyophilized to successfully obtain the product of the bacterial thiazole synthase.

2.4.6 1D-^{1}H NMR and 2D-dqf-COSY NMR analyses

To prepare the thiazole tautomer alcohol **22** for 1D-1H NMR and 2D-dqf-COSY NMR studies, the product **14** that eluted out from the HPLC purification at 2.8min was collected and treated with 1 unit of calf intestinal phosphatase for 20 min at room temperature to form **22**. This was then buffer exchanged into a low concentration of volatile ammonium acetate buffer by HPLC. The following linear gradient was used at a flow rate of 3 mL/min: Solvent A is water, solvent B is 25 mM NH_4OAc, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 10% A, 90%B; 6 min: 15% A, 20% B, 65% C; 8 min: 15% A, 20% B, 65%; 11 min: 100% B; 14 min: 100%B. The collected fractions were then pooled and lyophilized to obtain the thiazole tautomer alcohol **22**, which was then used for 1D-^{1}H NMR and 2D-dqf-COSY NMR studies. A Shigemi NMR tube (susceptibility-matched for D_2O) was used for all the experiments, which were carried out on a Varian INOVA 600 MHz instrument equipped with a 5 mm triple gradient inverse-detection HCN probe.

2.4.7 Preparation of HPLC standards of 14, 22, 15 and 23:

Compound **14** was obtained by the following procedure - *S. cerevisiae* THI4p (thiazole synthase) plasmid was overexpressed in BL21(DE3) cells grown in LB media with shaking at 37 °C until the OD_{600} reached 0.3. At this point, protein overexpression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration = 2 mM) and cell growth was continued at 15 °C for 16 h. The cells were not allowed to grow till OD_{600} reached 0.6, as larger quantities of adenylated **14** are obtained from THI4p at lower OD's of the cell culture. The protein was then purified by Ni-NTA chromatography at a lower temperature (4°C). It was not desalted as the process of desalting results in the loss of the bound small molecules out of the protein. As we were interested in higher yields of the small molecules bound, the

protein was only eluted by elution buffer and further denatured as follows: THI4p from 4 L of culture (~200 mg, 10 mL) was divided into twenty 500 µL aliquots and heat denatured (100 °C, 2 minutes). The precipitated protein was removed by centrifugation and the supernatants were combined and filtered through a 10 kDa MW cut off microcon filter to obtain the stock of small molecules free of all proteins. Adenylated **14** was purified by HPLC by observing its absorbance at 254nm and 300nm using the following linear gradient at a flow rate of 3 mL/min: solvent A is water, solvent B is 100 mM KPi, pH 6.6, solvent C is methanol. 0 min: 100% B; 3 min: 10% A, 90%B; 17 min: 34% A, 60% B, 6% C; 21 min: 35% A, 25% B, 40% C; 23 min: 100%B and the collected fractions were pooled. A second HPLC purification, using a low concentration of volatile ammonium acetate buffer, was performed on the pooled fractions using the following linear gradient at a flow rate of 3 mL/min: Solvent A is water, solvent B is 25 mM NH_4OAc, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 10% A, 90%B; 6 min: 15% A, 20% B, 65% C; 8 min: 15% A, 20% B, 65%; 11 min: 100% B; 14 min: 100%B. The collected fractions were then lyophilized to yield micromolar quantities of adenylated **14**. The quantity of 14 obtained could be estimated by quantifying the AMP that would be released in an 1:1 ratio with the thiazole tautomer phosphate **14** upon treatment with nucleotide pyrophosphatase. **14** was then treated with 1 unit nucleotide pyrophosphatase at pH 7.2 to yield **14 (**Figure 2.6**)** and further with 1 unit calf intestinal phosphatase in phosphate buffer, pH 7.8 for 20 min. to yield **22.** Adenylated compound **15** is also found bound to *S. cerevisiae* THI4p and **15** was obtained by the same purification procedure as described above for the preparation of compound **14**. Compound **23** was synthesized by carboxylation of thiazole alcohol **26** using the reported literature procedure[17].

Acknowledgement. We acknowledge the help of Frank Schroeder with the NMR dqf-COSY experiment and David Hilmey for the synthesis of the thiazole carboxylate alcohol **22**.

REFERENCES

(1) R. F. Butterworth, *Nutr. Res. Rev*. **2003**, *16*, 277-283.

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

(3) J. M. Soriano, J. C. Moltó and J. Mañes, *Nutr. Res.* **2000**,*20*, 1249-1258

(4) T. P. Begley, D. M. Downs, S. E. Ealick, F. W. McLafferty,A. P. G. M. Van Loon, S. Taylor, N. Campobasso, H.-J.Chiu, C.Kinsland, J. J. Reddick, *J. Xi, Arch. Microbiol.* **1999**, *171*, 293-300.

(5) I. D. Spenser, R. L. White, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1032-1046.

(6) T. P. Begley, *Nat. Prod. Rep.* **1996**, *13*, 177-185.

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

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

(9) P. C. Dorrestein, H. Zhai, F.W. McLafferty, T. P. Begley, *Chem. Biol.* **2004**, *11*, 1373-1381.

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

(11) P. C. Dorrestein, H. Zhai, S. V. Taylor, F. W. McLafferty, T. P.Begley, *J. Am. Chem. Soc.* **2004**, *126*, 3091-3096.

(12) A. Chatterjee, C. T. Jurgenson, F. C. Schroeder, S. E. Ealick,T. P. Begley, *J. Am.Chem.Soc* **2006**, *128*, 7158-7159

(13) A. Chatterjee, F. C. Schroeder, C. T. Jurgenson S. E. Ealick and T. P. Begley, *J.Am.Chem.Soc* **2008**, *130*, 11394-11398

(14) S.V.Taylor, L.D.Vu, T.P.Begley, U.Schoerken, S.Grolle, G.A. Sprenger, S. Bringer-Meyer and H. Sahm *J. Org.Chem.* **1998**, *63*, 2375-2377

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

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

(17) N. Haginoya, S. Kobayashi, S. Komoriya, T. Yoshino, M. Suzuki,T. Shimada, K. Watanabe, Y. Hirokawa, T. Furugori and T. Nagahara *J.Med.Chem* **2004**, *47*, 5167-5182

(18) J. Rexroad, C. M. Wiethoff , L. S. Jones and C. R. Middaugh , *Cell Preservation Technology* **2002**, *1*, 91- 104

CHAPTER 3

Biosynthesis of the thiamin thiazole in *Bacillus subtilis*: TenI aromatizes the thiazole tautomer generated by B.Subtilis thiazole synthase

3.1 Introduction

Thiamin is an important cofactor in prokaryotes, lower eukaryotes and plants via a complex pathway. The biosynthesis of thiamin in microorganisms has been studied extensively and most of the genes involved have been characterized. (Figure 1.1) Detailed mechanistic studies have shown that the thiazole ring and the pyrimidine ring are biosynthesized by two separate unique mechanisms in bacteria and S. cerevisiae. In *Bacillus subtilis*, the thiazole moiety is formed by an oxidative condensation of glycine, 1-deoxy-D-xylulose-5-phosphate (DXP) and cysteine (Figure 2.1)[1-3] and HMP-PP **17** is produced by rearrangement of aminoimidazole ribonucleotide followed by phosphorylation[4-6]. Thiamin phosphate **18** is then formed by the coupling of the pyrimidine and the thiazole heterocycles (Figure 3.1).

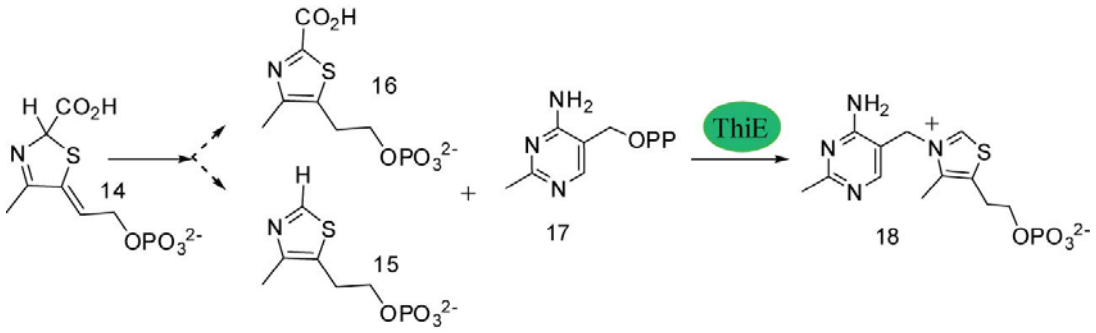


Figure 3.1: Product of ThiG **14** is converted to either **15** or **16** and coupled to HMP-PP to form thiamin phosphate **18**

A final phosphorylation gives thiamin pyrophosphate, the biologically active form of the cofactor.

In *B.subtilis* and many other bacteria, the genes *tenA* and *tenI* were found clustered together or fused with the genes involved in thiamin thiazole biosynthesis[7]. Studies indicated the involvement of TenA and TenI in regulation of extracellular enzyme production, though both these genes were not essential for the cell growth or extracellular enzyme production[8]. Neither *tenA* nor *tenI* shared any homology with any known regulatory proteins. Also, both TenA and TenI are known to be strongly

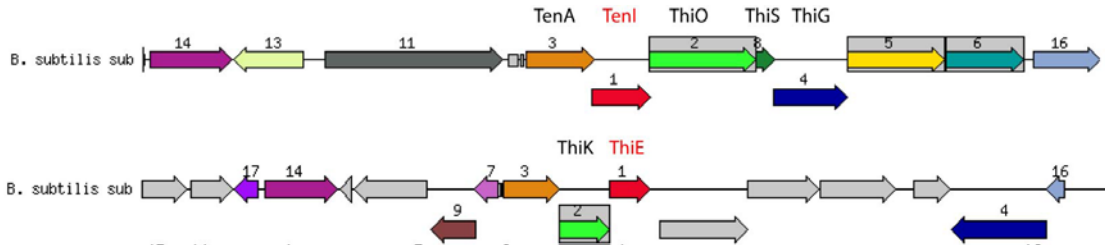


Figure 3.2: Gene neighborhood of TenI and ThiE in *B. subtilis*. In many organisms, TenI clusters with the thiamin biosynthesis genes and is found closely associated with TenA, which has been annotated as thiaminase II.

repressed by thiamin[9]. Subsequently, TenA was functionally[10] and structurally[11] characterized to be a thiaminase II involved in salvaging 4-amino-5-aminomethyl-2-methylpyrimidine to HMP-OH. However, no enzymatic role for TenI has been reported to date.

Orthologs of TenI have been detected in bacilli, clostridiae and various other classes of bacteria but it is found to be less widely distributed than TenA. *B. subtilis* TenI is a 22,783 Da protein[8] and interestingly, it shares a very strong primary sequence

homology with ThiE, the thiamin phosphate synthase particularly in the active site region. This strongly suggested that TenI may bind thiamin phosphate or one of the heterocyclic components of thiamin, though no thiamin phosphate synthase -like coupling activity was seen for TenI[12]. A crystal structure for TenI exists with a thiazole phosphate modeled into the active site, but the role of thiazole phosphate or its association with TenI was stated to be unclear[11]. In the following chapter, we functionally characterize TenI and describe its role in the prokaryotic thiazole biosynthesis pathway.

3.2 Results/ Discussion

3.2.1 TenI accelerates the rate of thiazole formation

The in-vitro reconstitution of *B. subtilis* thiazole pathway[13] was performed in the presence and the absence of TenI. Two sets of reconstitution reactions were set up, each one containing His-tagged proteins ThiF, NifS, ThiO, ThiSG and TenI overexpressed in *E. coli* BL21(DE3). ThiS-COOH (in complex with ThiG), NifS and ThiF were incubated with L-cysteine in the presence of dithiothreitol and ATP. At this point, to one reaction TenI was added, and to the other an equal volume of buffer was added. This reaction was allowed to incubate for 1.5 h to form ThiS-COSH. To each of the reactions, glycine and ThiO were added. To quantitate the thiazole, thiochrome phosphate would need to be produced which would involve making thiamin phosphate, hence HMP-PP and the thiamin phosphate synthase ThiE were added. DXP was added to initiate the reaction. 50 µL aliquots of each of the reactions were quenched with 10% TCA at time points of 0 min, 1 min, 2 min, 5 min, 10 min, 20 min, 60 min and 120 min.

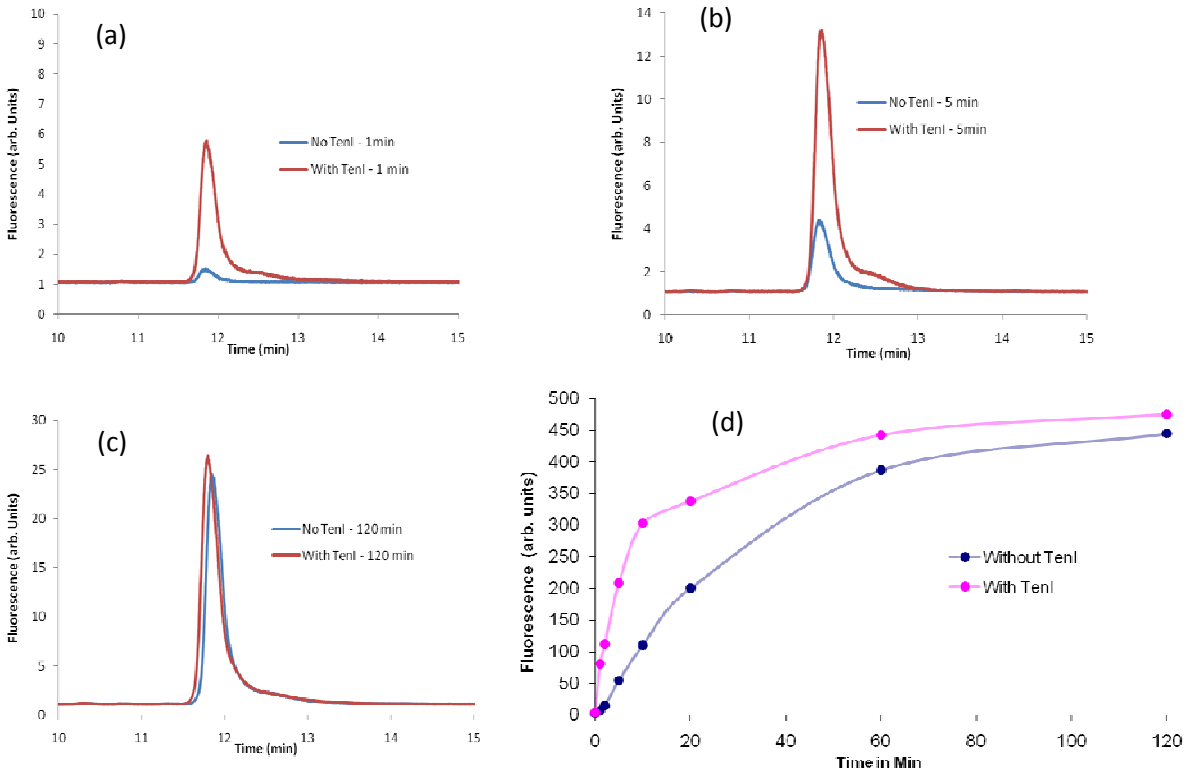


Figure 3.3: Thiochrome analysis of the thiazole reconstitution reaction in the presence and absence of TenI. (a), (b), (c) The red trace shows the thiochrome formed in the presence of TenI and blue trace shows the formation of thiochrome in the absence of TenI at different time points. (d) The rate of formation of thiazole in the presence of TenI is higher at initial timepoints of the reconstitution, however the final amount of thiazole formed is the same.

The thiamin phosphate formed at each of these timepoints for both the reactions was then oxidatively modified in the presence of basic $K_3Fe(CN)_6$ to thiochrome phosphate, neutralized, filtered and analyzed by fluorescence detection using RP-HPLC. It could be seen that in the absence of TenI, the initial rate of production of thiazole was slower than in the presence of TenI, but the total amount produced at the

end of 60 min was the same (Figure 3.3). Hence, TenI is not necessary for the production of thiamin, but its presence accelerates the rate of its production.

3.2.2 TenI does not affect the rate of glycine oxidase ThiO or thiamin phosphate synthase ThiE but does affect the rate of the thiazole synthase ThiG reaction

To analyze TenI's role in the thiazole reconstitution, the effect of addition of TenI was studied on different parts of the reaction. The glycine oxidase activity of ThiO was probed in the presence and absence of TenI by monitoring hydrogen peroxide formation as previously described[14]. It was seen the rate of formation of hydrogen peroxide with increasing concentrations of glycine in the presence of TenI was identical to the rate in the absence of TenI (Figure 3.4 a).

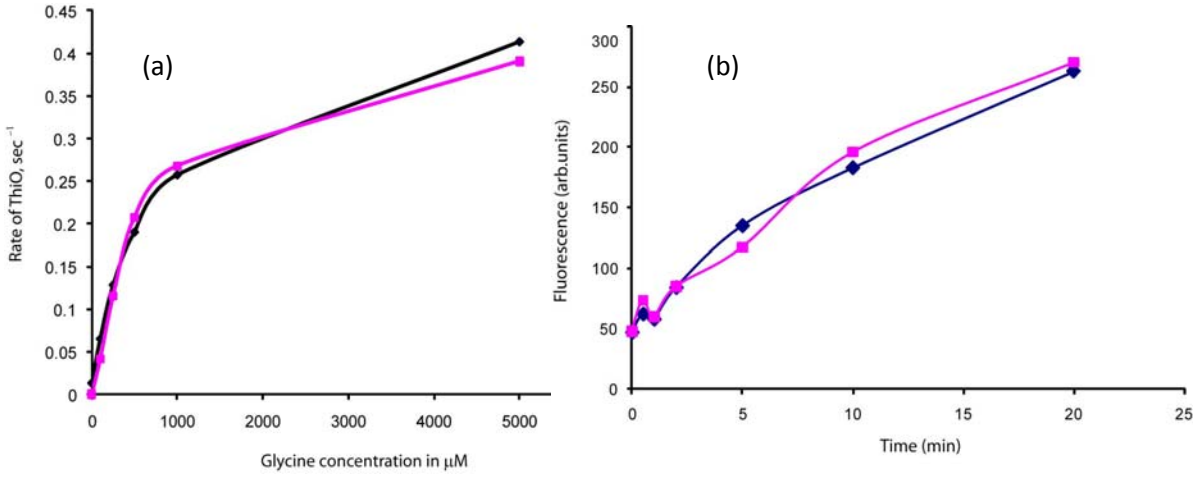
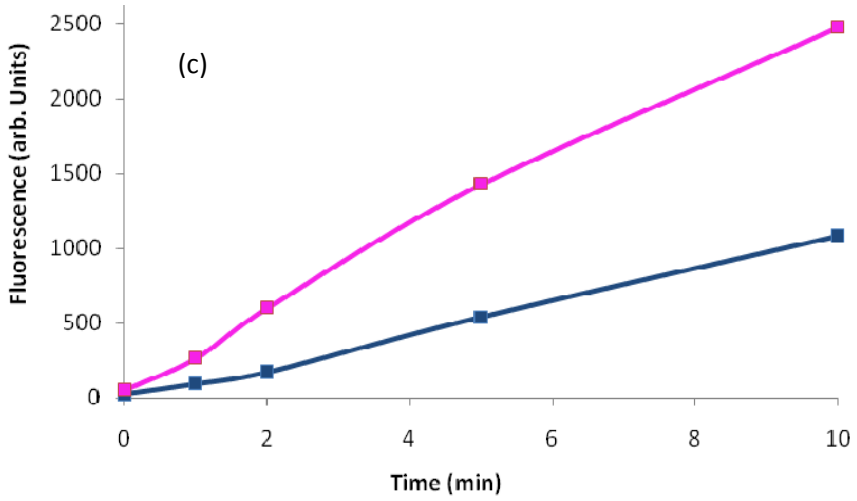


Figure 3.4: Blue trace is the reaction in the absence of TenI and pink trace is the reaction in the presence of TenI. (a) The presence of TenI does not affect the rate of glycine oxidase ThiO. (b) The rate of coupling of HMP-PP and thiazole phosphate by thiamin phosphate synthase ThiE is not affected (c) The rate of the thiazole synthase ThiG is affected by the presence of TenI

Figure 3.4 (Continued):



The possible involvement of TenI in the thiazole-HMP-PP coupling reaction by ThiE was similarly probed in the presence and absence of the enzyme using the thiochrome reaction[1]. The two reactions were quenched at different time points and the amount of thiochrome was quantified by measuring its fluorescence. TenI was seen to have no effect on the rate of coupling of thiazole phosphate to HMP-PP (Figure 3.4 b).

TenI was then added to the reconstitution reaction after formation of ThiS-COSH and after addition of glycine and ThiO, right before the step where the thiazole synthase ThiG gets involved in the reaction, and the formation of thiazole product was monitored by quenching the reaction at time points and monitoring the thiamin phosphate formed by the thiochrome assay. A clear increase in the rate of production of thiochrome could be seen at the initial stages of the reaction in the presence of TenI (Figure 3.4 c).

3.2.3 ThiE couples the thiazole phosphate (Thz-P) 15, thiazole carboxylate phosphate (Thz-C-P) 16 and the thiazole tautomer phosphate (Thz-T-P) 14 but the coupling is fastest for the Thz-C-P

As mentioned previously, TenI is not required for the formation of thiamin phosphate in the *in vitro* reconstitution system. This would mean that the thiamin phosphate synthase, ThiE can couple the Thz-P, Thz-C-P as well as the Thz-T-P with the HMP-PP to form thiamin phosphate. Also, we noted that the reconstitution in the presence of TenI is faster than in its absence and the total amount of thiamin phosphate formed at the end of the reconstitution reaction is the same. The effect of the rate difference can be explained by the fact that even though ThiE can couple the all three forms of thiazole substrate with HMP-PP to form thiamin phosphate and subsequently thiochrome phosphate, the rate of coupling of the three substrates varies. To compare the rates of coupling for the three thiazole substrates, three separate reaction with HMP-PP, $MgCl_2$ and purified ThiE was set up with Thz-T-P **14**, Thz-P **15** and Thz-C-P **16** instead of the Thz-T-P **14** were set. All three reactions were quenched at time points of 0, 1, 2, 5,15, 30, 180 min and the thiamin phosphate formed was converted to thiochrome phosphate for fluorescence analysis. It could be clearly seen that the formation of thiamin phosphate occurred fastest with the Thz-C-P, followed by Thz-P and then Thz-T-P during the initial time points and then became equal at 180 min. (Figure 3.5). Thus TenI helps to convert the thiazole tautomer phosphate to either thiazole carboxylate phosphate or thiazole phosphate, after which ThiE takes this up with HMP-PP and makes thiamin phosphate.

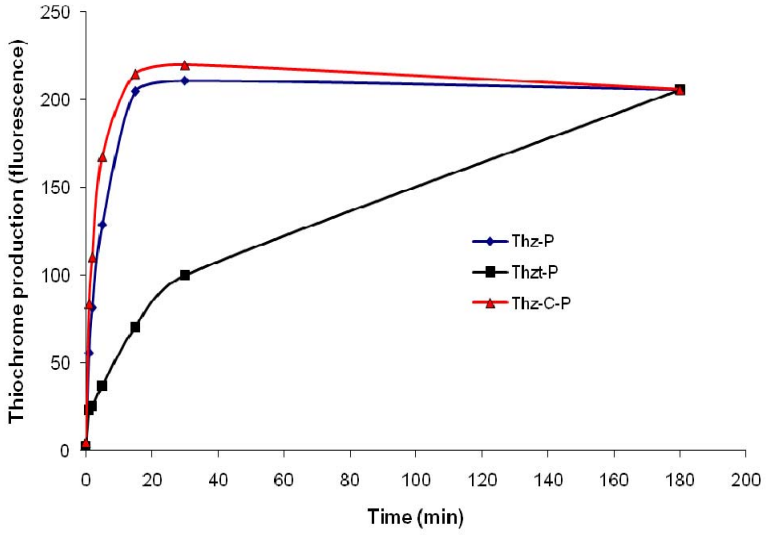


Figure 3.5: Relative rates of coupling of Thz-T-P **14,** Thz-P **15** and Thz-C-P **16** with HMP-PP by ThiE

3.2.4 TenI aromatizes the thiazole tautomer to thiazole carboxylate phosphate

We have shown previously that the product of the *B.subtilis* thiazole synthase is the thiazole tautomer **14** (Figure 2.1)[13], which is an unstable molecule with a half life of 5.3 days at room temperature (unpublished results). As this molecule does not spontaneously aromatize to give either Thz-C-P **16** or thiazole phosphate **14**, TenI was thought to be involved in this aromatization. To test this hypothesis, we needed substantial quantities of the substrate **14**. However, attempts to synthesize the thiazole tautomer **14** have been unsuccessful till date.

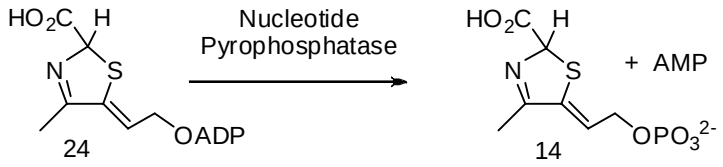


Figure 3.6: Making Thz-T-P **14** from **24**

Our source for **14** yet relies on **24,** one of the metabolites that copurifies out with the *Saccharomyces cerevisiae* thiazole synthase, an enzyme that catalyzes very different chemistry[15,16]. Release of these metabolites from the *S. cerevisiae* thiazole synthase by heat denaturation, followed by purification of **24** by RP-HPLC and further treatment with nucleotide pyrophosphatase, generated small amounts of an authentic sample of the required Thz-T **14** (Figure 3.6).

This was then incubated with TenI, and the UV-Vis profile of the reaction mixture was analyzed by RP-HPLC to observe the disappearance of the Thz-T-P (Figure 3.7, Peak A) to produce a new peak (Figure 3.7, peak B) that migrated very close to it.

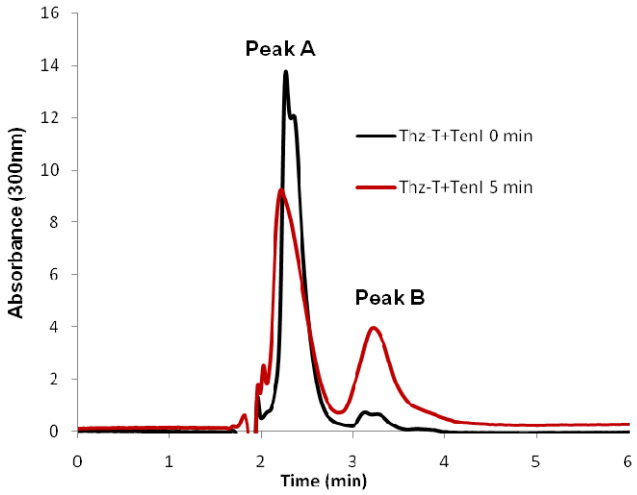


Figure 3.7: Thz-T-P **14** (peak A) converts to peak B in the presence of TenI

Peak B was confirmed to be Thz-C-P **16** and not Thz-P **15** by comigration experiments on the HPLC using synthetically prepared standards of **15** and **16**. To show that TenI can actually aromatize the product of the thiazole reconstitution reaction, a set of reconstitution reactions were performed where TenI was added to one reaction and not to the other. The reaction mixtures were analyzed by strong anion

exchange HPLC at 295nm. It could be seen that the product of the reconstitution reaction without TenI comigrated with the Thz-T **14** standard, whereas the product for the reconstitution reaction with TenI comigrated with the Thz-C-P **15** standard (Figure 3.8).

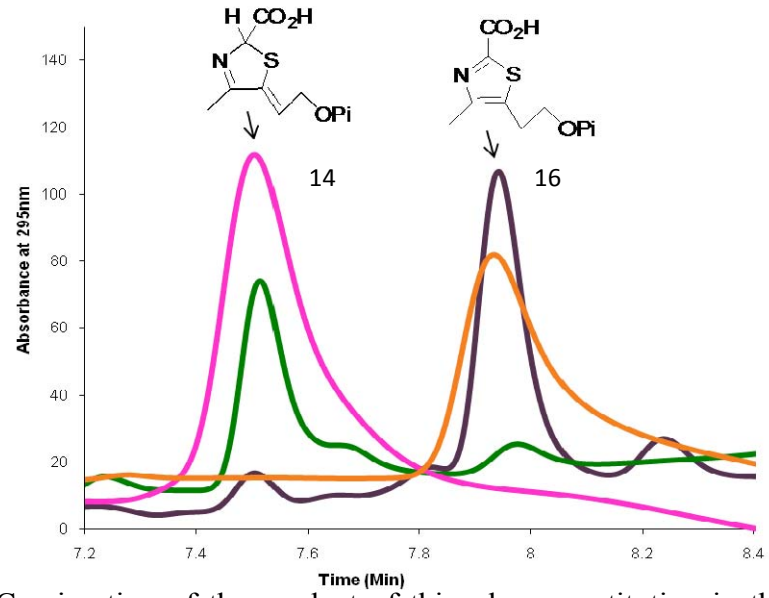


Figure 3.8: Comigration of the product of thiazole reconstitution in the absence of TenI (green) with the standard for Thz-T-P (pink) and comigration of the product of thiazole reconstitution in the presence of TenI (purple) with the standard for Thz-C-P (orange)

To further confirm the reaction product, standards of **14**, **15** and the two above reconstitution reactions were dephosphorylated by treatment with alkaline phosphatase and the resulting alcohols were reanalyzed by reverse phase HPLC. Again, the dephosphorylated product of the reconstitution reaction without TenI comigrated with dephosphorylated thiazole tautomer alcohol **22** standard, whereas the product for the reconstitution reaction with TenI comigrated with thiazole carboxylate alcohol **28** standard (Figure 3.9).

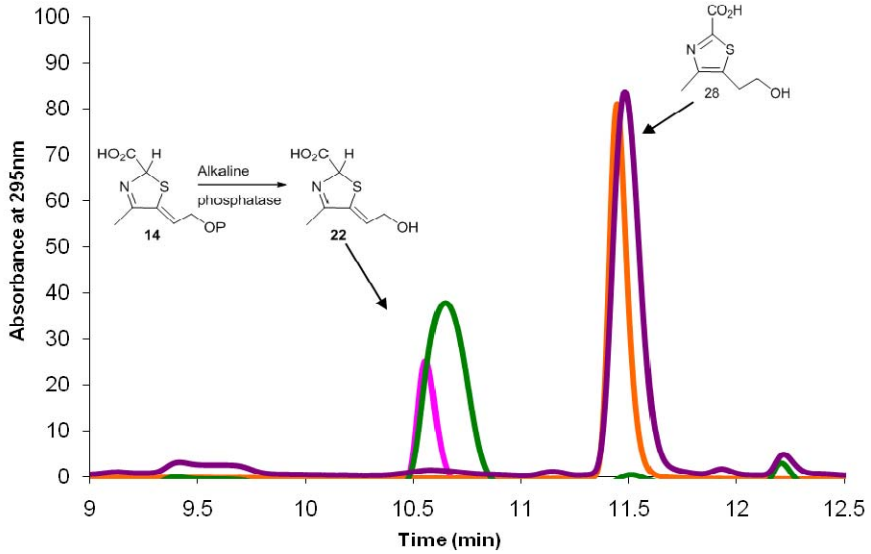


Figure 3.9: Comigration of alkaline phosphatase treated reconstitution reaction products - the product of thiazole reconstitution in the absence of TenI (green) with the standard for Thz-T-OH **22** (pink) and the product of thiazole reconstitution in the presence of TenI (purple) with the standard for Thz-C-OH **28** (orange)

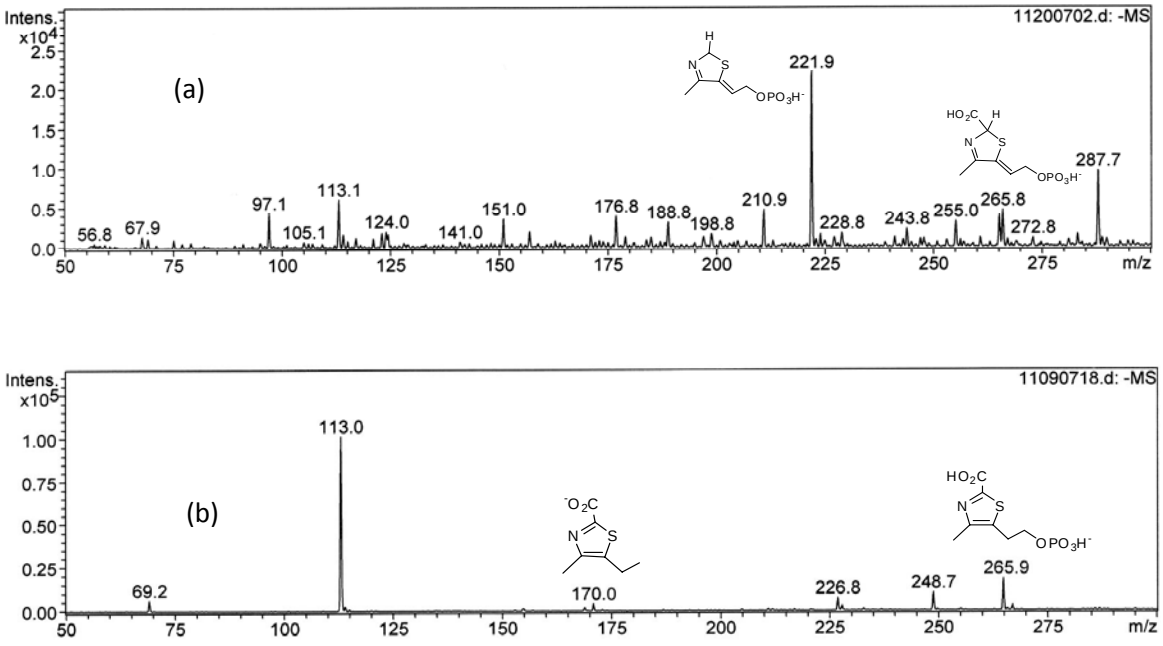


Figure 3.10: Negative mode ESI-MS of (a) Thz-T-P and (b) Thz-C-P

This conversion was verified by a negative mode ESI-MS analysis of the Thz-T-P and then of the product when Thz-T-P was treatd with TenI. It can be seen that the starting material and the product have the same m/z of 265.9, but the fragmenting pattern is different, and the fragmenting pattern of the product matches with that of synthesized Thz-C-P (Figure 3.10).

3.2.5 Thiazole carboxylate phosphate associates with TenI and has a dissociation constant of 32 μM

To study the affinity of TenI with Thz-C-P **15**, the K_d of the ligand-protein complex was determined by isothermal titration calorimetry. 4mM thiazole carboxylate phosphate was titrated in using an injector over 25 injections into the calorimeter reservoir containing 212 μM TenI and the heat of binding was determined for each titration (Figure 3.11 a). From the curve for the heat of binding after the ligand saturated the enzyme, the K_D was calculated to be 32 μM (Figure 3.11 b).

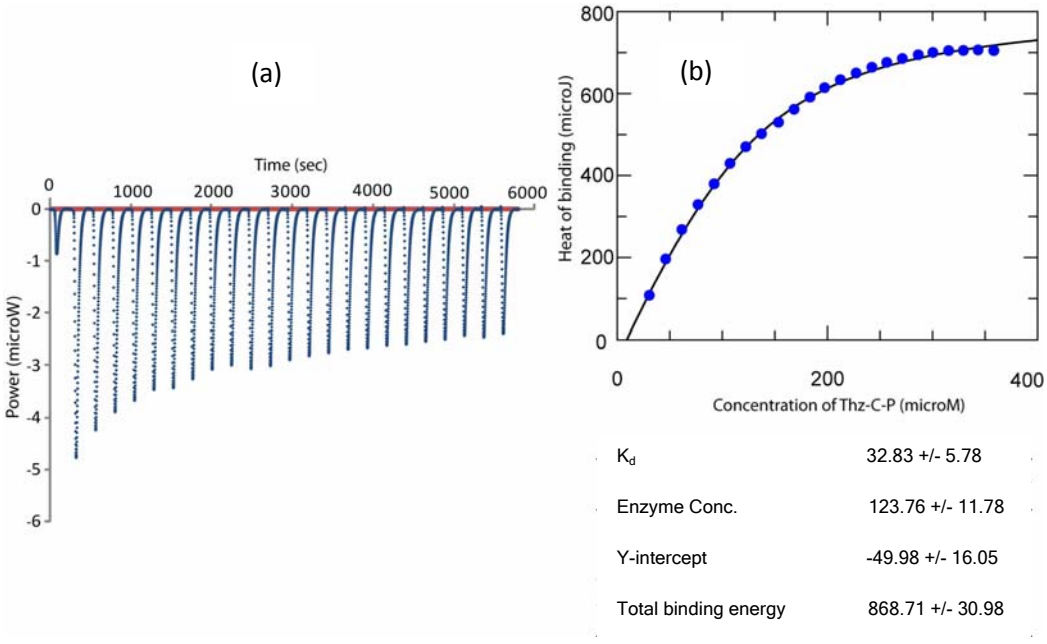


Figure 3.11: (a) Data showing the heat of binding of the ligand Thz-C-P to the active site of TenI and (b) a plot to obtain the dissociation constant of the Thz-C-P with TenI

3.2.6 Crystallization of TenI-Thz-C-P Complex (performed by Ying Han and Yang Zhang, Ealick group)

There previously existed a crystal structure for TenI with no small molecule bound to the active site (PDB: 1YAD REF). In order to analyze the active site of TenI and to look for residues that participate in this aromatization reaction, a crystal structure of TenI with the thiazole carboxylate phosphate bound to it was obtained. Crystals of ligand free TenI were grown from 1.65 – 1.75 M ammonium sulfate, 100 mM bicine, pH 8.7 – 9.6, 2% PEG400 (w/v), and 8 mM L-cysteine by hanging drop vapor diffusion method as previously described (*1*), and were used to obtain the product complex by soaking experiments. In order for the product thiazole carboxylate phosphate (TCP) to bind, the crystals were first dialyzed into 2.38 M sodium malonate (pH 7.0) (Hampton Research) to remove the sulfate ions, by gradually increasing sodium malonate concentration and decreasing ammonium sulfate concentration in 30 steps with 5 min incubation for each step. Subsequently the crystals were soaked overnight in 2.38 M sodium malonate, 21.5 mM Bicine (pH 9.0), 0.01% PEG400 (w/v), 0.5 mM L-cysteine, 4% glycerol and 11.5 mM TCP, followed by flash freezing in liquid nitrogen.

X-ray Data Collection and Processing: The X-ray intensity data of the TenI-TCP complex were measured at A1 beamline of the Cornell High Energy Synchrotron Source (CHESS) using a Quantum 210 CCD detector (Area Detector Systems Corp.). Data were collected over 180° using a 10 s exposure time and 1° oscillation per frame with a crystal to detector distance of 200 mm. The data were integrated and scaled using HKL2000 (*2*) and the structure was determined and refined (Experimental methods). The data processing statistics and refinement statistics are summarized in Table 3.1.

Structure Determination and Refinement:. The structure of TenI-TCP complex was determined by fourier synthesis using previously reported TenI structure (PDB: 1YAD) as the starting model. First round of rigid-body refinement with the starting model by REFMAC5 (*3*) reduced the *R*-factor to 0.286 (R_{free} 0.294). TCP was modeled in based on clear electron density. The model was refined through iterative cycles of restrained refinement by REFMAC5 and PHENIX (*4*), and manual rebuilding in Coot (*5*). Refinement statistics are shown in Table 3.1.

It can be seen that the asymmetric unit contains four monomers of TenI (Figure 3.12). The TenI momoner has a $(\beta\alpha)_8$ barrel fold (REF). Also, as previously reported, the most similar structural alignments by DALI search are with proteins from the FMN-dependent oxidoreductase and phosphate binding enzymes (FMOP) superfamily. The FMOP superfamily is characterized by a conserved phosphate binding pocket. The highest alignment of TenI was to the *B. subtilis* thiamin phosphate synthase ThiE, however it differs structurally from ThiE because of a critical substitution of Leu119 in ThiE for Gly119 in TenI, which does not permit the binding of substrates in the right conformation for coupling to form the thiamin phosphate.

Table 3.1: TenI crystal structure parameters

Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	98.05, 105.42, 219.26
a, b, g (°)	90, 90, 90
Resolution (Å)	50 – 2.23 (2.23 – 2.31)*
I / s*I*	18.9 (4.8)
Completeness (%)	99.8 (100)
Redundancy	6.1
Refinement	
Resolution (Å)	50 – 2.23 (2.23 – 2.31)
No. reflections	55266
R_{work} / R_{free}	0.206 (0.247)
No. atoms	
Protein	6142
Ligand/ion	50
Water	427
B-factors	
Protein	20.0
Ligand/ion	26.7
Water	31.5
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.4

*Values in parantheses are for the highest resolution shell

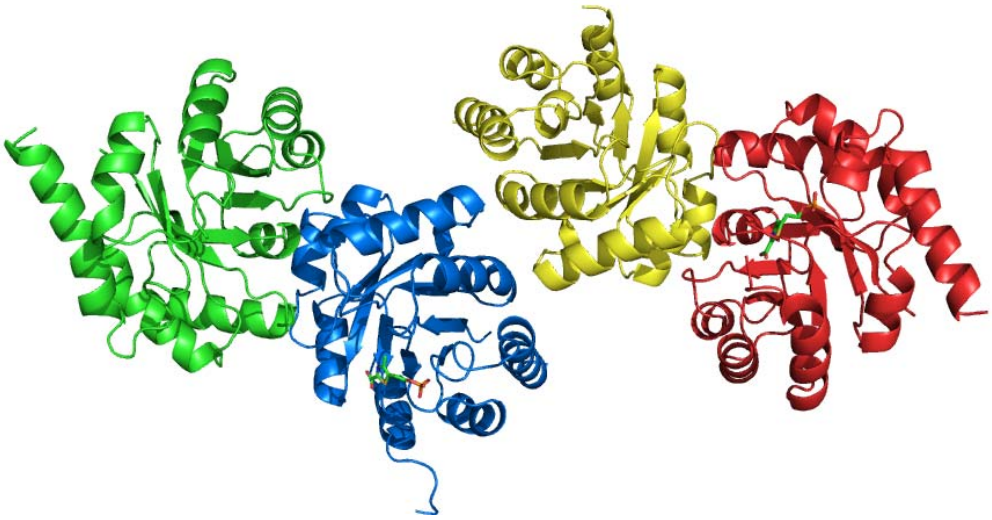


Figure 3.12: Assymetric unit of TenI structure contains four monomers

In the active site of the enzyme where the thiazole carboxylate phosphate is bound, two histidine residues, His102 and His122 are visible (Figure 3.13).

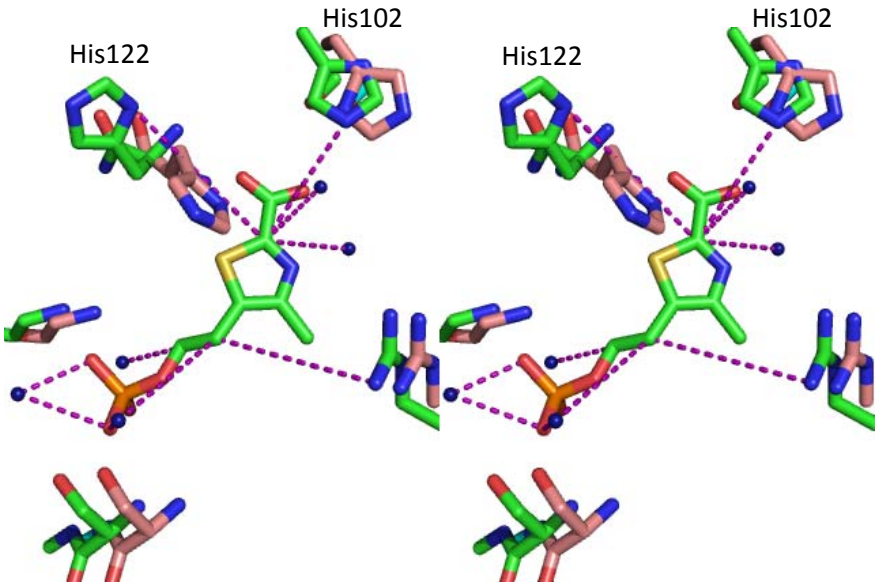
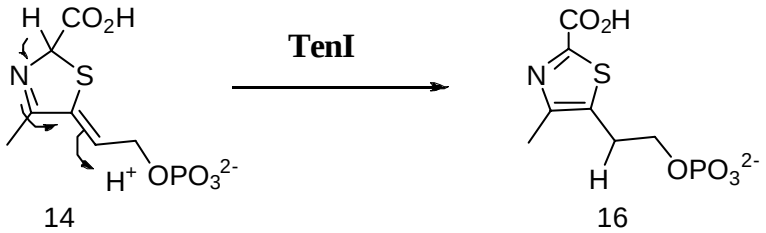


Figure 3.13: Stereoview of the active site of the apoenzyme TenI (pink) overlaid with the TenI structure with Thz-C-P **16** bound.

Upon overlaying the active site of the Thz-C-P bound protein with that of the apoenzyme crystal structure, we can see that almost all other residues in the active site remain at the same position, except the His122 residue. Also, a water molecule can be seen close to the Thz-C-P. It has been observed that His122 has slightly different conformations in the four chains of the crystals indicating flexibility of that residue inside the active site. This H122Q residue may be responsible for the deprotonation at the C-2 position of the thiazole tautomer phosphate and further protonation can occur either by water or by the phosphate group of the Thz-T-P[16]. Site-directed conserved mutants of His102 to glutamine (H102Q) and His122 to glutamine (H122Q) were made. Both the proteins were purified out using Ni-NTA chromatography and were soluble. The mutant H102Q showed activity of conversion of thiazole tautomer to thiazole carboxylate phosphate, but the H122Q mutant was mostly dead, as the thiazole tautomer was unaffected. The mutants H102A and H122A were also made, and H102A shows conversion of thiazole tautomer phosphate to thiazole carboxylate phosphate however, the H122A mutant does not show any conversion activity of the thiazole tautomer. Further biochemical and kinetic characterization of the protein TenI or the mutant H102Q could not be further characterized due to the lack of substrate.

Thus, the mechanistic proposal for the reaction by TenI is as shown in Figure 3.14. We propose that His 122 acts as a base to abstract the H^+ from the C-2 position of the Thz-T-P **14**.



14 16

Figure 3.14: Mechanistic proposal for TenI aromatization

From the position of His122 and its apparent flexibility in the active site, it seems that the His122 residue further provides the proton for reprotonation at the methylene position of the thiazole ring (C6 in Figure 3.14). The active structure shows the presence of both the His residues, His102 and His122 in the vicinity of Thz-C-P (16) which is bound in the active site (Figure 3.15).

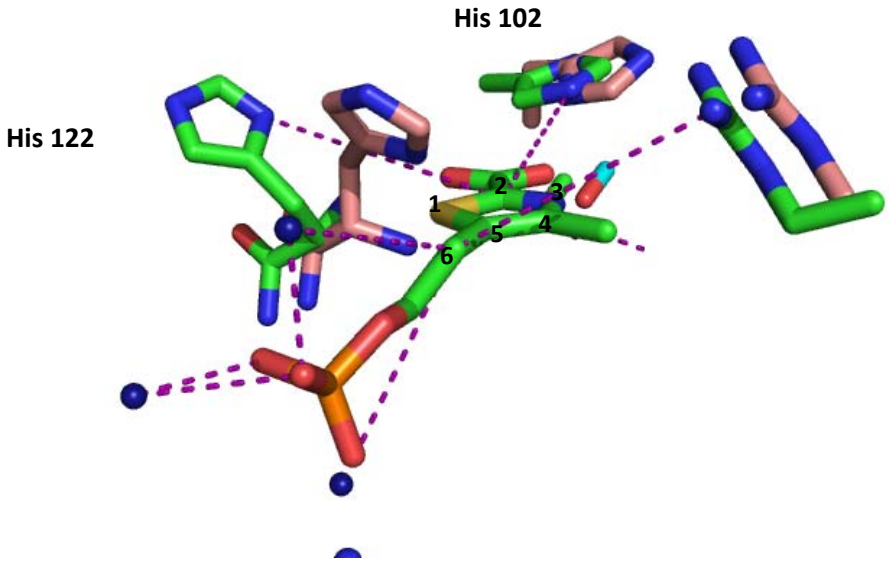


Figure 3.15: His102 and His122 in active site of TenI

If the abstraction of the proton at C2 position has to be done by the His residue(s), it appears that the Thz-T-P has an R-stereochemistry at the C2 center. This is consistent with observation for the *S. cerevisiae* thiazole tautomer-ADP is bound in the active site in a configuration such that the His237 (Figure 3.16). The His122 also appears in the vicinity of C6 and may be the residue to provide the product with the proton on aromatization. However, many structured water molecules appear in the active site too, which may be a likely proton donors too as they are spatially closer to the C6.

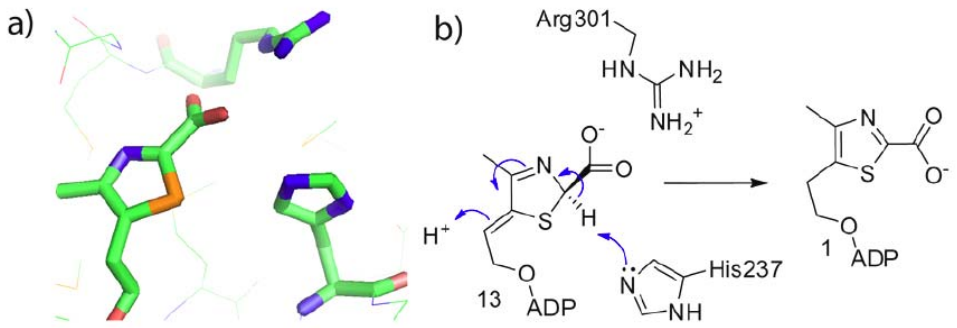


Figure 3.16: Determination of the absolute stereochemistry at C2 of the thiazole tautomer. a) Crystal structure of ADT bound to the active site of THI4p. The Arg301 residue of THI4p interacts with the carboxylate functionality of ADT. b) Proposed model for thiazole tautomer-ADP at the active site of THI4p, showing Arg301 interacting with the carboxylate functionality. In the proposed R configuration of thiazole tautomer-ADP, His237 and the beta phosphate of the ADP moiety are suitably positioned to catalyze the deprotonation/protonation reactions required for the aromatization of thiazole tautomer-ADP (as observed by A. Chatterjee and C. Jurgenson).

3.3 Conclusions

The functional role of the enzyme TenI has been elusive in thiamin biosynthesis even though the gene has been found associated with the thiamin biosynthetic cluster in many organisms. We have shown that TenI is the aromatase that takes the product of ThiG and aromatizes it to thiazole carboxylate phosphate, which is then further coupled to form thiamin phosphate. Further detailed characterization is unfortunately limited by the availability of the substrate. Structural characterization of TenI with the product, thiazole carboxylate phosphate bound helped identification of the active site and residues that may be involved in the aromatization reaction. Assignment of the

role of TenI reveals another interesting step in the complex biosynthesis of thiamin in prokaryotes.

3.4 Experimental Methods

3.4.1 Source of Chemicals:

All chemicals and snake venom nucleotide pyrophosphatase were purchased from Sigma-Aldrich Corporation (USA) unless otherwise mentioned. Calf intestinal phosphatase was obtained from New England Biolabs. LB medium was obtained from EMD Biosciences. Kanamycin, ampicillin and IPTG were purchased from LabScientific Inc. NTA resin was the NTA superflow by Qiagen. The microcon membrane filters were from Millipore. Analytical HPLC (Agilent 1100 instrument) was carried out using a Phenomenex Gemini C18 110A (150x4.6 mm, 5 μm ID) reverse phase column for thiochrome analysis, a Supelco LC-18-T (150x4.6 mm, 3 μm ID) column for thiazole reconstitution analysis and a Phenosphere Strong Anion-Exchange (SAX) 80A (250x4.6 mm, 5 μm ID) column for the anion exchange chromatography. HPLC purifications were carried out using a semi-prep Supelco LC-18-T (250x10 mm, 5 μm ID) column. HPLC grade solvents were obtained from Fisher Scientific. Previously synthesized stock of [1-^{13}C]-DXP [14] was used as the substrate of the thiazole reconstitution reactions and and Thz-C-P and Thz-C-OH (synthesized by Dr. David Hilmey, Begley lab) were used as the reference for TenI product .

3.4.2 Overexpression and purification of enzymes for reconstitution:

ThiSG, ThiF, NifS, ThiO, ThiE and TenI: *E. coli* BL21(DE3) containing the ThiSG overexpression plasmid (ThiG is co-purified with ThiS for stability) in pET16b was grown in LB medium containing ampicillin (40 μg/mL) with shaking at 37 °C until the

OD_{600} reached 0.6. At this point, protein overexpression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration = 2 mM) and cell growth was continued at 15 °C for 16 h. The cells were harvested by centrifugation and the resulting cell pellets were stored at -80 °C. To purify the protein, the cell pellets from 1L of culture were resuspended in 25 mL lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, pH 8) and lysed by sonication (Heat systems Ultrasonics model W-385 sonicator, 2 s cycle, 50% duty). The resulting cell lysate was clarified by centrifugation and the ThiSG protein was purified on Ni-NTA resin following the manufacturer's instructions. After elution, the protein was desalted using a 10-DG column (BioRad) pre-equilibrated with 50 mM Tris-HCl buffer, pH 7.8. The remaining proteins ThiF (pET22), NifS (pET16), ThiO (pET22) ThiE (pQE32 and pREP4) and TenI (pET28b) were overexpressed and purified in a similar manner.[15,16] NifS, ThiO and ThiE were stored in aliquots at -80 °C in 20% glycerol. ThiSG, TenI and ThiF were purified immediately before use.

3.4.3 Reconstitution of the thiazole synthase catalyzed reaction on an analytical scale (in the presence or absence of TenI):

All solutions were made with 50 mM tris buffer, pH 8. Final concentrations of the reactants are given in parentheses. Cysteine (0.35 mM), DTT (0.70 mM), ATP (0.60 mM) and MgCl2 (3.5 mM) were incubated with purified ThiSG (1.25 μM), ThiF (1.24 μM) and 70 μL NifS (1.38 μM) for 1.5 hours. Total volume of this solution was 425 μL. Glycine (6.50 mM), DXP (0.33 mM), $MgCl_2$ (3.5 mM) and ThiO (6.8 μM) were then added to this reaction mixture and the final volume of the reconstitution mixture now was 610μL. TenI was added in the reconstitution reaction to a final concentration of 10 μM to check for the acceleration of rate of thiazole formation. In the control reaction, the same volume of buffer was added into the reaction. This mixture was

incubated for an additional 2 hours. For reactions where timepoints for the formation of thiazole were being measured, the reaction was quenched at various timepoints like 0 min, 1 min, 2 min, 5 min, 10 min, 20 min, 60 min and 120 min. The reaction mixture was then analyzed for product formation using the thiochrome assay (see below). In this reconstitution, 16% of the DXP was converted to product. This is a 3-fold improvement over our previously reported reconstitution, and corresponds to about 12 turnovers by the thiazole synthase.

3.4.4 Thiochrome Assay:

The thiochrome assay involves conversion of the thiazole product of the reconstitution to thiamin phosphate (**18**) and further to thiochrome phosphate. The product of the thiazole reconstitution is reacted with HMP-PP (**17**) (0.5 mM) in the presence of thiamin phosphate synthase (ThiE) (1.00 μM). The reaction is allowed to stand at room temperature for 2 hours and then quenched with an equal volume of 10% TCA. Potassium acetate (50 μL of 4M) is added to 100 μL of the quenched reaction followed by oxidative cyclization to thiochrome phosphate (**10**) using 50 μL of a saturated solution of $K_3Fe(CN)_6$ in 7M NaOH. The oxidation reaction is neutralized after 1 minute with 6M HCl and analyzed by reverse phase HPLC with fluorescence detection (excitation at 365 nm, emission at 450 nm). The following linear gradient, at a flow rate of 1 mL/min, was used. Solvent A is water, solvent B is 100 mM K2HPO4, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 10% A, 90%B; 10 min: 25% A, 15% B, 60% C; 12 min: 25% A, 15% B, 60%; 15 min: 100% B; 17 min: 100%B.

3.4.5 Assay for ThiO activity in the presence and absence of TenI:

25 mL of assay solution containing 4mM phenol, 100mM 4-amino-antipyrene and 2units/mL HRP) was made. To 500 uL of the assay solution, 10mM, 5mM, 1mM, 500

μM, 250 μM and 100 μM glycine was added and the volume each time was adjusted to the 505 uL. To each reaction was added ThiO to a final concentration of 6.6 uM and TenI to a final concentration of 10uM to initiate the reaction. A parallel set of reactions were similarly done where only ThiO was added to a concentration of 6.6 uM and buffer was added in place of TenI to initiate the reactions. The rate of ThiO was measured with regard to the concentration of glycine in the presence and absence of TenI by measuring 500nm constant visible wavelength for 600sec for each sample and data for the rate of ThiO for different concentrations of glycine was generated.

3.4.6 Assay for ThiE activity in the presence and absence of TenI:

437 uM of Thz-P and 485uM HMP-PP were mixed with 10uM ThiE and 10uM TenI in a 500uL reaction solution with 2mM $MgCl_2$. A parallel reaction was set up with 437 uM of Thz-P and 485uM HMP-PP were mixed with 10uM ThiE and buffer instead of TenI in a 500uL reaction solution with 2mM $MgCl_2$. 100 uL of the two solutions were quenched at time points 0min, 0.5min, 1min, 2 min, 5min, 10min and 20min, thiochrome oxidation of the product was carried out and then fluorescence was analyzed by HPLC.

3.4.7 Assay for ThiE activity in the presence and absence of TenI:

Two sets o thiazole reconstitution reaction as mentioned previously was carried out. In one reaction, TenI was added to the reaction just before adding the DXP and in the other reaction, same volume of buffer was added to the reaction just before adding the DXP. Timepoints were taken at 0min, 1min, 2min, 5min and 10min.

3.4.8 Making Thz-T-P 14 and Thz-T-OH 22 from 24:

Compound **14** was obtained by the following procedure - overexpressed *S. cerevisiae* THI4p (thiazole synthase) protein was denatured as follows: THI4p from 4 L of culture (~200 mg, 10 mL) was divided into twenty 500 μL aliquots and heat denatured (100 °C, 2minutes). The precipitated protein was removed by centrifugation and the supernatants were combined and filtered through a 10 kDa MW cut off microcon filter. Adenylated **14** was purified by HPLC using the following linear gradient at a flow rate of 3 mL/min: solvent A is water, solvent B is 100 mM KPi, pH 6.6, solvent C is methanol. 0 min: 100% B; 3 min: 10% A, 90%B; 17 min: 34% A, 60% B, 6% C; 21 min: 35% A, 25% B, 40% C; 23 min: 100%B and the collected fractions were pooled. A second HPLC purification, using a low concentration of volatile ammonium acetate buffer, was performed on the pooled fractions using the following linear gradient at a flow rate of 3 mL/min: Solvent A is water, solvent B is 25 mM NH_4OAc, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 10% A, 90%B; 6 min: 15% A, 20% B, 65% C; 8 min: 15% A, 20% B, 65%; 11 min: 100% B; 14 min: 100%B. The collected fractions were then lyophilized to yield micromolar quantities of adenylated **14**. This was then treated with 1 unit nucleotide pyrophosphatase at pH 7.2 to yield **14** and further with 1 unit calf intestinal phosphatase in phosphate buffer, pH 7.8 for 20 min. to yield **22**

3.4.9 HPLC conditions for separation of Thz-T-P and Thz-C-P using analytical strong anion exchange column chromatography

The following linear gradient at a flow rate of 1 mL/min on the Phenosphere Strong Anion-Exchange (SAX) 80A (250x4.6 mm, 5 μm ID) column was used: solvent A is water, solvent B is 100 mM $(Et)_3NHOAc$, pH 7.8. 0 min: 100% A; 1 min: 100% A; 4 min: 100% B; 7 min: 100% A; 10 min: 100%A.

3.4.10 HPLC conditions for separation of Thz-T-OH and Thz-C-OH using the analytical reverse column chromatography

The following linear gradient at a flow rate of 1 mL/min on the Supelcosil LC-18-T (150x4.6 mm, 3 μm ID) was used: solvent A is water, solvent B is 100 mM KPi, pH 6.6, solvent C is methanol. 0 min: 100% B; 2 min: 100% B; 4min 10% A, 90%B; 9 min: 10% A, 25% B, 65% C; 14 min: 10% A, 25% B, 65% C; 16 min: 100%B; 20min 100%B.

3.4.11 ITC experiment to measure dissociation constant of TenI and its product Thz-C-P

A solution containing 50 mM tris, pH 7.6, 100mM NaCl, 8mM $MgCl_2$, 2mM TCEP and 123 μM TenI was prepared. A similar solution containing 50 mM tris, pH 7.6, 100mM NaCl, 8mM $MgCl_2$, 2mM TCEP and 4mM Thz-C-P was made. The TenI solution was pre-equilubrated to a stable temperatue in the calorimeter of the instrument and subsequently, the product (Thz-C-P) was added in increasing amounts into the enzyme TenI in 25 runs, 4 μL injections with 240 sec. intervals.

3.4.12 For crystallography: Expression and Purification of Bacillis subtilis

The recombinant plasmid TenIpET28a was transformed to the *E. coli* BL21(DE3) competent cells (Invitrogen). The transformed cells were grown at 37 °C with vigorous agitation (200 rpm) in Luria broth (LB) containing 30 μg/mL kanamycin to an OD_{600} of 0.7, at which point the cells were induced with 500 μM IPTG (Gold Biotechnology, Inc.) and allowed to incubate overnight at 22 °C under conditions of mild mixing (180 rpm). The cells were harvested by centrifugation (6,000 g) for 15 min at 4 °C and stored at -80 °C for later use. All purification steps were carried out at

4 °C. The cell pellet was suspended in 50 mL binding buffer (50 mM sodium phosphate, pH 7.0 and 300 mM NaCl), and lysed by sonication. The crude extract was centrifuged at 4 °C for 30 min at 50,000 g and the resulting supernatant was augmented with 5 mM imidazole and loaded onto a column containing 2 mL of TALON metal affinity resin (BD Biosciences) equilibrated with 50 mL binding buffer. The column was washed with 20 column volumes of binding buffer plus 5 mM imidazole, followed by 5 column volumes of binding buffer plus 10 mM imidazole. The six-His-tagged TenI was eluted from the column with elution buffer (50 mM sodium phosphate, pH 7.0, 300 mM NaCl and 300 mM imidazole). The recombinant TenI was further purified by a Superdex 200 gel-filtration column (Pharmacia) and eluted in the storage buffer (25 mM Tris-HCl, pH 8.5, 150 mM NaCl, and 1 mM thiamin-phosphate). The fractions containing pure TenI were combined and concentrated to 12 mg/mL using a 10 kDa cutoff concentrator (Amicon) and stored at -80 °C. Protein concentration was determined by the Bradford method with bovine serum albumin as the standard. The purity of TenI was determined by SDS-PAGE analysis and found to be greater than 99%.

REFERENCES

1. Park, J.-H.; Dorrestein, P.C.; Zhai, H.; Kinsland,C.; McLafferty F.W.; Begley, T.P. *Biochemistry* **2003**, 42, 12430-12438
2. Dorrestein, P.C., Zhai, H., McLafferty, F.W.; Begley, T.P. *Chemistry and Biology* **2004** 11, 1373–1381
3. Settembre, E.C.; Dorrestein, P.C., Zhai, H., Chatterjee, A,; McLafferty, F.W. Begley, T.P.; Ealick, S.E. *Biochemistry* **2004**, 43, 11647-11657
4. Newell P. C.; Tucker R.G.; *Biochem J.* **1968,**106, 279-87
5. Martinez-Gomez N.C; Downs D.M.; *Biochemistry* **2008**, 47, 9054-6.
6. Chatterjee A.; Li Y.; Zhang Y.; Grove T.L.; Lee M.; Krebs C.; Booker S.J.; Begley T.P.; Ealick S.E. *Nat Chem Biol.* **2008**, 4, 758-65
7. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand,M. S. *J. Biol.Chem.* **2002,** 277, 48949-48959.
8. Pang, A. S. H., Nathoo, S., and Wong, S. L. *J. Bacteriol.***1991**, 173, 46-54.
9. Lee, J.-M., Zhang, S., Saha, S., Santa Anna, S., Jiang, C., and Perkins, J. *J. Bacteriol.* **2001** 183, 7371-7380.
10. Jenkins, A.H.; Schyns, G.; Potot,S.; Sun, G.; Begley, T.P. *Nat. Chem.Bio* **2007** 3, 492-497
11. Toms, A.V., Haas, A.L., Park, J.-H., Begley, T.P. & Ealick, S.E. *Biochemistry* **2005** 44, 2319–2329
12. Chiu, H., Reddick, J., Begley, T., and Ealick, S. *Biochemistry* **1999** 38, 6460-6470.
13. Hazra. A.; Chatterjee, A.; Begley, T.P. *J.Am.Chem.Soc.* **2009** 131, 3225-9.
14. Nishiya, Y., and Imanaka, T. *FEBS Lett.* **1998** 438, 263-266

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

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

CHAPTER 4

Biosynthesis of the thiamin thiazole in *Bacillus subtilis*: Reversibility of ThiG

4.1 Introduction

In Chapter 2, we were able to show that the product of ThiG is the Thz-T-P **14,** and is not the Thz-P **15** or the Thz-C-P **16**. Thz-T-P **14** was interestingly very stable and was characterized extensively by HPLC, NMR and MS to reveal its structure[1]. We then observed that in the overall reconstitution reaction, the rate of thiamin formation in the presence of TenI is faster than in the absence of TenI. From the sequence homology and structural information we had about TenI, we hypothesized that it is required to aromatize Thz-T-P to a stable product, either Thz-P **15** or Thz-C-P **16**. We used purified Thz-T-P **14** to prove that TenI was the enzyme responsible for converting it to Thz-C-P **16**. The Thz-C-P is further taken up by ThiE along with HMP-PP to form thiamin phosphate (unpublished results) (Figure 4.1).

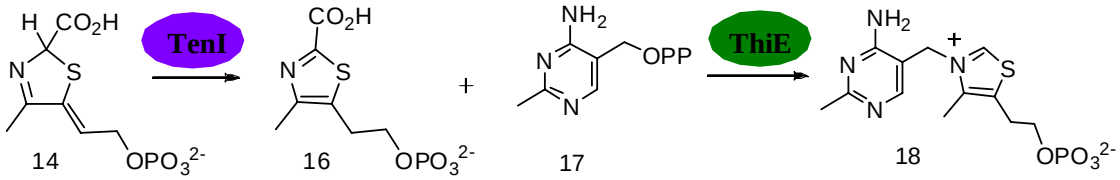


Figure 4.1: TenI aromatizes the product of ThiG to Thz-C-P 16, which is further coupled by ThiE with HMP-PP **17** to form thiamin phosphate **18**.

It was seen that ThiE can non-selectively take up Thz-T-P, Thz-C-P and Thz-P and couple it with HMP-PP to make thiamin phosphate. However, it is interesting to note that the rate of coupling is lowest for Thz-T-P, followed by Thz-P and is fastest for

Thz-C-P (Figure 3.4). This reiterates the role of TenI to convert Thz-T-P **14** to Thz-C-P **16**, hence accelerating the rate of formation of thiamin phosphate *in vivo*. In an effort to probe the stability of Thz-T-P **14**, it was added back to ThiG. On examination of the reaction, there appeared to be a reversible reaction of Thz-T-P with ThiG. This reversibility is discussed in this chapter.

4.2 Results/ Discussion:

In our attempts to characterize the role of TenI further, we decided to examine the reactivity of **14** with TenI in the presence and absence of ThiG (pure **14** was obtained as indicated in Figure 3.5). We wanted to measure the rate of conversion of thiazole tautomer to thiazole carboxylate phosphate by TenI in the presence and absence of ThiG. In this reaction, the purified enzymes ThiSG and TenI were obtained. In two separate reactions, Thz-T-P 14 was treated with TenI and ThiSG, and only TenI respectively. The concentration of the enzymes was approximately equal to the amount of Thz-T-P 14 present in the reactions. The components of the reaction were allowed to incubate and time points were obtained for 0min, 2min and 5min by quenching. The proteins were then filtered off and the filtrate was analyzed on a RP-HPLC column for the UV-Vis absorbance of its components. It was seen that when Thz-T-P was added in the absence of ThiG, some of it had converted to the Thz-C-P **16** (Figure 4.2 a). However when ThiG was present along with Thz-T-P and TenI, the thiazole tautomer disappeared and no other chromophoric signal was seen in the chromatogram (Figure 4.2 b).

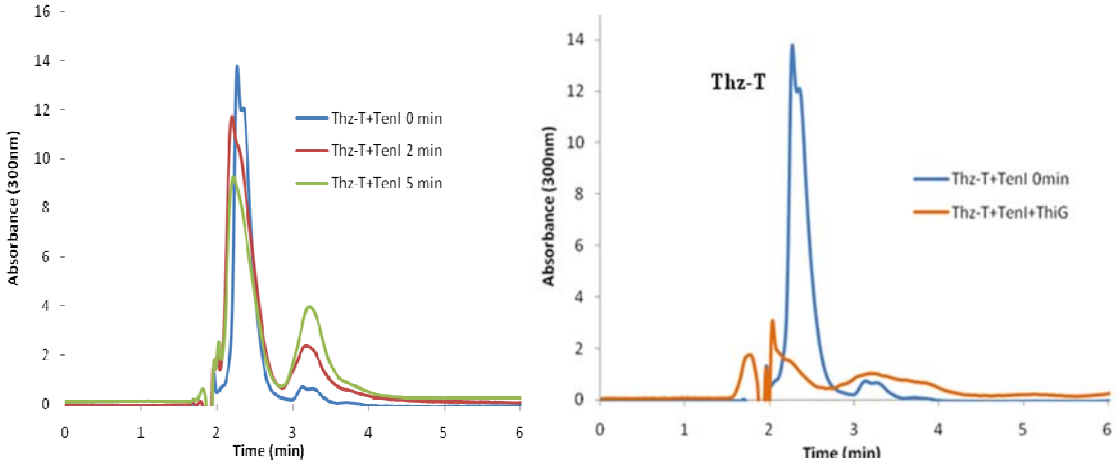


Figure 4.2: (a) The fate of Thz-T-P in the presence of TenI only (b) The fate of Thz-T-P in the presence of TenI and ThiG.

This reaction was performed multiple times varying the denaturing conditions by using 8M urea and quenching at shorter timepoints like 1 min, and each time, the thiazole tautomer peak disappeared without the formation of another product when ThiG was present.

This could either mean that ThiG was reacting with the Thz-T-P to produce a nonchromophoric species or that ThiG was binding irreversibly to the thiazole tautomer phosphate. In the case that the thiazole tautomer phosphate was irreversibly bound to ThiG, it would have to be an enzyme bound intermediate as it would have been released when the enzyme was heat denatured or treated with 8M urea. So the species formed by the reaction of ThiG with thiazole tautomer phosphate must be enzyme bound.

Thinking about the problem mechanistically, the thiazole tautomer is an unstable intermediate on the pathway of thiazole formation in *B. subtilis*. If ThiG was added back into the system, there is a possibility that the enzyme catalyzes the reversible reaction from thiazole tautomer phosphate to the previous intermediate **27,** a late intermediate on the well-studied thiazole biosynthesis pathway[2] (Figure 4.3). This can then further break down to release the glycine imine and form intermediate **28**, which will be bound to the protein. Interestingly, in all the bound intermediate structures, the phosphate group remains intact, and hence it is a functional group that can be used as a probe to test our hypothesis.

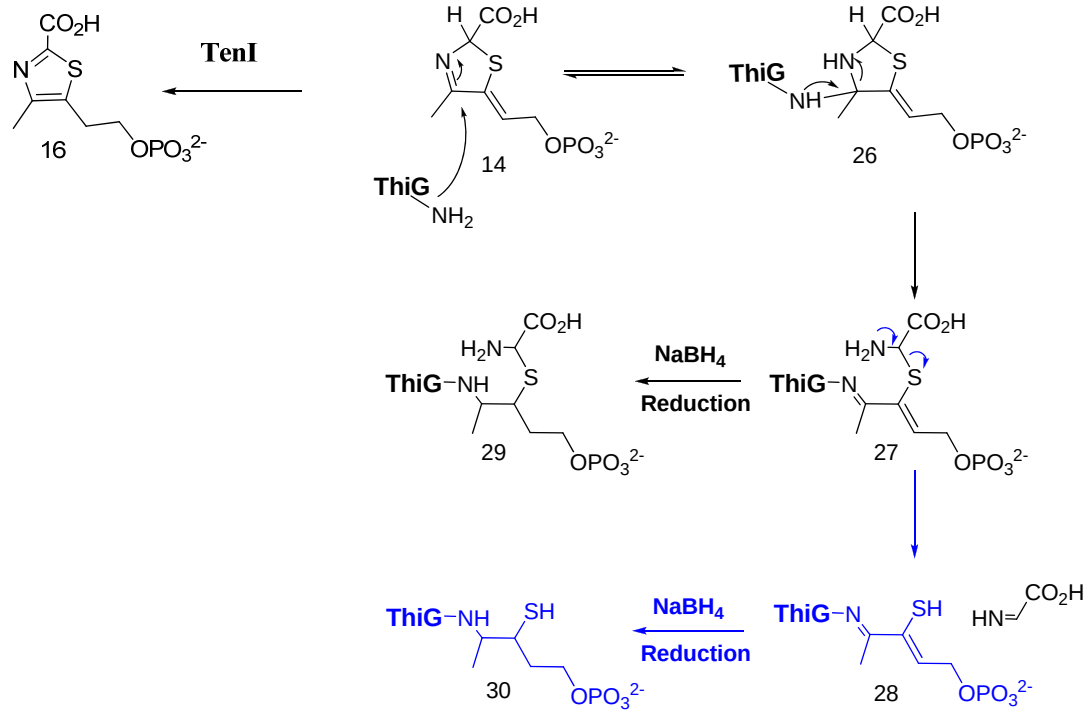


Figure 4.3: Putative reversibility of Thz-T-P with ThiG

The reversible imine bond in both the plausible intermediates would have to be reduced so as to stabilize it on the protein (species **29** and **30**) and then further analysis can be done. This hypothesis of ThiG-Thz-T-P reversibility was tested by analyzing for phosphates bound to the protein. This would be done initially by running a gel for the protein after reaction and staining it with a phosphate specific stain. If labeling of the intermediate bound ThiG protein was seen, that would indicate the presence of a phosphate somewhere on the protein. This protein could then be subjected to mass spectrometric analysis to reveal the increase in mass of the protein due to the bound adduct and then the site of phosphorylation.

His-tagged ThiSG protein was purified out by Ni-affinity chromatography. Thiazole tautomer phosphate was obtained as discussed previously and quantified by comparing the amount of AMP released by treatment with nucleotide pyrophosphatase with a known AMP standard (Figure 4.4).

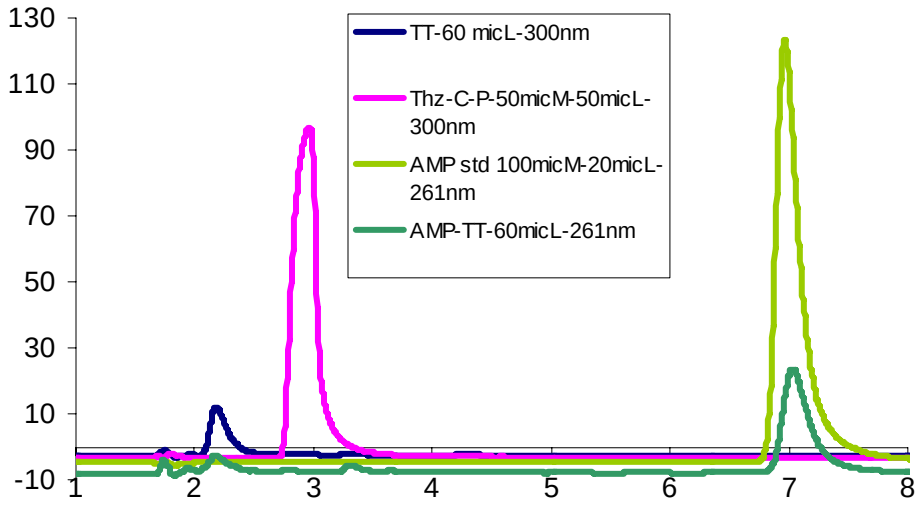


Figure 4.4: Quantitation of Thz-T-P by quantitating the AMP released on nucleotide pyrophosphatase treatment of **24**.

The Thz-T-P and ThiG were mixed in 1:1.2 ratio and kept at room temperature for 10min in 50mM Tris-HCl buffer, pH 7.4. A small aliquot of this was filtered off to analyze by HPLC to make sure that the Thz-T-P had disappeared. To the remaining reaction mixture was added 10mM $NaBH_4$ and this was allowed to react for 5 minutes. The $NaBH_4$ was then washed out from the ThiG protein by desalting the protein twice into 50mM NH_4OAc buffer, pH 7.4. Two other control reactions with no thiazole tautomer phosphate added and neither thiazole tautomer phosphate nor $NaBH_4$ added were also set up and washed similarly as the above mentioned sample. The filtrate of the main reaction as well as the control reactions were analyzed by UV-Vis absorbance on a RP-HPLC and the proteins were subjected to gel electrophoresis using a 12% Tris-Glycine gel. The proteins on the gel were then treated with phosphate-labeling stain, ProQ Diamond (Molecular Probes, Invitrogen) following the exact protocol for labeling phosphate labeled proteins and the gel was visualized by fluorescence scan (excitation 532nm, emission 580nm).

After the fluorescence scan, the gel was visualized by Coomassie stain and the two gels were compared for their protein content versus phosphate content by the two stains (Figure 4.5). It is quite clear that the phosphate stain in the lanes 3 and 4 are much darker as compared to the lanes 1 and 2. If one compares the corresponding Coomasie staining lanes, it can be clearly seen that the concentration of protein in lanes 3 and 4 are comparable to that in the lanes 1 and 2. So there is more phosphate bound to protein in lane 3 and 4 as compared to lanes 1 and 2. Another interesting point to note is that lane2, which has ThiG reduced with $NaBH_4$, has more phosphate staining as compared to lane A. This result has been seen consistently, and it indicates the presence of a phosphate bound species to the native protein ThiG, that gets fixed when treated with $NaBH_4$. This correlates very well with the finding that when the

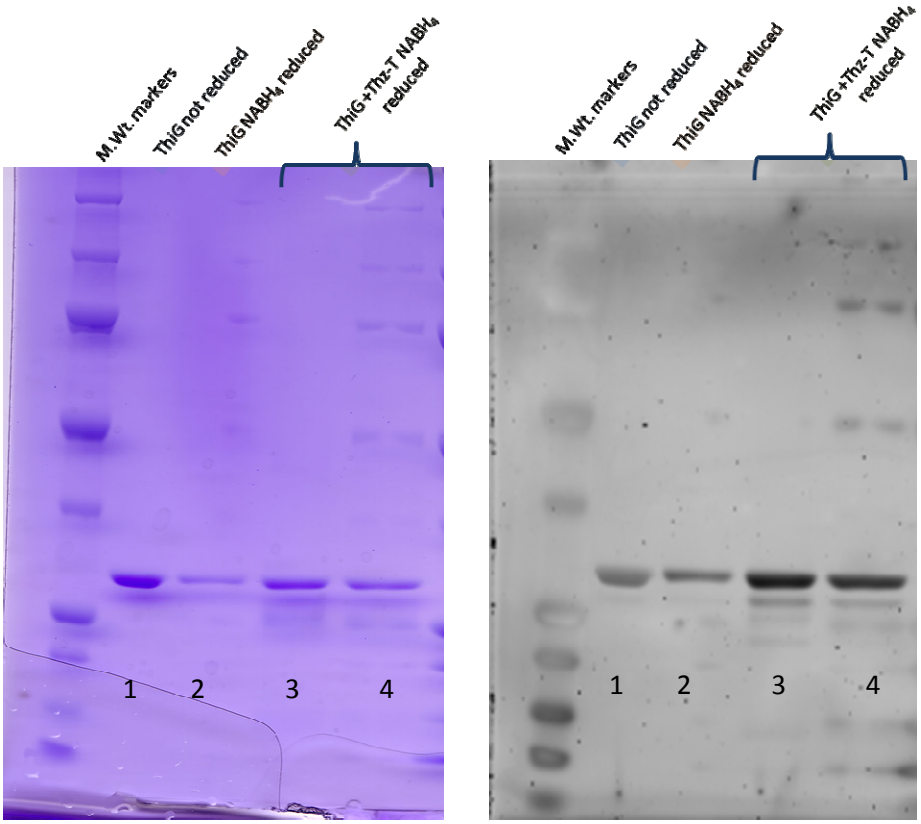


Figure 4.5: Coomassie stained and phosphostained gel of ThiG with Thz-T-P and other controls. Lane 1 in both gels indicates the pure ThiG protein without $NaBH_4$ reduction. Lane 2 indicates pure ThiG protein with nothing added, but reduced by $NaBH_4$. Lanes 3 and 4 are both the same full reaction sample where ThiG has been allowed to react with Thz-T-P 14 and then reduced with $NaBH_4$.

thiazole reconstitution reaction was done with all the reconstitution proteins and components without adding in DXP, the amount of thiazole formed, measured by thiochrome formation was greater than the basal levels of thiochrome observed in the remaining controls[3]. This also explains why ThiG in all the lanes gets labeled to some extent with the phosphostain, but is more in lane 2 and even more in lane 3 and 4. So there may be some advanced phosphorylated intermediate of DXP bound in the active site of ThiG, which causes the protein to be phosphostained, and forms thiazole and

consequently thiochrome on treatment with the other thiamin reconstitution components.

To verify this phosphostain result for ThiG, the ThiG mutant K96A which cannot form the imine bond with DXP[4] was used for a comparative study for phosphostaining along with the wtThiG protein. As we did not have a pure stock of the ThiG K96A mutant, ThiSG-K96A was used and in parallel, wtThiSG was used.

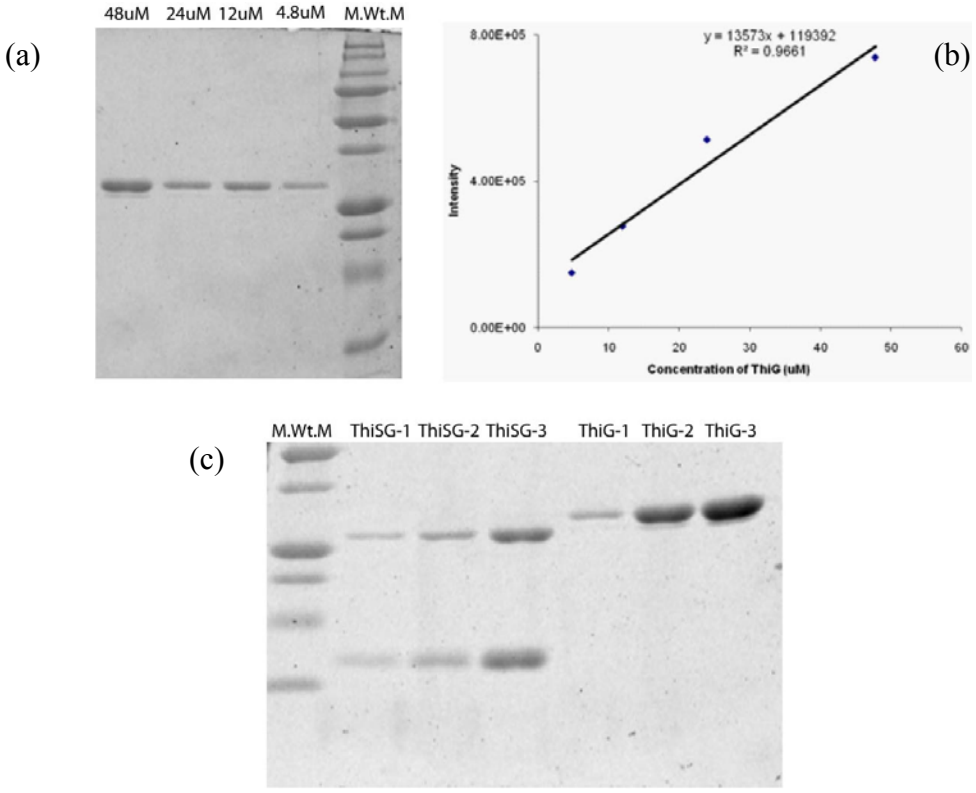


Figure 4.6: Fluorescence quantitation of the amount of ThiG in ThiSG using pure ThiG to make a standard curve by densitometry analysis. (a) Pure ThiG was run at various concentrations on a 12% Tris-glycine gel and the concentration of the protein in each lane was correlated by densitometry analysis to give the plot (b). In gel (c), the protein ThiSG was run at different concentrations along with the pure protein ThiG to correlate the amount of ThiG present in the ThiSG sample.

The amount of ThiG in the ThiSG was quantitated by densitometry analysis using a standard of pure ThiG and making a standard curve for the comparison (Figure 4.6).

Again, each of the wtThiG and the mutant K96A ThiG were set up in reactions with the thiazole tautomer phosphate as described previously, along with the controls a shown in Figure 4.7 for each.

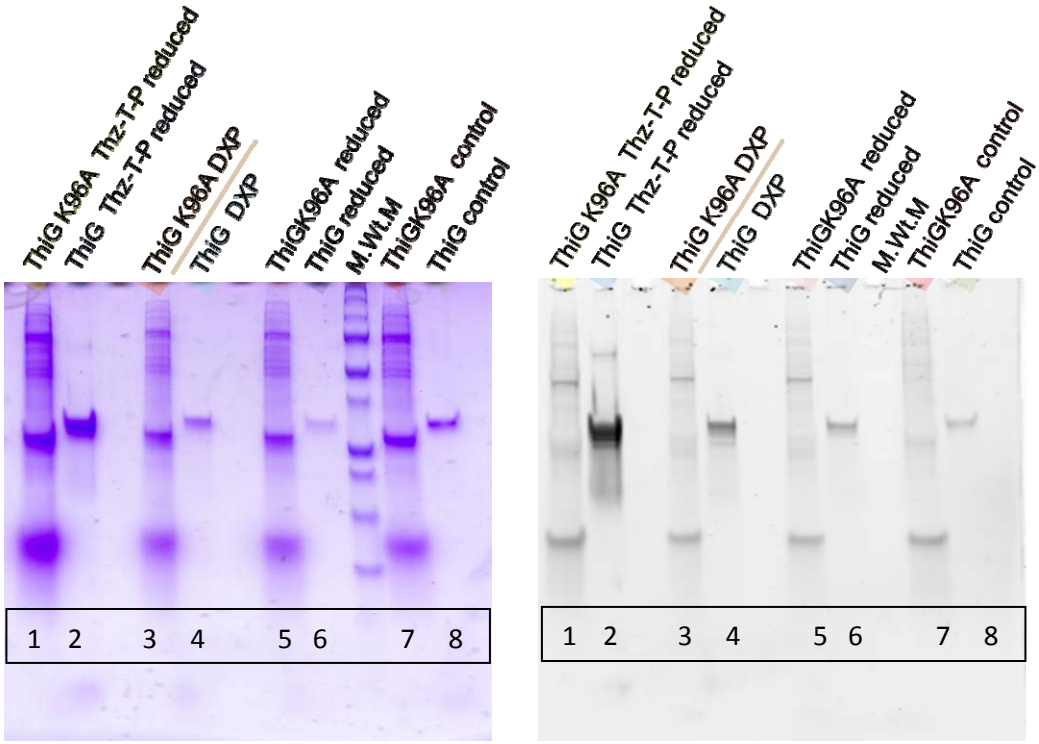


Figure 4.7: Coomassie stained and phosphostained gel of ThiG and ThiGK96A with Thz-T-P and other controls. The lanes on the gels from left to right are: 1,2-the ThiSGK96A and wtThiG respectively treated with Thz-T-P **14** and with $NaBH_4$ 3,4-the ThiSGK96A and wtThiG respectively treated with DXP without reduction 5,6 the ThiSGK96A and wtThiG respectively with reduction by $NaBH_4$ and 7,8 the ThiSGK96A and wtThiG respectively without reduction

As expected, the ThiG K96A mutant (lane 1) shows much less staining or non-specific staining than the wtThiG (lane 2), which clearly shows the presence of phosphate. This can be visualized clearly from the Coomassie stain that shows that the relative amounts of protein in each lane is comparable, even though a large difference can be seen in the phosphostained version of the gel (Figure 4.7)

Unfortunately, the phosphate stain is not very specific for phosphate residues only and non-specifically stained the ThiG protein in the control lanes even though the exact protocol of labeling was followed. As the manufacturer does not reveal the composition of the stain due to intellectual property issues, there is no way to even detect how this staining may be made more specific.

Preliminary top-down positive mode ESI-TOF MS analysis was performed on the pure ThiG samples treated with the thiazole tautomer and $NaBH_4$ comparing it to the control sample, where no thiazole tautomer was added to the reaction mixture. The expected molecular weight for ThiG is 29,440 Da. The expected molecular weight of the reduced species 29 is 29,710 Da and that of reduced species 30 is 29,636 Da (Figure 4.8).

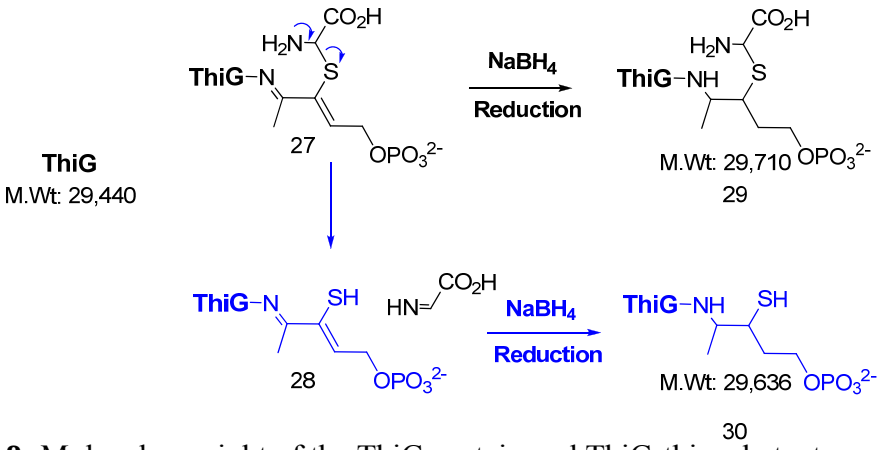


Figure 4.8: Molecular weight of the ThiG protein and ThiG-thiazole tautomer adduct 29 and 30 species expected to be observed by mass spectrometric analysis.

Both the reduced ThiG-thiazole tautomer sample and the control reduced ThiG sample sprayed very well on the ESI-TOF mass spectrometer and charged states could be clearly seen for both the samples (Figure 4.9 Inset a) and Inset b)). The deconvoluted mass spectrum for the reduced ThiG-thiazole tautomer (Figure 4.9 a)) showed a major peak at 29,636 Da corresponding to the adduct 30. The deconvoluted mass spectrum for the reduced control ThiG sample showed a peak at 29,440 Da (Figure 4.9 b)) and that for the reduced ThiG-thiazole tautomer shows a peak at 29,636 Da.

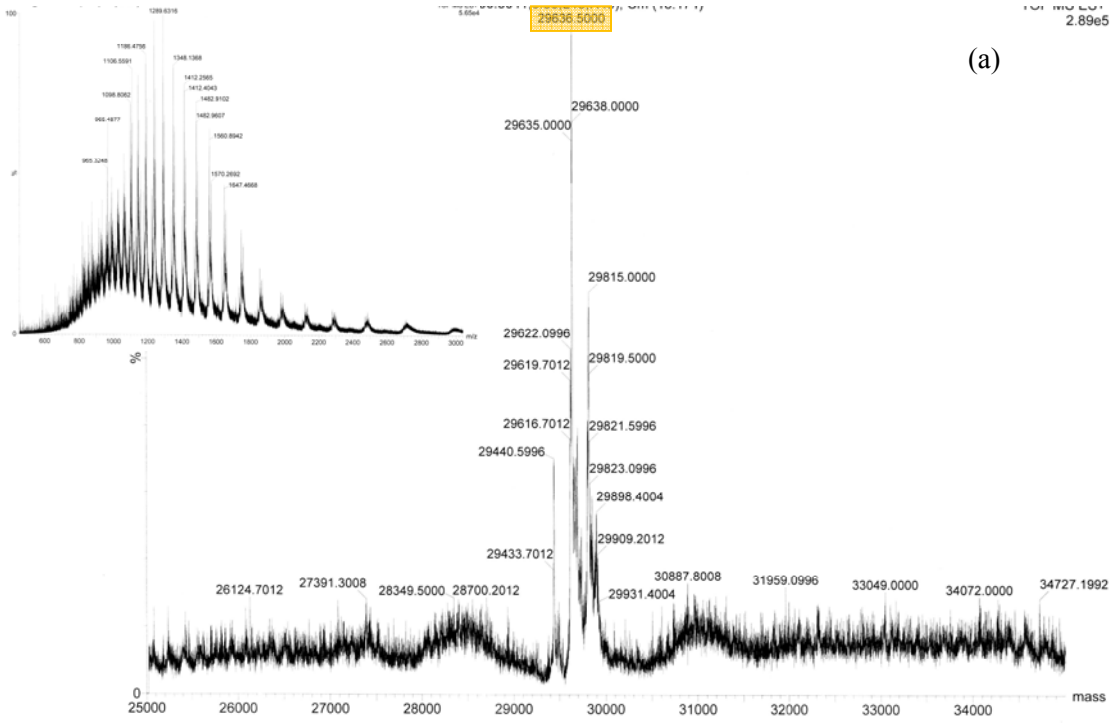
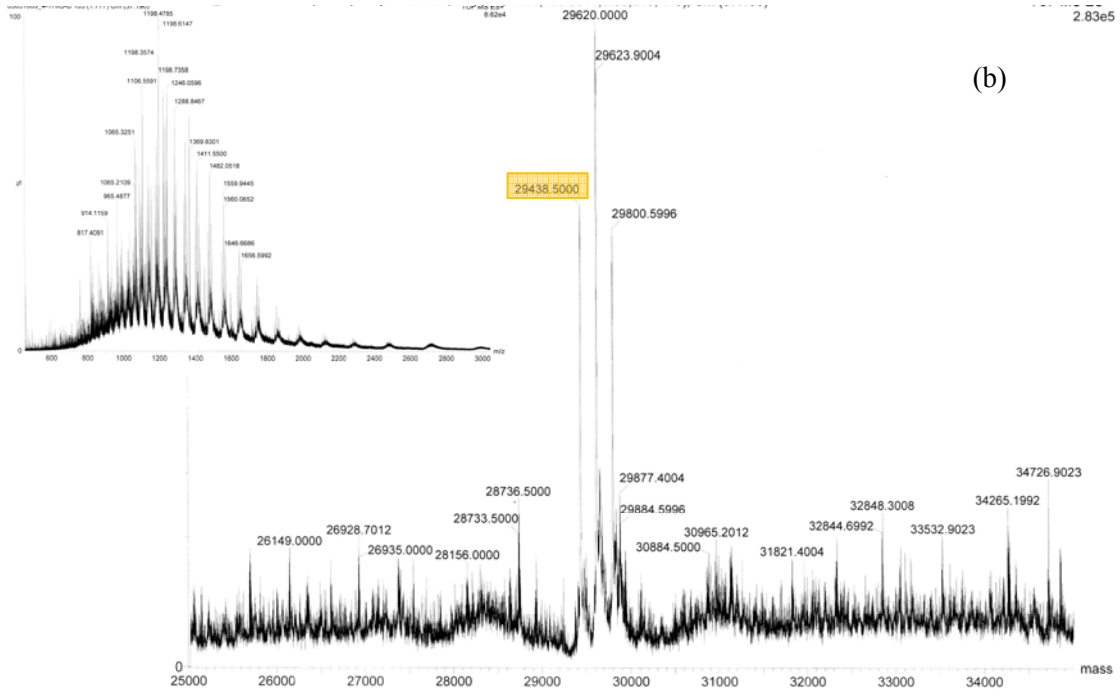


Figure 4.9: Positive mode ESI-TOF MS of the reduced thiazole tautomer treated ThiG reaction and the reduced conrol ThiG sample. (a) Deconvoluted mass spectrum of the reduced thiazole tautomer treated ThiG reaction. (Inset Figure a) Charged states for the reduced thiazole tautomer treated ThiG reaction. (b) Deconvoluted mass spectrum of the control reduced ThiG reaction. (Inset Figure b) Charged states for the control reduced ThiG reaction.

Figure 4.9 (Continued):



Other species are also visible on the mass spectrum and may be other interesting adducts of the protein ThiG. We will to investigate these species in further experiments.

4.3 Conclusions:

The product of ThiG **14** is a non-aromatized tautomeric form of thiazole carboxylate phosphate **16.** It has a half-life of 5 days and requires TenI to aromatize it to the thiazole carboxylate phosphate **16**. When 14 is added back to ThiG, it seems to disappear and gets attached irreversibly to the protein. The phosphate group on 14

remains bound on the protein, hence we could probe it using phosphate-staining reagent. Top down ESI-TOF MS has been done on the ThiG protein-intermediate bound samples to find out the mass of the modified protein. Further fragmentation analysis of the protein will reveal the amino acid that has been modified by the thiazole tautomer.

4.4 Experimental methods

4.4.1 Source of Chemicals:

All chemicals and snake venom nucleotide pyrophosphatase were purchased from Sigma-Aldrich Corporation (USA) unless otherwise mentioned. Calf intestinal phosphatase was obtained from New England Biolabs.. The microcon membrane filters were from Millipore. The Gel-fluorescence scanner Typhoon Trio imager was from GE healthcare life sciences (Piscataway, NJ). The Pro-Q® Diamond phosphoprotein gel stain was obtained from Invitrogen. Thz-T-P **14** was obtained as previously described in Chapter 3. Previously synthesized stock of [1-^{13}C]-DXP [14] was used for analysis with ThiG.

4.4.2 Overexpression and purification of enzymes for gel assays

ThiG, ThiSG and ThiSGK96A : *E. coli* BL21(DE3) containing the ThiG/ThiSG overexpression plasmid in pET16b was grown in LB medium containing ampicillin (40 μg/mL) with shaking at 37 °C until the OD_{600} reached 0.6. At this point, protein overexpression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration = 2 mM) and cell growth was continued at 15 °C for 16 h. The cells were harvested by centrifugation and the resulting cell pellets were stored at -80 °C. To purify the protein, the cell pellets from 1L of culture were resuspended in 25 mL

lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, pH 8) and lysed by sonication (Heat systems Ultrasonics model W-385 sonicator, 2 s cycle, 50% duty). The resulting cell lysate was clarified by centrifugation and the ThiSG protein was purified on Ni-NTA resin following the manufacturer's instructions. After elution, the protein was desalted using a 10-DG column (BioRad) pre-equilibrated with 50 mM Tris-HCl buffer, pH 7.8.

4.4.3 HPLC method for analysis of Thz-T-P and AMP

The following linear gradient at a flow rate of 3 mL/min: solvent A is water, solvent B is 100 mM KPi, pH 6.6, solvent C is methanol. 0 min: 100% B; 3 min: 10% A, 90%B; 17 min: 34% A, 60% B, 6% C; 21 min: 35% A, 25% B, 40% C; 23 min: 100%B

4.4.4 Gel Phosphoprotein staining

Fix the gel. Immerse the gel in ~100 mL of fix solution (prepared in step 1.1) and incubate at room temperature with gentle agitation for at least 30 minutes. Repeat the fixation step once more to ensure that all of the SDS is washed out of the gel. Gels can be left in the fix solution overnight.

Wash the gel. Incubate the gel in ~100 mL of ultrapure water with gentle agitation for 10 minutes. It is important that the gel be completely immersed in water in order to remove all of the methanol and acetic acid from the gel. Residual methanol or acetic acid will interfere with Pro-Q® Diamond phosphoprotein staining. Repeat this step twice, for a total of three washes.

Stain the gel. Incubate the gel in a volume of Pro-Q® Diamond phosphoprotein gel stain equivalent to 10 times the volume of the gel (e.g., 60 mL for Novex® precast minigels), with gentle agitation in the dark for 60–90 minutes. If directly comparing

multiple gels, it is important that the incubation time be the same for each gel. DO NOT stain overnight, as this will result in higher background staining.

Destain the gel. Destaining is important to reduce the gel background signal and to reduce the signal from nonspecific staining. Incubate the gel in 80–100 mL of destain solution (see step 1.2) with gentle agitation for 30 minutes at room temperature, protected from light. Repeat this procedure two more times. The optimal total destaining time is about 1.5 hours.

Wash the gel. Wash twice with ultrapure water at room temperature for 5 minutes per wash. If the background is high or irregular, the gel may be left in the second wash for 20–30 minutes and re-imaged.

Imaging the gel. Typhoon Trio (GE Healthcare) excitation 532nm, emission 580nm

REFERENCES

1. Hazra. A.; Chatterjee, A.; Begley, T.P. *J.Am.Chem.Soc.* **2009** 131, 3225-9.
2. Chatterjee, A.; Han, X., McLafferty, F.W.; Begley, T.P. *Angew. Chem. Int. Ed.* **2006** 45, 3507 –3510
3. Park, J.-H.; Dorrestein, P.C.; Zhai, H.; Kinsland,C.; McLafferty F.W.; Begley, T.P. *Biochemistry* **2003**, 42, 12430-12438
4. Dorrestein PC, Huili Zhai H, Taylor SV, McLafferty FW, Begley TP. *J Am Chem Soc.* **2004** 126, 3091-6

CHAPTER 5

The remarkable rearrangement reaction catalyzed by 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase: tracking the fate of C's and H's of the substrate AIR

5.1 Introduction:

The biosynthesis of thiamin pyrophosphate involves separate syntheses of the thiazole and the hydroxymethyl pyrimidine (HMP) precursors, which are then linked together and pyrophosphorylated to form the cofactor. Mechanistic details of thiamin-thiazole biosynthesis are well understood in prokaryotes (detailed in Chapter 2 and 3) and analysis of the eukaryotic system is well underway. In prokaryotes, 1-deoxy-D-xylulose-5-phosphate, cysteine and glycine or tyrosine are utilized by five proteins to construct the thiazole moiety, whereas in *S. cerevisiae*, just one gene product converts NAD and glycine to thiazole, obtaining sulfur from a source yet unknown[1-7]. In comparison, the mechanistic understanding of thiamin-pyrimidine biosynthesis in both prokaryotes and eukaryotes is still at its infancy. In yeast, a single gene product THI5p is implicated in HMP biosynthesis from PLP and histidine (Figure 5.1 b)[8-9], however attempts at reconstituting its acitivity in a purified system remain largely unsuccessful. In bacteria and some plants, the HMP-P synthase ThiC converts aminoimidazole ribonucleotide (AIR **1**), an intermediate in the purine nucleotide biosynthesis pathway, to hydroxylmethyl pyrimidine phosphate (HMP-P **2**)[10] (Figure 5.1 a). The thiC gene is found to cluster with thiamin biosynthetic genes in many prokaryotes and complements all HMP requiring mutants in *E.coli*, *S.typhimurium and B.subtilis*[11-12].

A careful analysis of the primary structure of ThiC across various species of bacteria and plants revealed the presence of a $C_{561}SMC_{564}GPKFC_{569}$ motif near the N-terminal of the protein, indicating the likely presence of an Fe-S cluster (5.1 c). Previous studies had shown that in the C561A, C564A and C569A mutant proteins in *Salmonella enterica* were unable to biosynthesize the thiamin pyrimidine, suggesting that HMP-P synthase contained an Fe-S cluster[13]. Also, when the Fe-S cluster biosynthesis genes were disrupted in *Salmonella enterica*, the organism became thiamine requiring[12]. *In vivo* and clarified lysate studies in the past using labeled AIR lead to the localization of atoms from the substrate that were incorporated into the product HMP-P. The results pointed towards a very unusual rearrangement; however the cofactor requirements and the fate of all atoms of AIR could not been elucidated owing to the complexities of dealing with *in vivo* systems or clarified lysates[14].

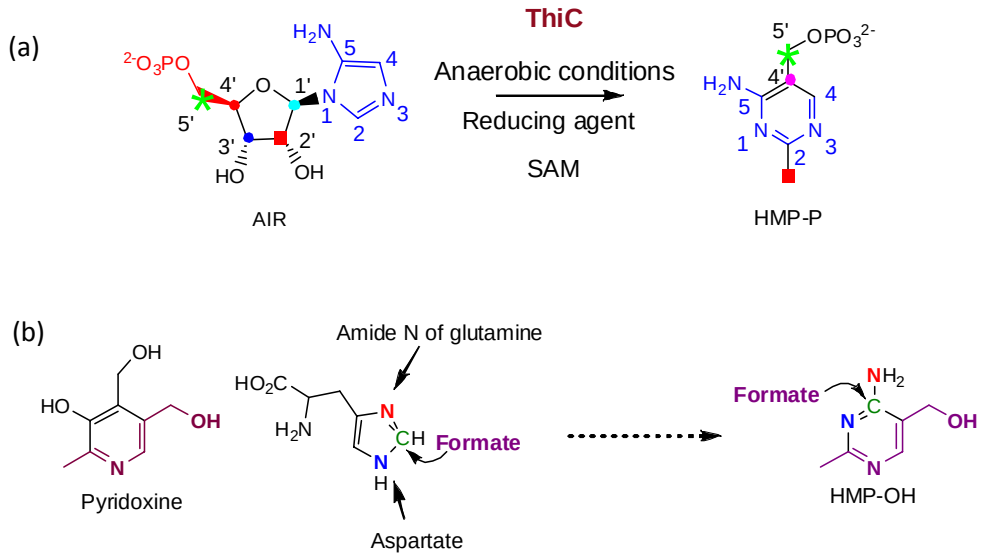
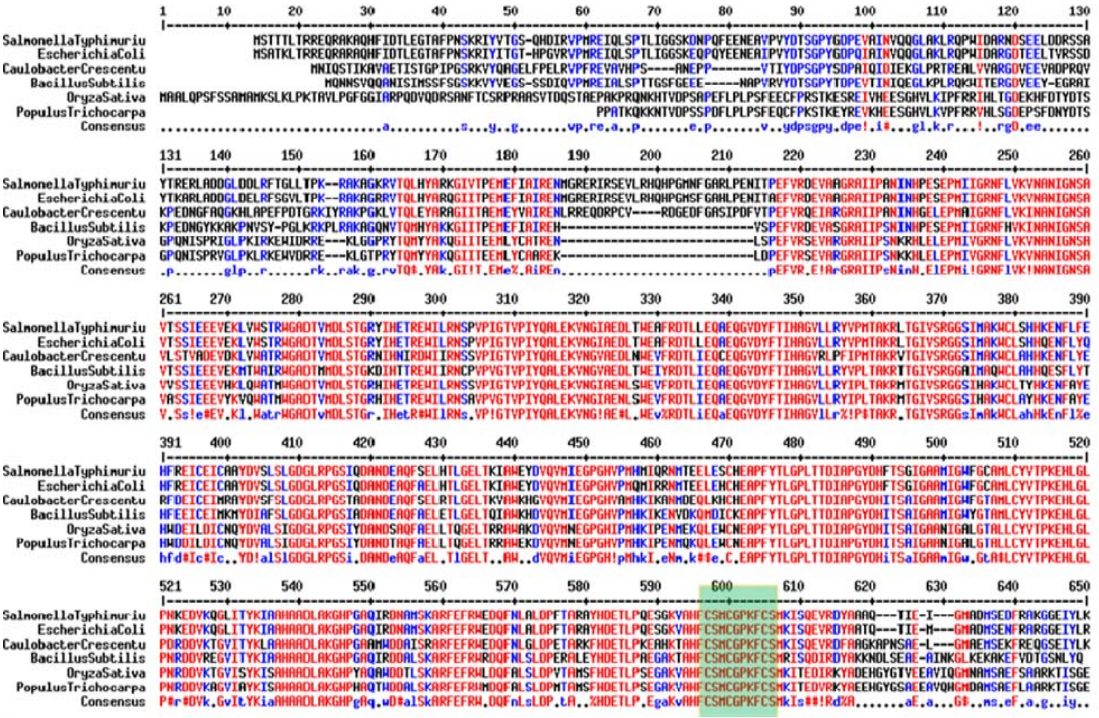


Figure 5.1: HMP biosynthesis in (a) *E. coli* and (b) *S. cerevisiae.* (c) Primary sequence alignment for ThiC protein of various origins shows a conserved Fe-S cluster motif C-X-X-C-X-X-X-X-C near the N-terminal of the protein

Figure 5.1 (Continued):

(c)



Recently, active wtThiC enzyme was obtained by cloning the thiC gene from *C. cresentus* into a pET16b plasmid, cotransforming this plasmid into B834(DE3) along with the pDB1282 (Fe-S cluster chaperone proteins, gift from Dennis Dean) and overexpressing the protein in minimal media containing Fe and S.

This protein was His-tagged and was purified under anaerobic conditions using Ni-NTA chromatography. The activity of ThiC was reconstituted in a defined biochemical system and *in vitro* studies on the reaction catalyzed by ThiC using labeled AIR has revealed the involvement of a rearrangement reaction of unprecedented complexity (Figure 5.2)[15].

The protein has an absorbance typical of Fe-S cluster proteins and was also found to be incredibly sensitive to oxygen.

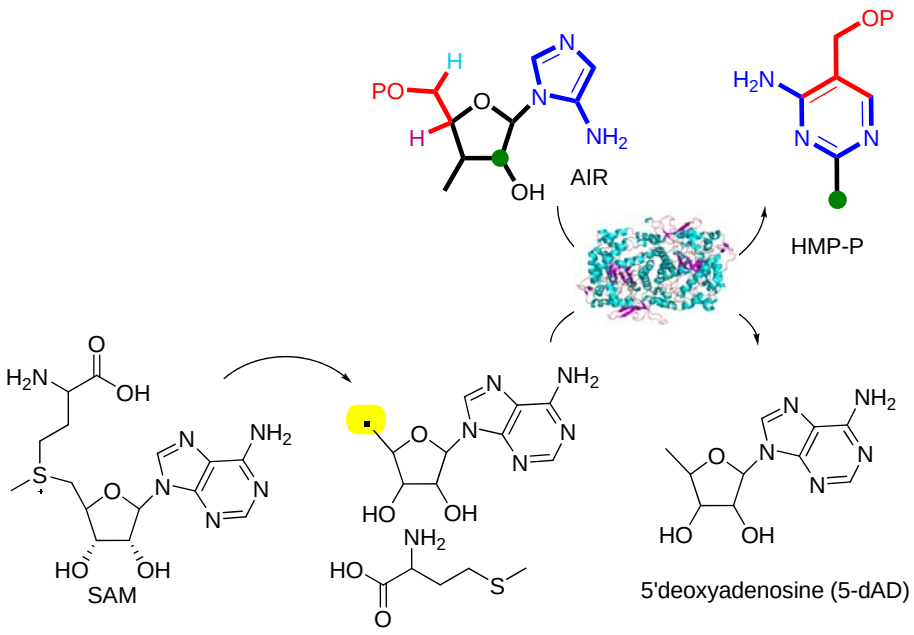


Figure 5.2: The complex rearrangement catalyzed by ThiC

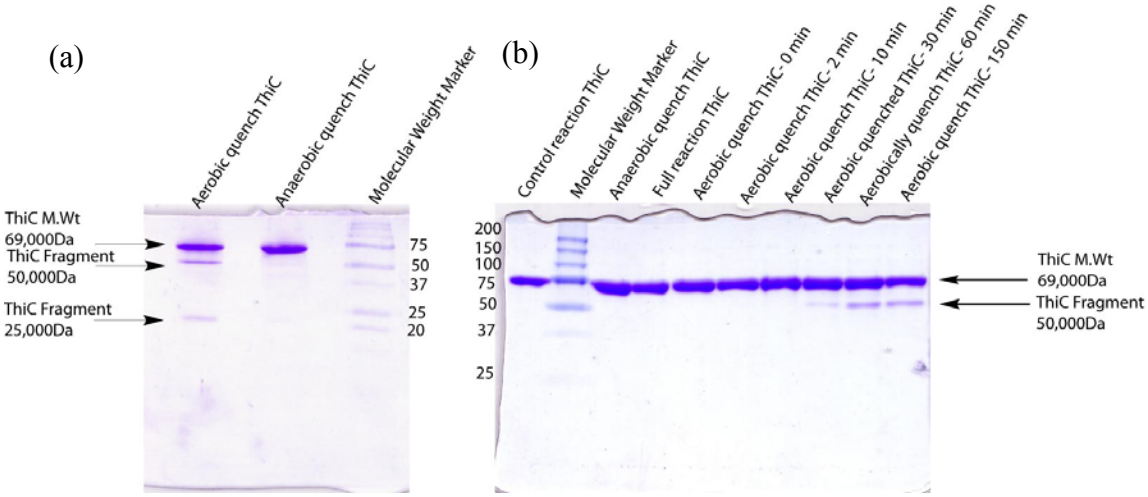
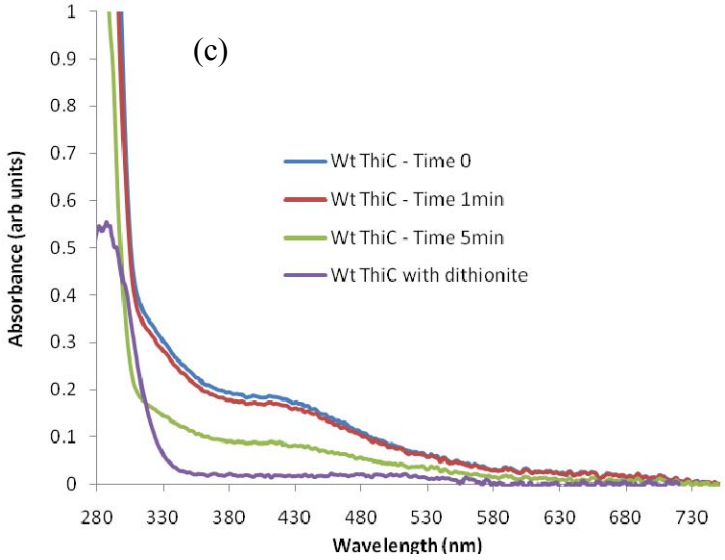


Figure 5.3: (a) SDS-PAGE analysis of ThiC protein that was purified anaerobically and quenched anaerobically or aerobically. A new band appears close to the anaerobic wtThiC when it is exposed to air (b) Appearance of ThiC-cleavage product on exposure to air (c) UV-Vis spectral traces of ThiC exposed to air over time. ThiC seems to change the Fe-S cluster state on exposure to air.

Figure 5.3 (Continued):



Purification in aerobic conditions lead to fragmentation (Figure 5.3 a) and inactivation of the protein (Figure 5.3 b). Spectroscopic, structural and biochemical studies established ThiC as a unique member of the 4Fe-4S cluster dependent radical-SAM superfamily of enzymes[14-16].

5.2 Results/ Discussion:

5.2.1 Determining the fate of the C1 atom of AIR:

Prior to the successful reconstitution of the ThiC catalyzed reaction[15], labeling studies could only elucidate the fates of the atoms of AIR that are incorporated into HMP-P (Figure 5.2)[14]. The low efficiency of conversion coupled with the complexity of the cell free extract, in which such assays were performed, precluded the fate determination of the substrate atoms that did not get incorporated in the product (C1 and C3 of the ribose moiety of AIR). With the defined ThiC reconstitution system, it was now possible to ascertain the reaction products originating from the C1 and C3 atoms of AIR. Also, enzymes belonging to the radical-SAM superfamily of proteins

initiate catalysis by generating the reactive 5-dAd radical (**4**), via the reductive cleavage of SAM **3** (Figure 5.2)[18-20]. Demonstration of ThiC as a radical-SAM protein raised the question regarding the precise mechanistic role of the 5-dAd radical to catalyze this remarkable rearrangement reaction.

To evaluate the fates of C1 and C3, corresponding singly ^{13}C-labeled AIR molecules were synthesized. [2]Using these as substrates, reactions were set up with ThiC, SAM

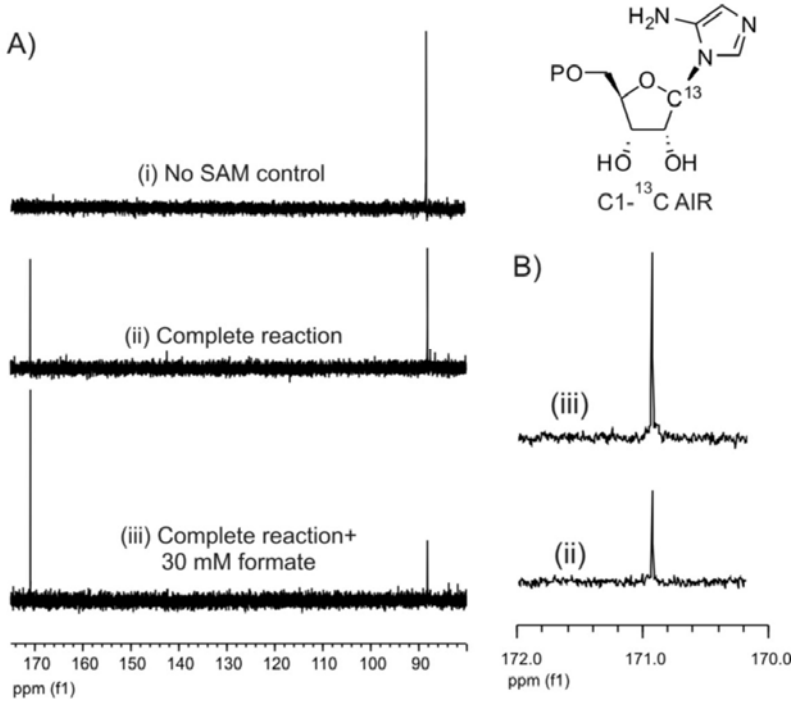


Figure 5.4: Studies with C1-^{13}C AIR. A) NMR analysis of the reaction mixture using C1-13C AIR as the substrate of ThiC reveals: ^{13}C signal of C1 of AIR (88 ppm), intact in the control reaction with no SAM (i); a new signal at 171 ppm along with the C1 of unreacted AIR (ii). Addition of 30 mM sodium formate to the reaction mixture enhances the intensity of the new peak at 171 ppm (iii). B) An expanded view of the NMR spectra around the formate peak for (ii) and (iii).

*, ** All work appearing in between * and ** has been performed by Abhishek Chatterjee

and dithionite. Upon the removal of the protein, the reaction mixtures were subjected to ^{13}C NMR analysis. A control reaction, where the cofactor SAM was omitted, was also analyzed for comparison. HPLC analysis showed approximately 45% conversion of AIR to HMP-P in the first reaction, while in the control it remained intact. The ^{13}C NMR analysis of the reaction mixture for the C1-^{13}C-AIR (Figure 5.4 a and b) revealed the generation of a large singlet carbon signal at 170 ppm, which was absent in the control. Doping the sample with sodium formate resulted in the appropriate enhancement in the intensity of the new signal, which suggests that C1 of AIR is converted to a formic acid**.

5.2.2 Determining the fate of the C3 atom of AIR:

Interestingly, no new signal was observed when similar experiments were performed with C3-AIR, even though significant conversion of AIR to HMP-P (40-50%) was confirmed by HPLC analysis. The reaction was also analyzed by ^{13}C NMR without removing the protein as previously described[32], in case a protein bound species is generated but evidence for no such species was found. Attempts were then made to detect formaldehyde in the reconstitution mixture, should this be the reaction product of C3, either by directly detecting the ^{13}C-NMR or using various trapping agents like hydroxylamine, dimedone[22-23] and Purpald[24].None of the trapping agents showed any significant changes in between the full reaction and the control reactions of the ThiC reaction mixture, where dithionite, AIR or SAM had been omitted. Thus, generation of formaldehyde during ThiC reaction was not detected. The other possibility was that C3 could leave as carbon dioxide (CO_2). However, CO_2 is highly soluble in water, and so we should have seen it in the NMR experiment as a bicarbonate signal. Subsequently, an anaerobic carbon monoxide (CO) trapping assay[25], which involves a specific change in the soret-region absorbance of hemoglobin (Figure 5.5 a, b) as a

result of its association with CO to form carboxyhemoglobin, was adapted to observe production of CO in the ThiC reconstitution assay. Interestingly, the generation of CO during the ThiC catalyzed conversion of AIR to HMP-P was clearly observed (Figure 5.5 c) and the change in the signal was proportional to the concentration of AIR used in the reaction mixture. Since the fates of all the other carbon atoms of the substrate have conclusively been ascertained previously by ^{13}C NMR and CO is sparingly soluble in aqueous solutions, it is reasonable to conclude that the C3 of AIR is converted to CO.

Many other methods to trap CO for direct detection were subsequently tried. Most conventional methods to trap CO require large CO pressure in the presence of a metal catalyst or non-aqueous solvents for facile reaction. Neither of these two conditions was achievable in the experiment we were performing. Hence, we set up the ThiC reconstitution reaction using ^{13}C-3C-AIR and bubbled out the CO gas generated in the ThiC reconstitution reaction using Argon, and then trapping it by various metal catalyzed reactions using tryptamine, benzylamine and iodobenzene[26-29]. We used ^{13}C3-AIR in the reconstitution so we could analyze the products by mass spectroscopy or NMR, however we could not conclusively form any of the CO-trapped compounds. An attempt to trap the labeled CO gas by bubbling it into hemoglobin and doing a ^{13}C-NMR on the protein-CO complex was made[30] but no significant signal for CO bound to the hemoglobin could be seen. The failure of these strategies can possibly be attributed to the fact that the amount of CO being generated is at best low micromolar quantities and the trapping efficiency of these methods[29] is very low (~10%). In the process of bubbling it out of an enzymatic reaction vial, incomplete trapping as well as transfer may occur.

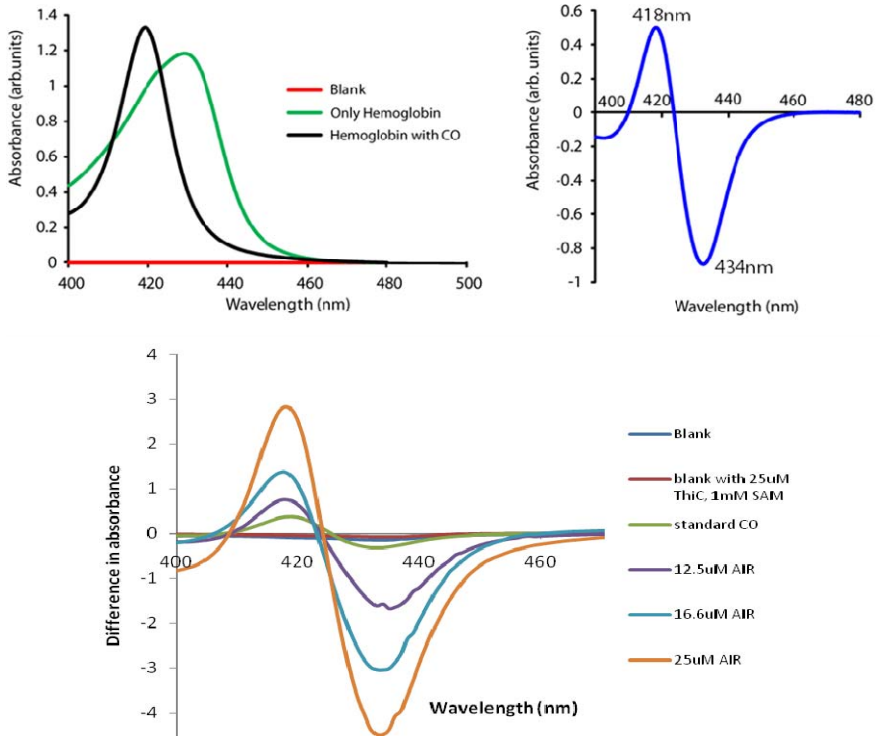


Figure 5.5: (a) Visible range absorbance of reduced anaerobic hemoglobin (green) and hemoglobin treated with CO saturated buffer (black). (b) Difference spectrum of the two hemoglobin traces shown in (a) (c) Change in the soret region spectrum of hemoglobin on addition of different amounts of AIR into the ThiC reaction mixture, subsequently releasing different quantities of CO.

Additionally, trapping of CO is not quantitative in any of the methods, so the final amount of CO that is trapped is not within detectable limits. Hence, none of these trapping strategies worked successfully.

5.2.3 Determining the chemistry of the 5'deoxyadenosyl radical in the rearrangement reaction

Having established the fates of the C1 and C3 atoms of AIR, we went on to investigate the role of the 5-dAd radical in the reaction. As a member of the radical-SAM

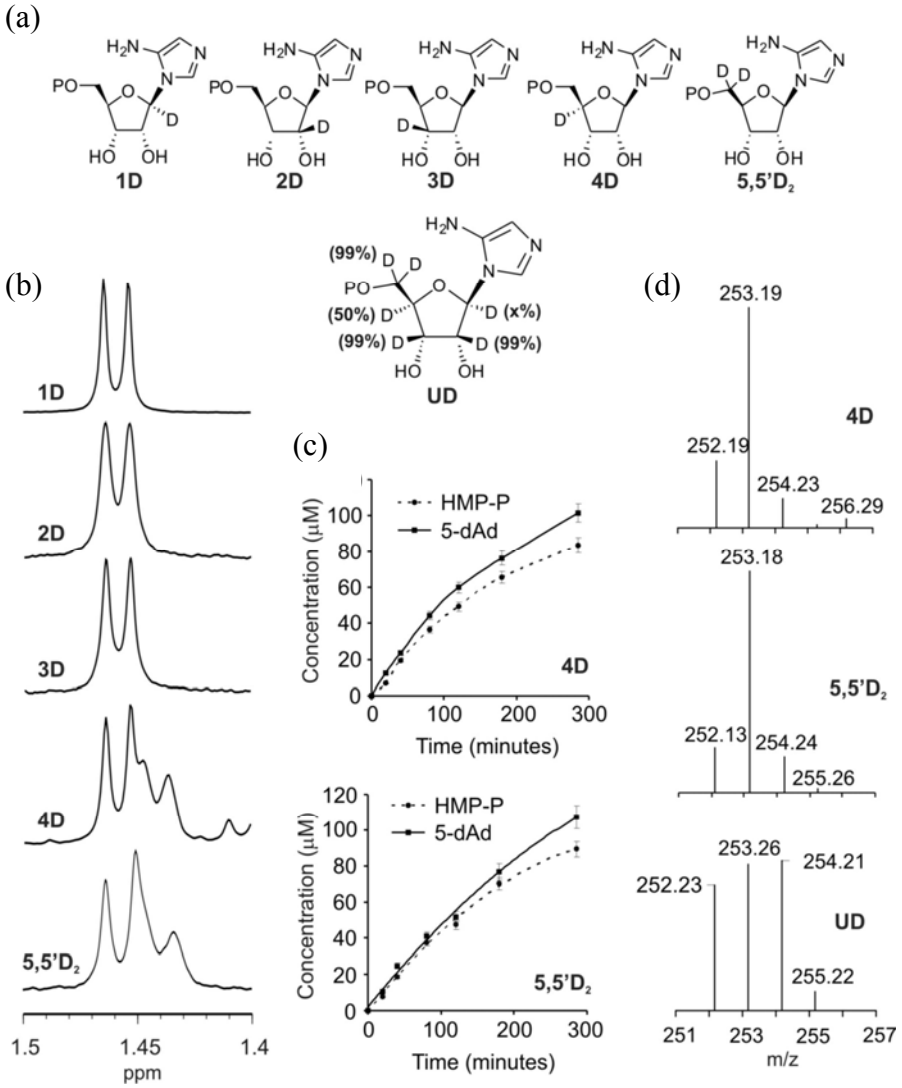


Figure 5.6: Studies with deuterated AIR.) Site-specifically deuterium labeled AIR molecules synthesized for this study. b) ^{1}H NMR signal for the 5′-methyl group 5-dAD isolated from ThiC reactions with labeled substrates. Incorporation of D in the methyl group for 4D and 5,5'D_2 AIR is noted by the presence of the additional, upfield shifted broader doublet. c) Rate of formation of 5-dAd and HMP-P measured over 5 hours reveal a 1:1 product ratio. d) MS analysis of 5-dAd produced in the early phase of the reaction reveal predominant mono-deuteration (m/z=253) with 4D and 5D and 50% di-deuteration (254 Da) with UD.

superfamily, the generation of the 5-dAd radical from SAM plays an intimate role in initiating catalysis. The radical may directly react with the substrate or, alternatively, abstract a hydrogen atom from the protein to generate a protein-bound radical, which in turn reacts with the substrate. When it directly reacts with the substrate, the 5-dAd radical may be used as a co-substrate or as a catalyst, in which case it is regenerated at the end of the reaction. Site-specifically D-labeled AIRs molecules were synthesized (by Sameh Abdelwahed and David Hilmey, Begley lab) and further enzymatically phosphorylated to produce labeled AIR (Figure 5.6 a).

*The reconstitution reaction with ThiC was performed with the five different AIR substrates, 5-dAd was HPLC purified and was then analyzed by ^{1}H-NMR. Incorporation of a deuterium atom at the 5′ position of 5-dAd has been shown to result in a small upfield shift of the 5′-H NMR signal[21]. Thus, ^{1}H NMR analysis of the isolated 5-dAd was initially used to identify possible deuterium incorporation at the 5′-position of 5-dAd**. Flavodoxin/flavodoxin reductase and NADPH was used to reduce the Fe/S cluster of ThiC, since the use of dithionite as the reducing agent was associated with high levels of uncoupled 5-dAd production. Surprisingly, we observed robust deuterium incorporation, when either of 5′-D_2 and 4′-D-AIR was used as the substrate, while no D incorporation was associated with 1′, 2′ or 3′- labeled substrates (Figure 5.6 b).

To understand whether both of these deuterium incorporation reactions are mechanistically relevant, we determined the product ratio (HMP-P:5-dAd) of the ThiC reaction with the flavodoxin/flavodoxin reductase and NADPH over a period of 5 hours using both D labeled substrates. A product ratio of nearly 1:1 was observed throughout the course of the reactions (Figure 5.6 c). This experiment was repeated

*,** All work appearing in between * and ** has been performed by Abhishek Chatterjee

using methyl viologen as a reducing agent[33] rather than flavodoxin/flavodoxin reductase and NADPH, which forms quantitative amounts of 5'deoxyadenosine as compared to HMP-P and again a product ratio of nearly 1:1 was observed throughout the reaction.

This product ratio is consistent with the use of 5-dAd radical as a co-substrate rather than a catalyst. Simultaneous inline ESI-MS analysis of the product 5-dAd revealed mono-deuterated 5-dAd as the predominant product in both cases (Figure 5.6 d). These results suggest that the transfer of the deuterium label from AIR to 5-dAd, observed for either of the deuterated substrates, occur as a direct consequence of the ThiC catalyzed conversion of AIR to HMP-P and not via an uncoupled reduction of SAM.

*To further demonstrate the transfer of the deuterium labels from two different positions of AIR to the same 5-dAd, we needed a substrate which is deuterium labeled both at 4 and 5 positions. Since the appropriately D-labeled ribose, the starting material for the synthesis of AIR, was commercially unavailable, we attempted its synthesis using catalytic H/D exchange of ribose. The resulting ribose molecule was mostly deuterated (>98%) [31] at 2, 3 and 5 positions and only partially deuterated at positions 4 (50%) and 1 (<1%). AIR was synthesized using this preparation of ribose as the starting material and the 5-dAd produced in a ThiC reaction, where it was used as a substrate, was analyzed by HPLC coupled ESI-MS analysis. As expected from the abundance of the deuterium labels in the substrate, a 1:1 distribution of mono:bis-deuterated 5-dAd (m/z=253 and 254 respectively) was observed (Figure 5.6 d)**. These results confirm that one hydrogen atom each from the 4 and 5 positions of AIR is incorporated in the 5′-position of 5-dAd in the course of the ThiC catalyzed

*, ** All work appearing in between * and ** has been performed by Abhishek Chatterjee

reaction. The initial 5′-dAd radical must be regenerated after it abstracts the first hydrogen atom from the substrate, and catalyze a second abstraction event, to explain this observation. Such a novel reaction pattern involving back and forth hydrogen atom abstraction and the use of 5-dAd **5** as a "radical shuttle" is unprecedented in the literature.

5.2.4 Stereochemistry of H-abstraction by 5'deoxyadenosyl radical at the 5'5''H_2-AIR:

Another interesting observation was that only one H was being abstracted from the 5' ribose position of AIR by the 5'deoxyadenosyl radical (Figure 5.2). It would be interesting to know which H i.e. whether the pro-R or the pro-S H at the 5' position (Figure 5.7 a) of was being abstracted with regard to the position of the AIR in the

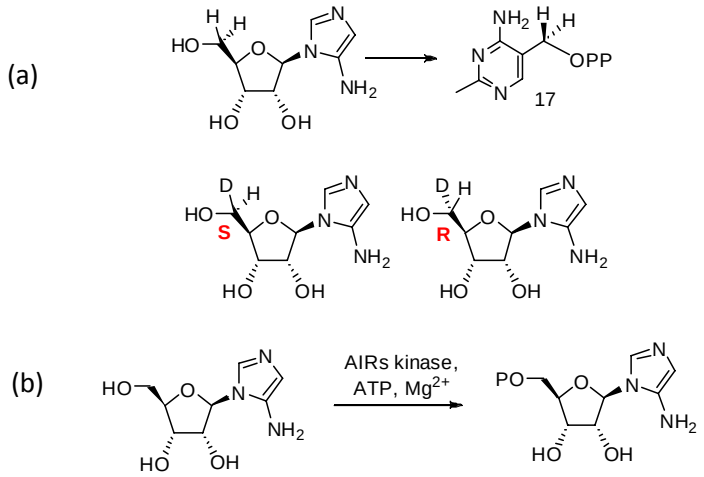


Figure 5.7: (a) One H atom from 5,5'-AIR is abstracted in the rearrangement (b) [5'S-^{2}H]AIRs and [5'R-^{2}H]AIRs were synthesized and further enzymatically phosphorylated to form the respective AIR.

active site of the crystal structure. Also, it would enable us to think about the position of the SAM with respect to the AIR and the residues involved in the catalytic mechanism of the reaction. To answer these questions, the [5'R-^{2}H]AIR and [5'S-^{2}H]AIR (Figure 5.7 b) were synthesized (by Sameh Abdelwahed, Begley lab).

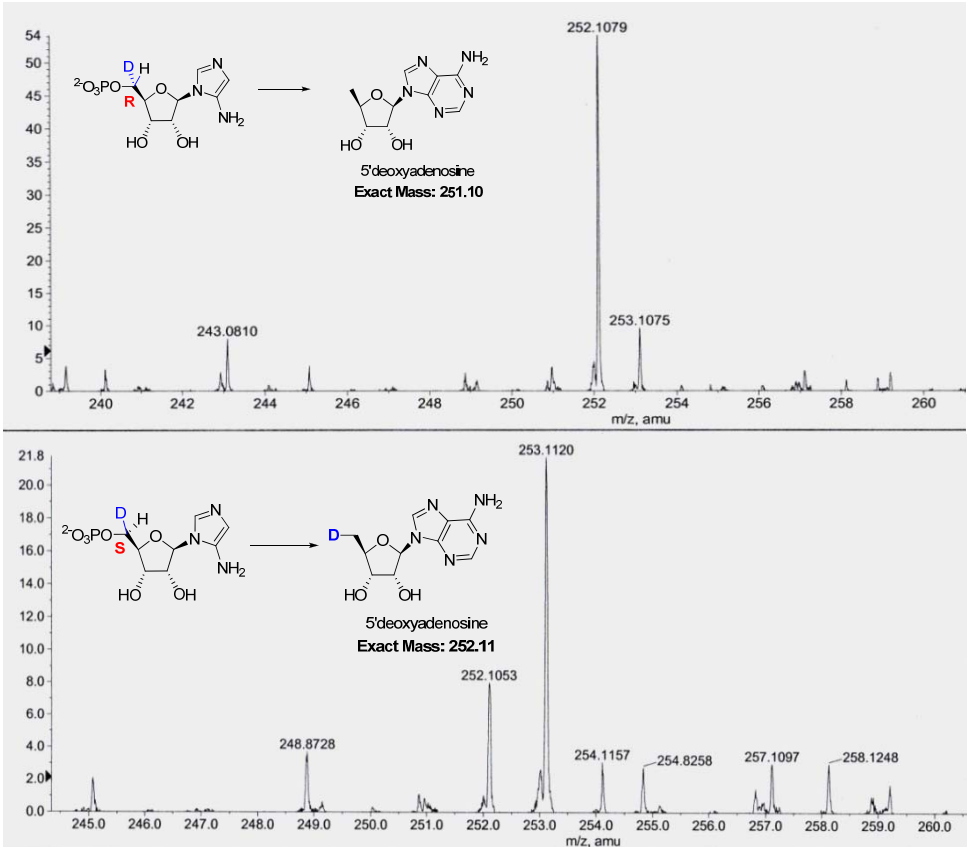


Figure 5.8: Positive mode ESI-MS of 5'deoxyadenosine from ThiC reaction with [5'S-^{2}H]AIR and [5'R-^{2}H]AIR.

Using these as substrates, reactions were set up with ThiC, SAM and methyl viologen. Upon the removal of the protein, the 5'deoxyadenosine in each reaction was purified by reverse-phase HPLC and analyzed by positive mode ESI-MS. A control reaction,

where the methyl viologen was omitted, was also analyzed for comparison and the 5'deoxyadenosine region was collected and analyzed by positive mode ESI-MS just to make sure no uncoupled product was being formed. The ESI-MS positive mode data clearly showed the incorporation of a D from the [5'S-^{2}H]AIR to form the monodeuterated 5'deoxyadenosine (m/z = 253) and incorporation of the H from the [5'R-^{2}H]AIR to form the unlabeled 5'deoxyadenosine (m/z = 252) (Figure 5.8). Analysis of the corresponding HMP-P showed the presence of a deuterium derived from the 5'R-^{2}H-AIR and unlabeled HMP-P from the 5'S-^{2}H-AIR.

5.2.5 Which H is abstracted from AIR first – the 4'H-AIR or the 5'H-AIR?

The fact that two H's of the AIR ring are found on the 5'deoxyadenine indicates the involvement of a "back-and-forth-radical shuttle" in the ThiC mechanism by the 5'deoxyadenosyl radical (Figure 5.9). This is an unprecedented reaction of the deoxyadenosyl radical and it may happen in the following manner - one H had to be abstracted from the AIR, followed by a rearrangement reaction within the radical-intermediate. After the rearrangement, the radical will be transferred, either directly or mediated by the protein to the 5'deoxyadenosine to regenerate the 5'deoxyadenosyl radical. This 5'deoxyadenosyl radical then abstracts an H again to make a second radical-intermediate, which completes the rearrangement to form the HMP-P and subsequently the 5' deoxyadenosine which contains two of the AIR H's - the [4'-H]AIR and the [5'-H]AIR.

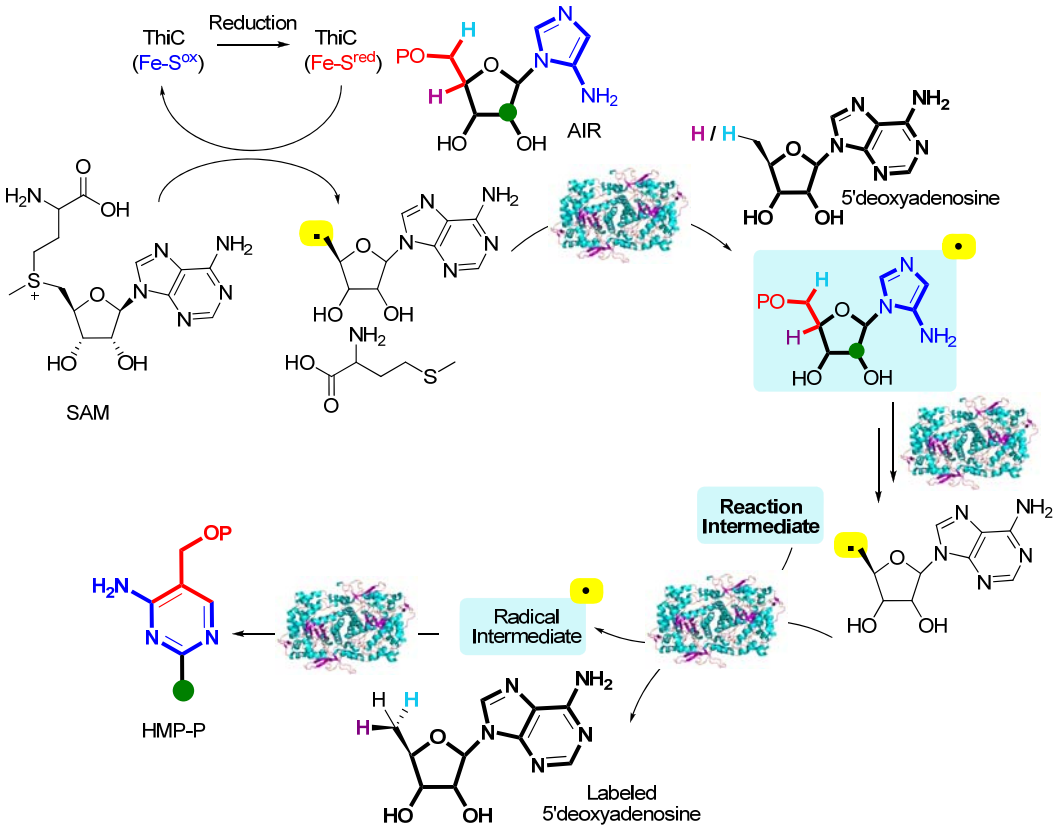


Figure 5.9: ThiC reaction pathway including the two radical abstractions by 5'deoxadenosyl radical.

A question one may ask in this regard is the order of abstraction of the two H atoms from the AIR i.e. does the 4-H get abstracted in the first step or does the 5-H get abstracted. This turns out to be mechanistically very important as the order of H atom abstraction can lead us to formulate a plausible mechanism for the rearrangement and then look for interesting intermediates on the pathway. To figure out the order of abstraction, one needs to either be able to: i) Rapid quench - set up a ThiC reconstitution reaction individually for AIR substrate labeled at the 4-H and the 5-H position, stop the reaction at a point where only the first H abstraction by the

5’deoxyadenosyl radical has occurred for the substrate and check for which D was abstracted by analyzing the 5’deoxyadenosine produced or ii) Competition experiment - compare the rate at which a substrate labeled with D at either at the 4-H or 5-H position is taken up as compared to the unlabeled substrate by ThiC in a competition experiment.

We proceeded to answer the question about the order of abstraction using the competition experiment. For this, we assume that the first step of radical abstraction is essentially irreversible and is the first ‘committed’ step in the rearrangement reaction. Figure 5.10 illustrates the predicted reaction pathway of the ThiC rearrangement where we see that the step involving H-atom abstraction by a radical to form ‘Radical Intermediate “a”’ will be relatively high energy and the backward reaction may be very low, thus making the reaction essentially ‘irreversible’.

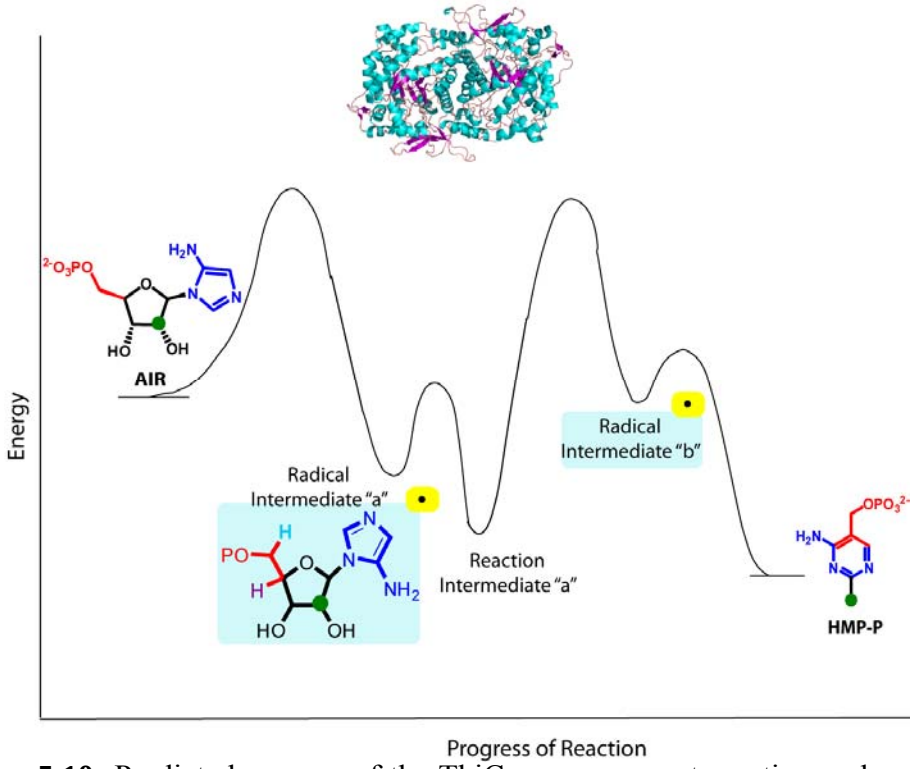


Figure 5.10: Predicted progess of the ThiC rearrangement reaction and energy of the intermediates formed on the pathway.

A known ratio of unlabeled AIR substrate and D-labeled AIR substrate at a specific position (4' or 5'S on the ribose ring) is taken. The ThiC reconstitution reaction is set up with this AIR mixture along with methyl viologen and SAM. The 5'dAdo radical formed by the enzyme can be predicted to choose to pick up the H-substrate over the D-labeled substrate in the first committed step in the reaction. Once the substrate is 'committed' to the reaction in the active site of the enzyme as it cannot go back after the first H-atom abstraction, the enzyme then has to take the substrate to completion and hence the second radical abstraction from the substrate follows along. Subsequently, the products HMP-P and 5'deoxyadenosine are released. The 5'dAdo was analyzed by inline-LC-ESI-MS and the ratio of labeled to unlabeled

Table 5.1: Experimental parameters for the isotope effect 'competition' experiment for [4'-^{2}H] AIR and [5'S-^{2}H] AIR

(a)

Reaction number	Ratio of H-AIR: [4'-^{2}H] AIR	Volume of ~30mM AIR mixture uL	20mM SAM ul	Methyl viologen ul	Volume of ~150uM ThiC uL	Ratio of k_H/ k_D	Isotope effect value
1.	1.01	6	3.5	5.5	350	1.55	1.56 +/- 0.30
2	1.63	6	3.5	5.5	350	1.86	
3.	1.46	6	3.5	5.5	350	1.27	
Final conc	N/A	493 uM	192 uM	Determined by color	144 uM	N/A	

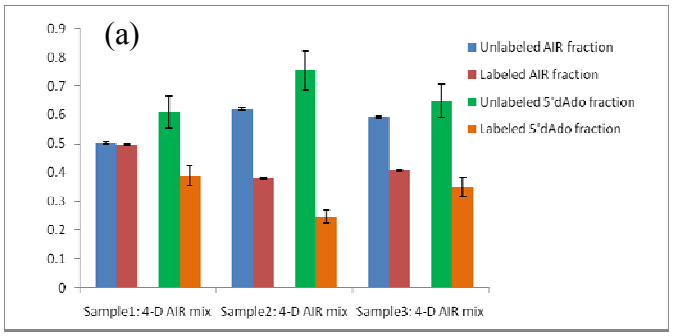
Table 5.1 (Continued):

(b)

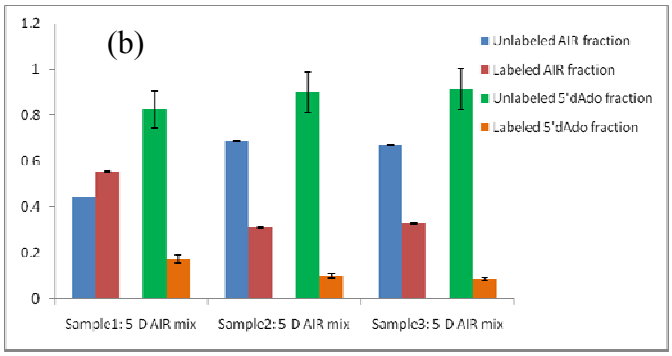
Reaction number	Ratio of H-AIR:[5'S-^{2}H] AIR	Volume of ~30mM AIR mixture uL	20mM SAM	Methyl viologen	Volume of ~150uM ThiC uL	Ratio of k_H/k_D	Isotope effect value
1.	0.804494	6	3.5	5.5	350	5.98	5.13 +/- 0.92
2.	2.216052	6	3.5	5.5	350	4.15	
3.	2.000785	6	3.5	5.5	350	5.24	
Final conc	N/A	493 uM	192 uM	Determined by color	144 uM	N/A	

5'deoxyadenosine was measured. Each LC-ESI-MS detection was performed five times and three different ratios of unlabeled AIR substrate and ^{2}H-labeled AIR substrate were used to validate this experiment and generate the isotope effect number.

If the ratio of ^{2}H-labeled 5'dAdo to unlabeled 5'dAdo is the same as that of the substrate ratio of [4'-^{2}H] AIR and [5'S-^{2}H] AIR that was provided to the enzyme, that ribose-H was abstracted second. If the ratio of ^{2}H-5'dAdo to H-5'dAdo is significantly lower than that of the substrate ratio of [4'-^{2}H] AIR and [5'S-^{2}H] AIR, that ribose-H was abstracted first.



$$\frac{k_H}{k_D} = 1.56 \pm 0.30$$



$$\frac{k_H}{k_D} = 5.13 \pm 0.92$$

Figure 5.11: Isotope effect for (a) [5'^{2}H-AIR] versus (b) [4'^{2}H-AIR] with ThiC. Blue = unlabeled AIR, red = labeled AIR, green = unlabeled 5'dAdo, red = labeled 5'dAdo. The relative amount of unlabeled to labeled substrate is shown in each bar graph against the amount of unlabeled to labeled 5'dAdo produced.

Hence, if there is a significant isotope effect seen in a labeled and unlabeled substrate mixture, we can determine the first step for the reaction. It is important for our analysis to first of all ensure that the amount of HMP-P and 5'deoxyadenosine are stoichiometrically produced, ensuring no uncoupled production of unlabeled 5'deoxyadenosine. All the reactions that were performed for the competition experiment are shown in Table 5.1 (a) and 5.1 (b). For each reaction, an HPLC analysis was done after removing the methyl viologen to quantitate the amount of HMP-P produced versus the amount of 5'deoxyadenosine produced. The ratios of D-labeled to unlabeled 5'deoxyadenosine were found to be ~1:1. The ratio of the unlabeled to labeled substrate AIR (Column 2, Table 5.1 (a) and (b)) and the

corresponding ratio of unlabeled to labeled 5'deoxyadenosine (Column 7, Table 5.1 (a) and (b)) indicates that there is a H/D isotope effect for the [5-^{2}H]AIR but not for the [4-^{2}H]AIR i.e. the unlabeled AIR substrate is specifically picked up over the [5-^{2}H]AIR substrate by the enzyme, but there is no such preference in the case of the [4-^{2}H]AIR substrate as compared to the unlabeled AIR. This can be seen from the values showing a clearly significant isotope effect of 5.13 +/- 0.92 for the [5-^{2}H]AIR in the reaction (Column 8, Table 5.1 (b)) and a very small isotope effect of 1.56 +/- 0.30 for the 4D-AIR in the reaction (Column 8, Table 5.1 (a)) (Figure 5.11).

Thus, when the 5'dAdo radical in the enzyme ThiC encounters a choice between abstraction of a H or a D (because we have a mixture of [5-^{2}H]AIR and 5H-AIR), the enzyme chooses to pick up the H-labeled AIR preferentially, and continue its reaction. Hence, the amount of unlabeled 5'deoxyadenosine is greater. In the case of the 4D-AIR and H-AIR mixture, 5'dAdo radical in the enzyme ThiC encounters a H at the 5-position for both the [4-^{2}H]AIR and the [5-^{2}H]AIR. It picks up the 5-H non-selectively from either of these molecules, and then the molecule, whether labeled or unlabeled at the 4-position is committed to the reaction. Hence, when the next H abstraction from the 4-H position happens, an equal amount of labeled and unlabeled 5'deoxyadenosine is formed.

The small isotope effect for 4D-AIR can be attributed to the leaking out of intermediates from the ThiC active site as the reaction proceeds, the 5'deoxyadenosine doing only the first H-abstraction and not completing the second abstraction. This is a possibility especially when the second radical abstraction has to occur and there is a significant isotope effect to abstract a D over a H from the 4-H-AIR position Both of these reasons can explain the small isotope effect observed for the 4D-AIR for the 5'deoxadenosine measurement.

5.3 Conclusion:

We have been able to determine the fate of the C1 and C3 atoms of AIR. We showed that C1 leaves as formate by 13C-NMR and that C3 leaves as carbon monoxide by a spectrophotometric assay with reduced hemoglobin in anaerobic conditions. Further, we were able to show that the 5'deoxyadenosyl radical formed from SAM can abstract two H's from the substrate, and that the H at 5-H-AIR is abstracted first, followed by the H at 4-H-AIR. Another interesting point to note is that the cleaving of the protein that is observed only occurs on exposure to air, and is nonspecific with regard to the reaction. Gel analysis of the protein after the reaction in anaerobic conditions did not show any cleaving of the protein. All these observations are essential data to help us to formulate a plausible mechanism for this complex rearrangement reaction.

5.4 Experimental Methods

All chemicals were purchased from Sigma-Aldrich Co. unless otherwise mentioned. HPLC analysis was performed using an Agilent 1100 instrument equipped with a diode array detector. LB medium was obtained from EMD Biosciences. Ampicillin and isopropyl β-D-thiogalactoside (IPTG) were purchased from LabScientific Inc. Chloramphenicol was purchased from USB (Ohio). Deuterated ribose was purchased from Omicron Biochemicals (South Bend, IN). AIR and HMP-P were synthesized as previously described[14]. A 150x4.6 mm Supelco LC-18-T column was used for analytical purposes, whereas a 250x10 mm semi-prep Supelco LC-18 column was used for the isolation of 5-deoxyadenosine. ESI-MS coupled HPLC analysis was performed on an Agilent 1100 instrument equipped with an in line Micromass Quattro ESI-mass spectrometer. For the isotope effect experiments, ESI-MS coupled HPLC analysis was performed on a Hewlett Packard 1100 HPLC equipped with a ThermoFisher DecaXP ion trap mass spectrometer.

5.4.1 Protein expression and purification for ThiC:

A full length clone of *Caulobacter crescentus* ThiC in pET28 vector was co-transformed with pDB1282, a plasmid encoding genes responsible for the Fe/S cluster biogenesis machinery, into *E. coli* B834(DE3) cell line. A 10 mL overnight culture of the resulting strain in LB was used to inoculate 1.5 L of sterilized M9 minimal medium, supplemented with 100 mg/L ampicillin and 40 mg/L kanamycin, and it was allowed to grow at 37 °C till the OD_{600} reached 0.1-0.2. At this point, ferrous ammonium sulfate, L-cysteine and L-arabinose were added to the final concentrations of 100 μM, 500 μM and 0.2% (w/v) respectively and the growth was allowed to continue with slow shaking (50 rpm). Once the OD_{600} reached 0.6, the cultures were cooled with ice-water and IPTG was added to a final concentration of 10 μM. The protein overexpression was continued at 15 °C with slow shaking for 18 hours and subsequently the cells were cooled again to 4 °C for 2-3 hours before harvesting and storing under liquid nitrogen till future use.

For protein purification, the cell pellet (~10 g) was transferred to an anaerobic chamber (Coy Laboratories) and allowed to thaw. The pellet was resuspended with 50 mL loading buffer (200 mM Tris-HCl, 1 mM DTT, pH 7.6) and 25 mg lysozyme and 1000 U of benzonase were added to it. After two hours of incubation at room temperature with continuous mixing (magnetic stir-plate/bar) the resuspended cells were further lysed by sonication (Misonic XL-2000; 5x1 min cycles at the highest power setting) on ice. The crude lysate was clarified by centrifugation and the clarified lysate was loaded onto two 5 mL HisTrap™ (GE healthcare) column (arranged in tandem) pre-equilibrated with the loading buffer. The column was then washed with 50 mL wash buffer (loading buffer with 20 mM imidazole). The bound protein was eluted with elution buffer (loading buffer with 300 mM imidazole). The colored

fractions were desalted into the storage buffer using a Bio-Rad 10 DG desalting column (200 mM Tris-HCl, 4 mM DTT, 40% glycerol, pH 7.6). Aliquots of this protein preparation were stored under liquid nitrogen till further use. Protein concentration, iron and sulfide content were measured as described previously[15].

5.4.2 Protein expression and purification for AIRs kinase, Flavodoxin and flavodoxin reductase:

Overexpression plasmids for the 6xHis tagged recombinant proteins were transformed in *E. coli* BL21 (DE3) cell strain. 10 mL overnight cultures of the resulting strain in LB was used to inoculate 1.5 L LB medium supplemented with 100 mg/L ampicillin (flavodoxin) or 40 mg/L kanamycin (flavodoxin reductase). Protein expression was induced with 1 mM IPTG at OD600 of 0.6 and the culture was supplemented with 100 uM riboflavin (for flavodoxin and flavodoxin reductase). After 16 hours of overexpression at 15 C, cells were harvested, frozen and stored at -80 C.

To purify the protein, the cell pellets were resuspended in 25 ml lysis buffer (10mM imidazole, 300mM NaCl, 50mM NaH_2PO_4, pH 8) and lysed by sonication on ice (Heat systems Ultrasonics model W-385 sonicator, 2s cycle, 50% duty). The resulting cell lysate was clarified by centrifugation and the protein was purified on Ni-NTA resin following the manufacturer's (Qiagen) instructions. After elution, the protein was desalted using a 10 DG column (GE Healthcare) pre-equilibrated with the 50mM potassium phosphate buffer, 2 mM DTT and 30% glycerol, pH 8.0. Aliquots of both proteins were frozen with liquid nitrogen and stored at -80 C for future use.

5.4.3 Preparation of AIR from AIRs:

A typical reaction mixture consisted of 25 µL 100 mM AIRs, 10 µL 400 mM ATP, 5 µL 1 M $MgCl_2$, 900 µL 25 mM KPi (pH 7.5) and 100 µL of the AIRs kinase stock.

The reaction mixture was allowed to incubate at room temperature for 1 hr and was lyophilized till dry. The residue was transferred into the glove box and resuspended in degassed water to a final volume of 100 µL (final AIR concentration of ~25 mM). This AIR stock was directly used for ThiC reactions.

5.4.4 NMR experiments with deuterated AIR:

A typical 2 mL reaction mixture consisted of 80 uM flavodoxin, 40 uM flavodoxin reductase, 3 mM NADPH, 300-500 uM ThiC (depending on the protein preparation), 500 uM AIR and 1 mM Sam-Cl. The reaction was allowed to proceed for 16 hours anaerobically and heat quenched subsequently. The product 5-dAd was analyzed and isolated by HPLC as previously described. Isolated 5-dAd was subjected to ^{1}H NMR analysis (Varian 600 MHz; only the 5-dAd resulting from 5,5′-D_2 AIR was analyzed using a Varian 500 MHz instrument).

To monitor the time dependent formation of the 5-dAd from 4-D and 5,5′-D_2 AIR, 100 µL aliquots were removed at desired time intervals from the reaction mixture, heat quenched and filtered using 10 kDa MW cut-off membrane filters (microcon). The filtrate was analyzed by HPLC-coupled ESI-MS to obtain the product ratio as well as the mass of 5-dAd.

5.4.5 NMR experiment with C1-^{13}C AIR and C3-^{13}C AIR:

A typical reaction mixture (1 mL) consisted of 600 µM ThiC + 600 µM AIR + 2 mM SAM-Cl + 10 mM dithionite. The reaction mixture was allowed to incubate for 30 minutes at room temperature anaerobically and was heat quenched. Precipitated protein was removed by centrifugation and to the supernatant was added D_2O (Cambridge Isotope Laboratories) to a final concentration of 10 % (v/v). It was transferred to a Shigemi NMR tube, susceptibility matched for D_2O and was analyzed

by ^{13}C NMR (Varian 500 MHz). A control reaction, where SAM-Cl was omitted, was also set up and processed in parallel, identical to the reaction mixture.

5.4.6 LC-MS analysis of deuterated 5-dAd from ThiC reaction mixture:

Reactions were set up as described in section D. However, only 100 µL reaction mixture was set up with each individual deuterated substrate. Upon the completion of the reaction, the protein was heat denatured and the precipitated mass was removed by centrifugation. The supernatant was filtered through a 10 KDa MW cut-off filter and the filtrate was used directly for LC-MS analysis. A different HPLC protocol was used for these experiments. Following program was used with a linear gradient at 0.5 mL/min: Solvent A: H_2O, solvent B: Methanol. 0 min: 100% A; 5 min: 100% A; 25 min: 35% A, 65% B; 30 min: 100% B; 35 min: 100% A.

5.4.7 Standard curve for hemoglobin-CO detection:

All buffers and reaction solutions were made using degassed water and inside the anaerobic chamber. 3.5 µM hemoglobin solution was made in 10mM MOPS, 0.9% NaCl buffer, pH 7.2 and 2mM final concentration of dithionite was added to it to convert all oxy hemoglobin to hemoglobin and to convert any methemoglobin (ferric) to the reduced (ferrous) species[25]. The solution was then allowed to stand for 1 hour. A saturated carbon monoxide(CO) standard solution was made by bubbling in CO gas (99% pure) into a sealed round-bottomed flask containing 10mM MOPS, 0.9% NaCl buffer, pH 7.2. This standard saturated solution of CO was then added in increasing volumes (0, 1,2,5,10, 20,50, 100 µL) to 8 fractions of 1mL each of the reduced hemoglobin solution and the total volume was made up to 1100 µL in each case by adding the required volume of the MOPS-NaCl buffer. The difference profile of the

hemoglobin-CO was taken with respect to the hemoglobin with no hemoglobin added to it and a change in the soret region of the absorbance is observed (Figure 5.4.1).

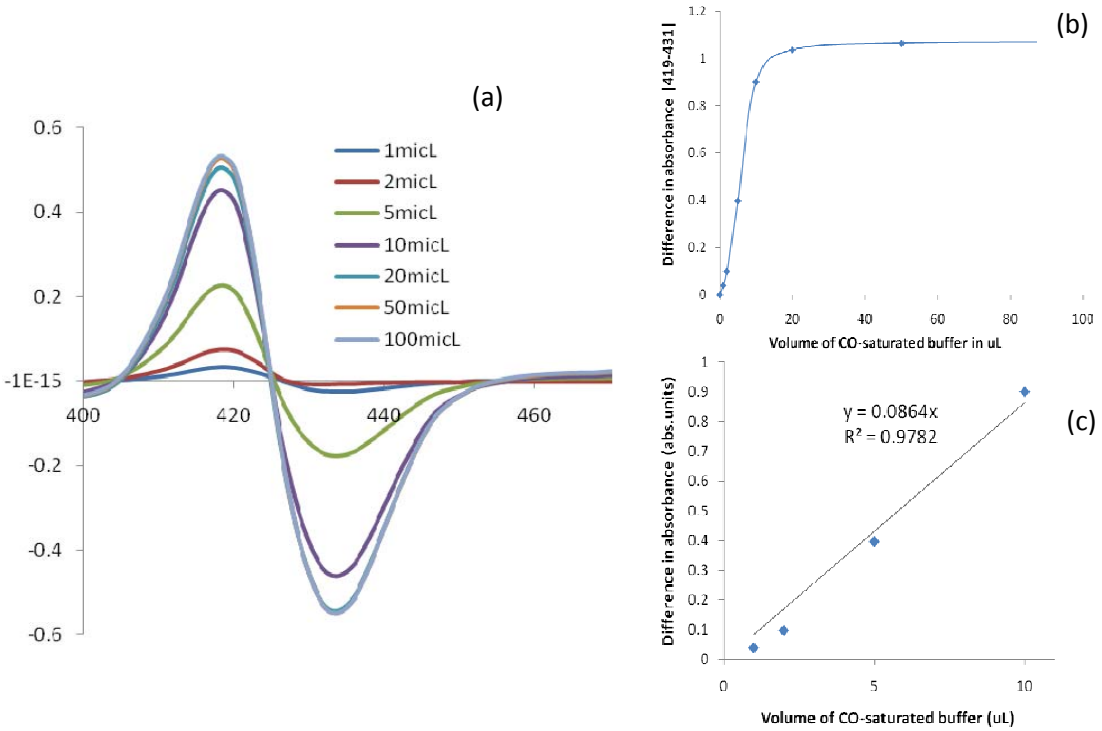


Figure 5.4.1: (a) Change in absorbance in the soret region of the spectrum of hemoglobin when bound to increasing amounts of saturated CO buffer. (b) Graphical representation of (a) plotting the difference in absorbance values i.e |419-431| with respect to the volumes of saturated CO buffer added. (c) Plot of the linear region of the difference spectrum with volume to obtain the linear equation to quantify the amount of CO bound to hemoglobin.

5.4.8 Detection of CO released in ThiC catalyzed reconstitution reaction:

A typical reaction mixture (100 mL) consisted of 100 µM ThiC + 12.5 µM AIR + 1 mM SAM-Cl + 10 mM dithionite. An identical control reaction with all components except AIR was also set up. The reaction mixtures was allowed to incubate for 30

minutes at room temperature anaerobically and 1mL of reduced hemoglobin was added to each - the full reaction mixture as well as to the control reaction mixture. The two solutions were then transferred into air-tight UV-Vis cuvettes in the anaerobic chamber and sealed before measuring the absorbance of the full reaction using a Cary 300 Bio double beam UV-Vis spectrophotometer, where the control reaction was used as the background. The same procedure was repeated for increasing concentrations of AIR in the full reaction mixture - 16.6 μM and 25μM of AIR and the UV-Vis absorbance profile for the change in the soret region of hemoglobin die to binding with CO was recorded.

5.4.9 Attempts at detection of formaldehyde as a product of the ThiC reaction by 4-amino - 5 - hydrazino - 3 - mercapto - 1,2,4 –triazole (Purpald):

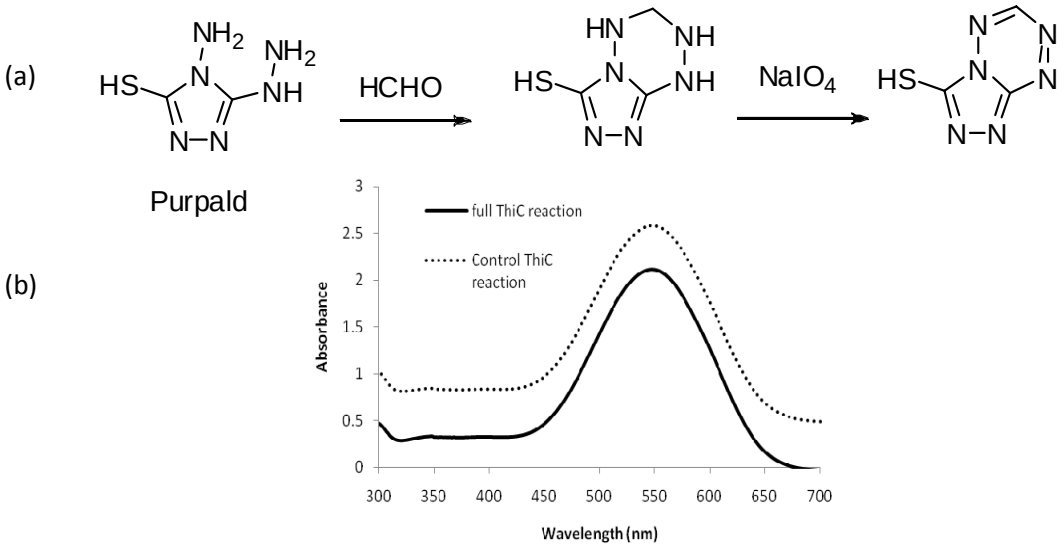


Figure 5.4.2: (a) Scheme for reaction of dimedone with formaldehyde and Purpald with formaldehyde (b) UV-Vis absorbance profile of full ThiC reaction and the control ThiC reaction with Purpald.

Purpald is a reagent that reacts with formaldehyde and is further oxidized to form an intense purple color with characteristic absorbance at 355nm and 550nm (Figure 5.4.2 (a)). A standard solution of 5mM formaldehyde was reacted with 10mM Purpald and further oxidized with 10mM $NaIO_4$ to form the Purpald-formaldehyde adduct which has maxima at 355nm and 550nm. A typical reaction mixture (100 mL) consisted of 200 µM ThiC, 400 µM SAM, 200 µM AIR and 500 µM dithionite and a control reaction consisted of ThiC without SAM. Both these reactions were then treated with 10mM of the reagent Purpald, an aldehyde-trapping agent, followed by oxidation by 10mM $NaIO_4$. It was observed that for the ThiC reactions, the absorbance profile of Purpald looked identical in the case of the control reaction as well as the full reconstitution reaction as shown in Figure 5.4.2 (b). Also, no significant peak was seen at 355nm. This reaction was repeated twice with other controls for the ThiC reaction, yielding identical results.

Hence, we can see there is no formaldehyde being produced in the ThiC reconstitution reaction. However, as observed from the absorbance of the reaction-Purpald adduct, some other aldehyde seems to be forming, as Purpald is showing an absorbance for complexing with an carbonyl-aldehyde group which has to be investigated further.

REFERENCES

1. Jordan, F. *Nat. Prod. Rep*. **2003**, 20, 184-201.

2. Begley, T. P. *Nat. Prod. Rep.* **2006**, 23, 15-25.

3. Settembre, E., Begley, T.P. & Ealick, S.E *Curr. Opin. Struct. Biol.* **2003**, 13, 739-747.

4. Chatterjee, A., Jurgenson, C.T., Schroeder, F.C., Ealick, S.E. & Begley, T.P. *J. Am. Chem. Soc.* **2006**, 128, 7158-7159.

5. Chatterjee, A., Jurgenson, C.T., Schroeder, F.C., Ealick, S.E. & Begley, T.P. *J. Am. Chem. Soc.* **2007**, 129, 2914-2922.

6. Jurgenson, C.T., Chatterjee, A., Begley, T.P. & Ealick, S.E. *Biochemistry* **2006**, 45, 11061-11070.

7. Kriek, M. *et al. J. Biol. Chem.* **2007**, 282, 17413-17423.

8. Tayuza, K.; Yamada, K.; Kumaoka, H. *Biochimica et Biophysica Acta* **1989** 990, 73 -79

9. Tayuza, K.; Yamada, K.; Kumaoka, H. *Methods in Enzymology* **1997** 279, 97-108

10. Newell, P.C. & Tucker, R.G. *Biochem. J.* **1968**, 106, 279-287.

11. C. Costello, Ph.D. dissertation, Cornell University, **1996**.

12. Y. Zhang and T. P. Begley, *Gene*, **1997**, 198, 73–82.

13. Dougherty, M.J.; Downs D.M.. *Microbiology* **2006**, 152, 2345-2353

14. Lawhorn, B.G., Mehl, R.A. & Begley, T.P. *Org. Biomol. Chem.* **2004**, 2, 2538-2546.

15. Chatterjee, A., Li, Y., Zhang, Y., Grove, T.L., Lee, M., Krebs, C., Booker, S.J., Begley, T.P. & Ealick, S.E. *Nat Chem Biol.* **2008**, 4, 758-65.

16. Martinez-Gomez, N.C., Downs, D.M. *Biochemistry*. **2008**, 47, 9054-9056.

17. Chatterjee, A. Ph.D. dissertation. Cornell University **2009**

18. Frey, P.A. & Booker, S.J. *Adv. Protein Chem*. **2001**, 58, 1-45.

19. Wang, S.C. & Frey, P.A. *Trends Biochem. Sci.* **2007**, 32, 101-110.

20. Sofia, H.J., Chen, G., Hetzler, B.G., Reyes-Spindola, J.F. & Miller, N.E. *Nucleic Acids Res* **29**, 1097-1106.

21. Frey, M., Rothe, M., Wagner, A.F. & Knappe, J. *J Biol Chem*. **1994**, 269, 12432-7.

22. Wang,J.; Stolowich, N.J.;,Santander, P.J.; Park, J.H.; Scott A.I. *Proc. Natl. Acad. Sci.* **1996** 93, 14320–14322

23. Romerol M.A., Blunden, G., Carpenter B.G., Tyihfik E. Chromatographia **1999**, 50, 160-166

24. Avigad G. Analytical Biochemistry **1983**, 134, 499-504

25. Bonam, D., Murrell, S.A., Ludden, P.W. J. Bact. **1984**, 159, 693-699

26. Al-Qahtani, M.H.; Pike, V.W. *J. Chem. Soc*., *Perkin Trans.* **2000**, 1, 1033–1036

27. Al-Qahtani, M.H.; Pike, V.W. J. *Labelled Cpd. Radiopharm* 2000, 43, 825-835

28. Lidström,P., Kihlberg, T. and Långström, B. *J. Chem. Soc., Perkin Trans.* **1997** 1, 2701-2706

29. Y. Andersson and B. Långström, *J. Chem.Soc.*, *Perkin Trans. 1*, 1995, 287.

30. Loupiac, C., Pin, S., Vezin, H., Alpert, B. *Chemical Physics Letters* **2001**, 344, 457

31. Foldesi, A, Nilson, F, Peder R., Glemarec, C; Gioeli, C; Chattopadhyaya, J. *Tetrahedron* **1992**, 48(41), 9033-72.

32. Hanes, J.W., Keresztes, I., Begley, T.P. Nat Chem Biol. **2008**,7,425-30

33. Xiao Y, Chu L, Sanakis Y, Liu P. *J Am Chem Soc* **2009** Jul 29;131(29):9931

CHAPTER 6

The remarkable rearrangement reaction catalyzed by 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase: Mutagenesis studies on ThiC and studies on the activity of the mutants

6.1 Introduction

After having investigated the fate of every C from AIR, and establishing the two H-atom radical shuttle done by the 5'deoxyadenosyl radical, it was yet difficult to establish a mechanism for the rearrangement of ThiC without knowing the structures of any of the intermediates on the pathway. Additionally, ThiC is a very fast enzyme, hence observing any intermediate steps of the reaction is also not currently feasible. One of the strategies to look for intermediates is to create important site directed mutants of the protein. This strategy has many advantages – i) active site mutants may not be able to convert the substrate fully to the product but may be partially active. This is very useful for getting the reaction to stop at a particular step in the reaction, thus identifying intermediates on the pathway[1-2] ii) It has been observed in the past that active site mutants come bound with relevant metabolites which can be purified and analyzed for their structure. These metabolites may be intermediates of the reaction or relevant molecules associated with the physiological role of the protein and can give us an idea of the functioning of the enzyme. iii) It has been observed that mutants of enzymes may be able to do the entire reaction but at a much slower rate. This would help in enzymatic characterization of the reaction.

6.2 Results/ Discussion

With these possibilities in mind, site directed mutants were created for the *C.cresentus* ThiC protein at C474S, E413Q, Y440F, Y277F, double mutant H417A-H481A, M248L, R377K and C333A. The mutants are as shown in Figure 6.1. Of these, mutants C474S, E413Q, Y277F and H417A-H481A were taken up for the present analysis.

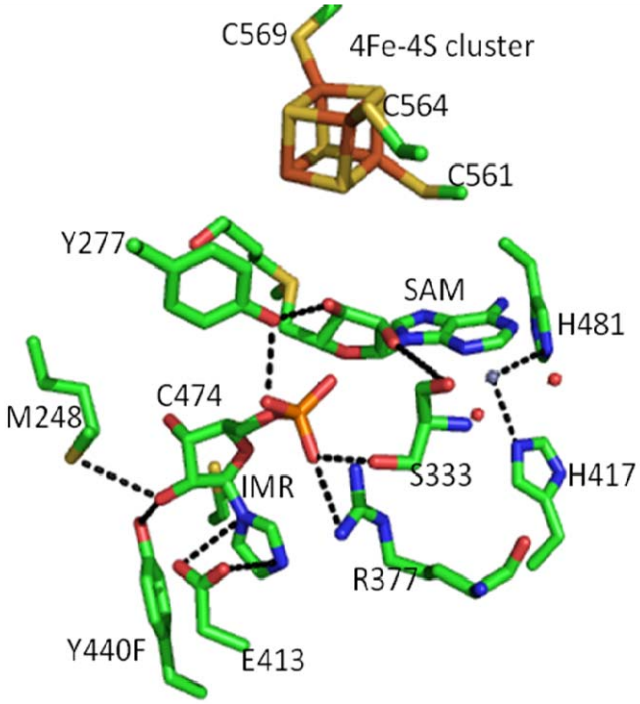


Figure 6.1: Active site residues of ThiC that have been mutated for mechanistic investigation.

The mutant plasmids C474S, E413Q, Y277F and H417A-H481A were cotransformed into B834(DE3) along with the pDB1282 plasmid (Fe-S cluster chaperone proteins) and were subsequently overexpressed in minimal media. The proteins were purified

and analyzed by gel electrophoresis (Figure 6.2). It can be seen that the H417A-H481A mutant does not cleave when exposed to air.

All of the mutants were then analyzed for reconstitution activity. Each of the five mutants was set up for a reconstitution reaction with AIR, SAM and dithionite or

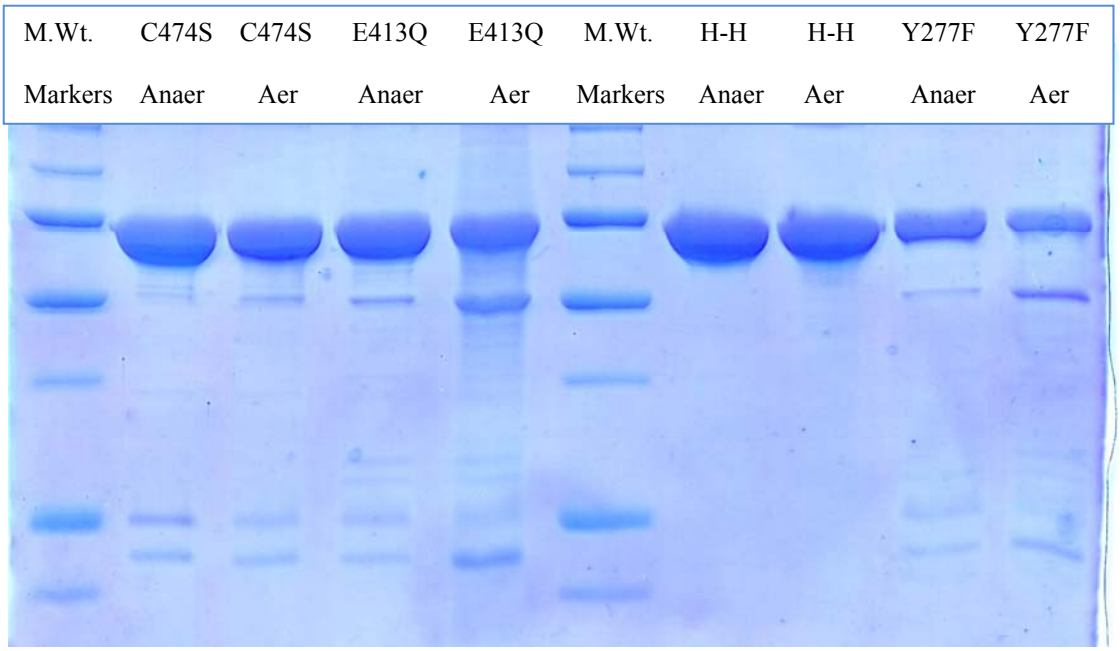


Figure 6.2: SDS-PAGE analysis of mutants of ThiC in anaerobic and aerobic conditions

methyl viologen. All appropriate controls with AIR, SAM and the reducing agent missing were performed. HPLC analyses of all these reactions were done to look for the formation of HMP-P and 5'deoxyadenosine.

Interestingly, two of the mutants Y277F and C474S showed the production of 5'deoxyadenosine but no formation of HMP-P by RP-HPLC analysis (Figure 6.3). The 5'deoxyadenosine peaks were collected and analyzed by ESI-MS to confirm their identity. One of the possibilities for observing production of 5'deoxyadenosine is that the reaction was being carried on to some extent in the enzyme, but not to completion.

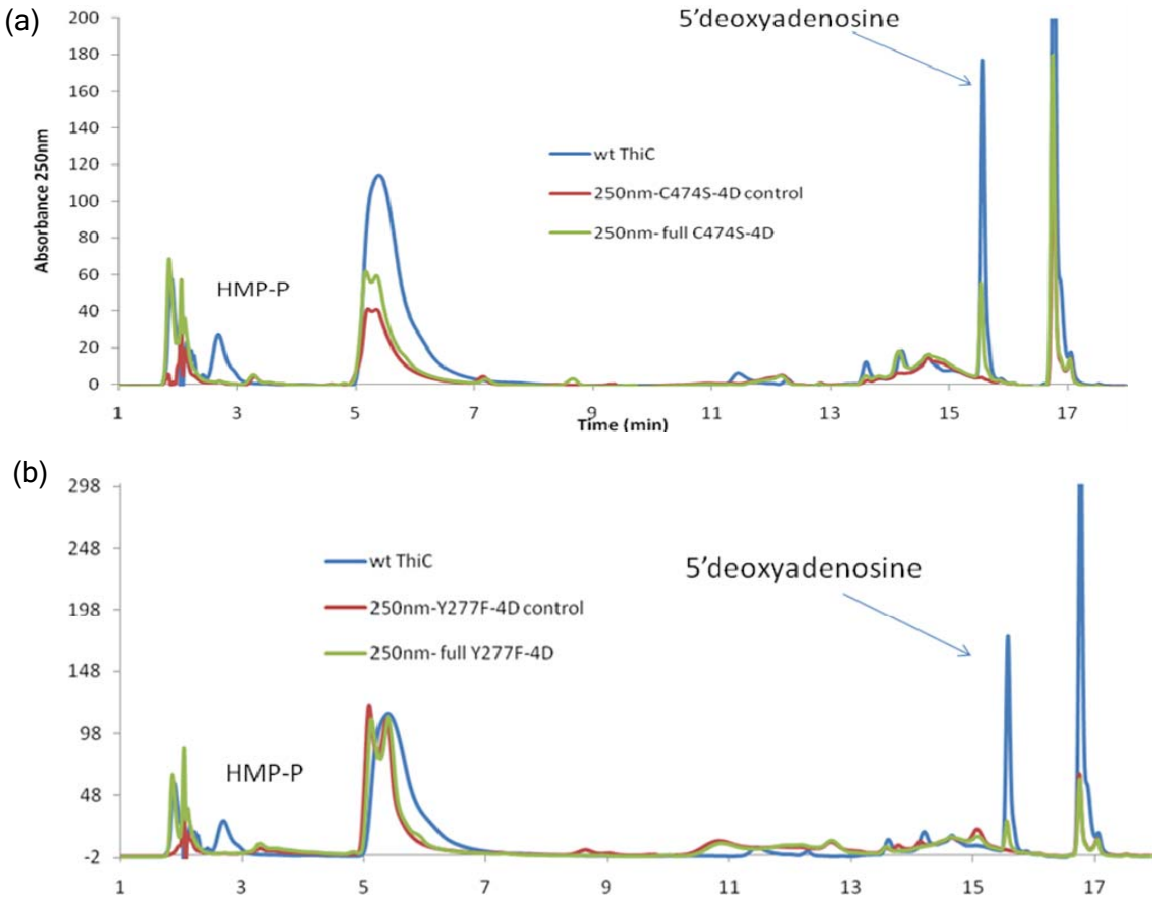


Figure 6.3: HPLC chromatogram showing production of 5'deoxyadenosine in the reconstitution reaction by the mutants (a) C474S and (b) Y277F

These mutants would be of great value in finding interesting intermediates on the pathway. The other possibility is that there is uncoupled production of 5'deoxyadenosine in these mutants, as they were both in the active site. The mutations could have perturbed the active site adequately to produce uncoupled 5'deoxyadenosine. To distinguish between these two possibilities, the reconstitution reaction was carried out using the two mutants Y277F and C474S and the two labeled AIR molecules, [4'-^{2}H] AIR and [5'-^{2}H]AIR.

In the case of the C474S mutant, we could clearly see the production of 5'deoxyadenosine in the case of both [4'-^{2}H]AIR and [5'-^{2}H]AIR substrates by HPLC.

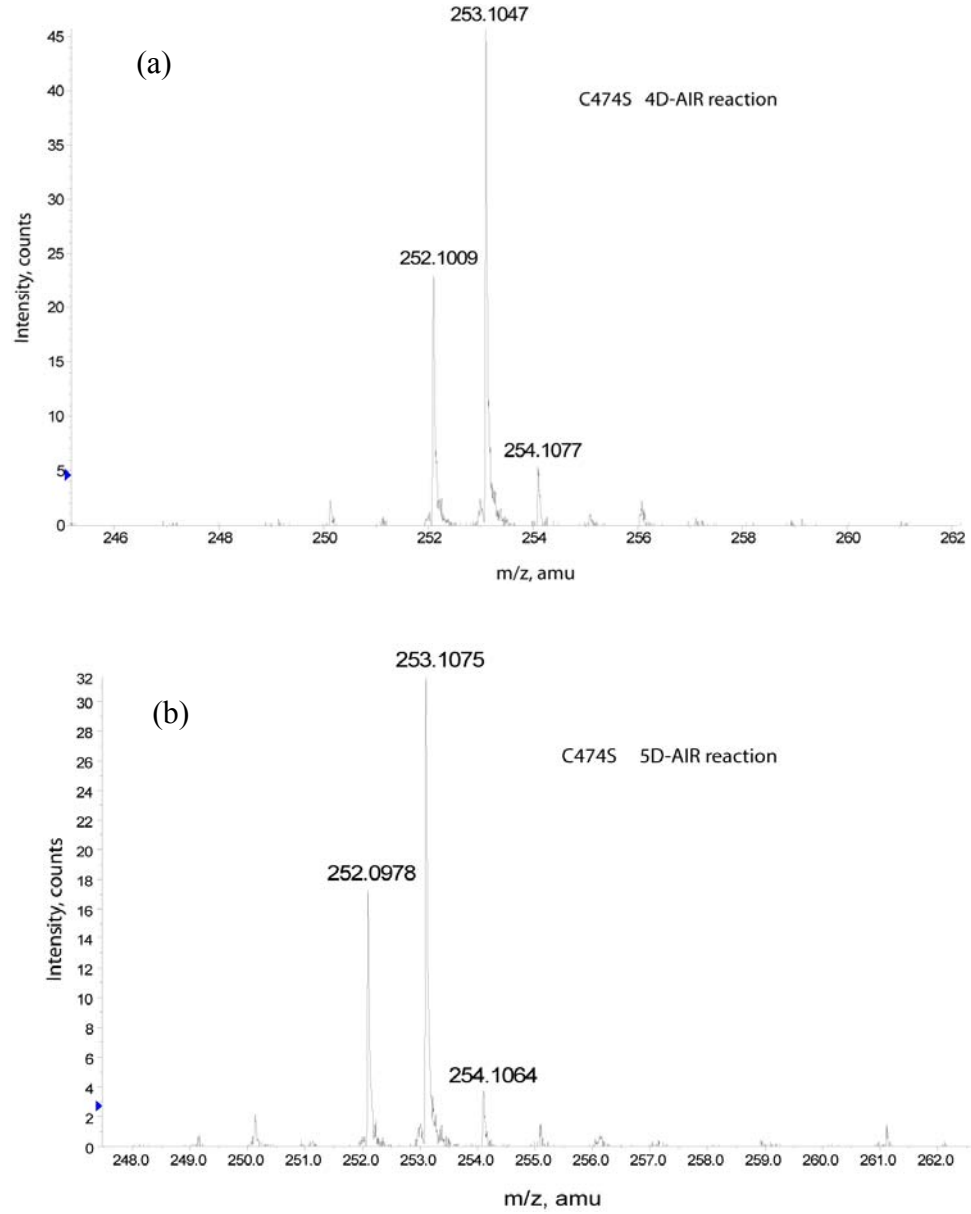
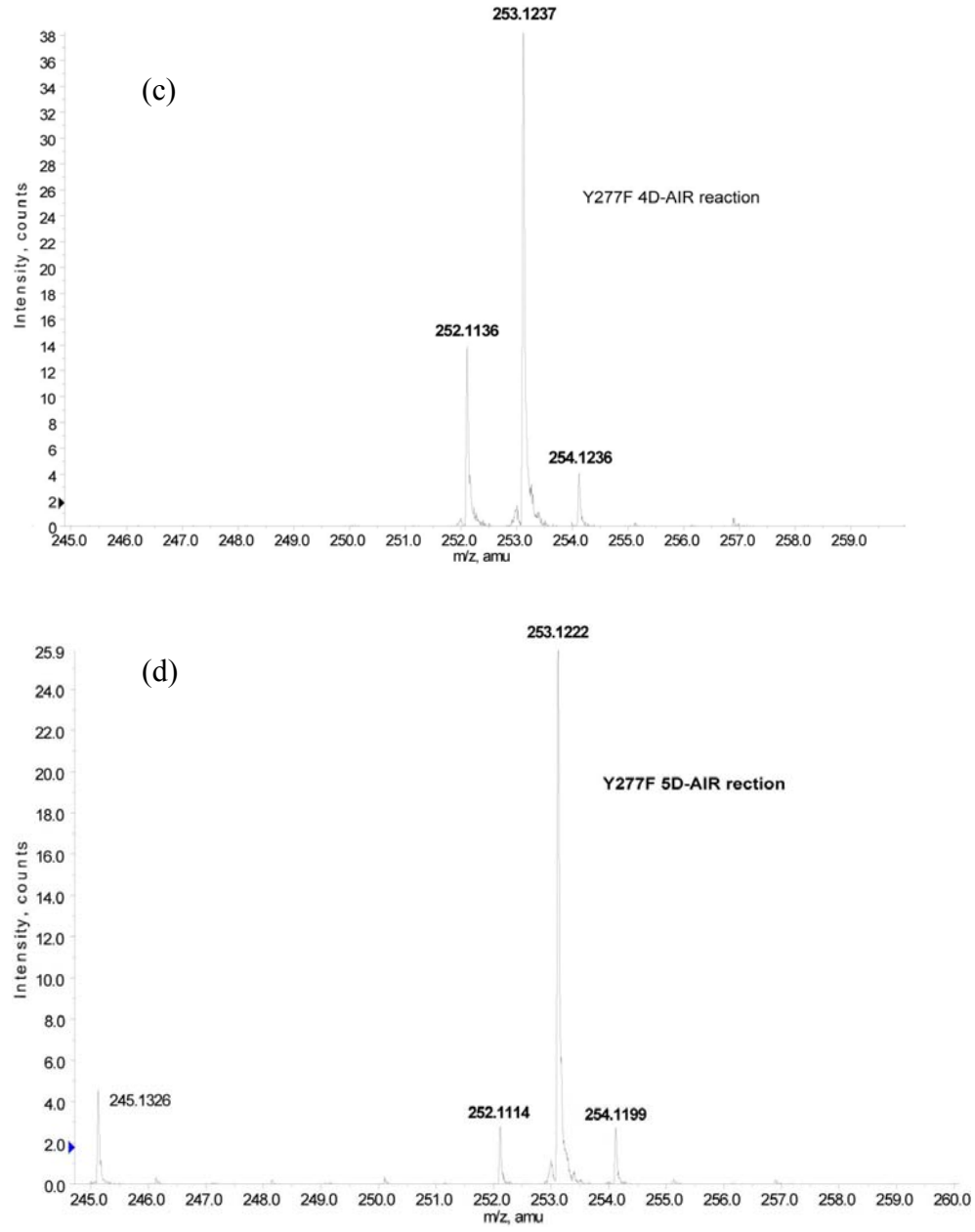


Figure 6.4: ESI-MS of labeled 5'deoxyadenosine of (a) C474S with [4'-^{2}H]AIR and (b) C474S with [5'S-^{2}H]AIR and (c) Y277F with [5'S-^{2}H]AIR and (d)Y277F with [4'-^{2}H]AIR .

Figure 6.4 (Continued)



The 5'deoxyadenosine peak was collected in both the reactions and analyzed by positive mode ESI-MS. The MS analysis showed the incorporation of ^{2}H from the [4'-^{2}H]AIR and [5'-^{2}H]AIR into the 5'deoxyadenosine (Figure 6.4).

In the case of the Y277F mutant, we could clearly see the production of 5'deoxyadenosine in the case of both [4'-^{2}H]AIR and [5'-^{2}H] AIR substrates by HPLC, however the quantity produced is lesser than in the C474S mutant. The ESI-MS results show that the D is being abstracted from the [5'-^{2}H -AIR] (Figure 6.4) as well as the [4'-^{2}H]AIR. Interestingly, the amount of uncoupled 5'deoxyadenosine (mass of 252.10 Da) is consistently greater in the case of the [4'-^{2}H] AIR as compared to the [5'-^{2}H] AIR. The ratio of labeled to unlabeled 5'deoxyadenosine when [4'-^{2}H] AIR was used was 2.8 whereas the ratio when [5'-^{2}H] AIR was used was 8.4.

This result leads us to an interesting corollary of which H is abstracted first from the AIR by the 5'deoxyadenosyl radical. As we have shown, the 5-H is abstracted first. This is proved here, as in the case of the [5'-^{2}H] AIR, the D is being abstracted first. So if there is any leaking out of 5'deoxyadenosine as the reaction goes forward, only the labeled 5'deoxyadenosine will be observed. In the case of the [4'-^{2}H] AIR, the H is abstracted first to form the unlabeled 5'deoxyadenosine. So, if there is any leaking out of 5'deoxyadenosine as the reaction goes forward, only the unlabeled 5'deoxyadenosine will be leaked out. Clearly, in the case of this mutant, there is leaking out observed, clarified by the reaction done with the [4'-^{2}H] AIR. Further work needs to be done on this mutant to establish the extent to which it has catalyzed the rearrangement.

6.3 Conclusion

The ThiC rearrangement reaction is not only complex but also very fast, hence difficult to track. No intermediates have yet been successfully trapped from the wt ThiC protein reaction. Hence, we made active site mutants for ThiC, hoping to make a

mutant that was either partially active, or very slow at doing the reaction. In the process, we were able to identify the three mutants E413Q, C474S and Y277F that have interesting biochemical properties. The mutant E413Q has small molecules bound in its active site which have been isolated and their structural characterization is in progess. The mutants C474S and Y277F appear to be partially active, doing the ThiC rearrangement reaction till a particular step of the mechanism, without actually going to completion. Further analysis of these mutants and other active site mutants not yet been tested for activity may be able to provide us with a snapshots of the complex mechanism of the ThiC reaction.

6.4 Experimental Methods

6.4.1 Materials: All chemicals were purchased from Sigma-Aldrich Co. unless otherwise mentioned. HPLC analysis was performed using an Agilent 1100 instrument equipped with a diode array detector. LB medium was obtained from EMD Biosciences. Ampicillin and isopropyl β-D-thiogalactoside (IPTG) were purchased from LabScientific Inc. Deuterated ribose was purchased from Omicron Biochemicals (South Bend, IN). AIR and HMP-P were synthesized as previously described[14]. A 150x4.6 mm Supelco LC-18-T column was used for analytical purposes, whereas a 250x10 mm semi-prep Supelco LC-18 column was used for the isolation of 5-deoxyadenosine. ESI-MS coupled HPLC analysis was performed on an Agilent 1100 instrument equipped with an in line Micromass Quattro ESI-mass spectrometer. For the isotope effect experiments, ESI-MS coupled HPLC analysis was performed on a Hewlett Packard 1100 HPLC equipped with a ThermoFisher DecaXP ion trap mass spectrometer.

6.4.2 Analysis of mutant ThiC protein in anaerobic versus aerobic conditions:

The protein was removed from -80 ^{o}C and allowed to come to room temperature inside the anaerobic chamber (Coy laboratories). Two separate microcentrifuge tubes were taken with 5 μL of the protein in each. One tube was taken out of the glove box and exposed to air for 4 hours. The other one kept inside the anaerobic chamber for the same time. After 4 hours, each sample was quenched with gel sample buffer containing 50mM β-mercaptoethanol. The sample inside the anaerobic chamber was heat-quenched inside the anaerobic chamber and the sample exposed to air was quenched outside the chamber. The two samples for each mutant protein were then analysed by SDS-PAGE by running on a 12% Tris-glycine gel.

6.4.3 Reconstitution of the ThiC reaction and analysis of the 5'deoxyadenosine peak

A typical reaction mixture (100 mL) consisted of 100 μM ThiC + 12.5 μM AIR + 1 mM SAM-Cl. Methyl viologen was added till the blue color of the reagent could be seen i.e. in slight excess after complete reduction of the cluster had occurred. An identical control reaction with all components except AIR was also set up. The reaction mixtures were allowed to incubate for 30 minutes at room temperature anaerobically. The reactions were then analyzed using RP-HPLC and after identifying the production of 5'deoxyadenosine in the full reaction, and its absence in the control sample, the 5' deoxyadenosine was collected and subject to ESI-MS analysis. This reaction was done for the [5'S-^{2}H]AIR and the [4-^{2}H]AIR

REFERENCES

1. Chatterjee, A., Jurgenson, C.T., Schroeder, F.C., Ealick, S.E. & Begley, T.P. *J. Am. Chem. Soc.* **2006**, 128, 7158-7159.

2. Chatterjee, A., Jurgenson, C.T., Schroeder, F.C., Ealick, S.E. & Begley, T.P. *J. Am. Chem. Soc.* **2007**, 129, 2914-2922.

CHAPTER 7

Summary and Outlook

7.1 Summary

The advances in the detailed investigation of thiamin biosynthesis presented in this thesis increase to a larger extent the mechanistic puzzle involved in thiamin biosynthesis. Even though prokaryotic thiamin thiazole biosynthesis already had a lot of detailed work done and a strong foundation already had been established, we stumbled upon many surprises at various points in our research. 'Unstable' intermediates, reversibile enzymatic pathways, clustered enzymes similar to known ones but with unexplored functions and much more was encountered, and we studied them in detail and learnt an immense amount of very novel chemistry. The thiamin pyrimidine biosynthetic pathway had a lot to explore with regard to absolutely unprecedented radical chemistry, handling of air sensitive proteins and rigorous and detailed investigation of each glimpse that the reaction conditions have allowed us till date. There are many ways of advancing the research discussed in this thesis, and we discuss them here:

7.2 Thiamin thiazole biosynthesis:

7.2.1 Product of B. subtilis thiazole synthase:

We established that the product of the thiazole synthase from the thiazole reconstitution reaction was the thiazole tautomer phosphate. We characterized it by HPLC and detailed 1D and 2D-NMR analysis.

7.2.2 TenI

The thiazole tautomer phosphate molecule is surprisingly stable, however, did not convert by itself to the thiazole carboxylate phosphate or thizole phosphate on being in buffer at room temperature, instead it degraded to an uncharacterized product. Hence, we suggested the involvement of an enzyme TenI to specifically take up the thiazole tautomer phosphate and convert it to thiazole carboxylate phosphate. TenI was a protein found clustered with thiamin biosynthesis proteins in prokaryotes, but without an assigned role yet. Its curious similarity to the protein sequence of ThiE, the thiamin phosphate synthase, but its inability to bind thiamin phosphate had provided us adequate evidence that it is involved in the pathway at some later stage. We characterized the activity of TenI using limited amount of thiazole tautomer phosphate substrate that we could get by isolation from the *S. cerevisiae* thiazole synthase THI4. A crystal structure was then obtained for TenI bound to the product of the aromatization reaction (Ying Han, Yang Zhang, Steven E. Ealick), and basic mutational analysis studies revealed the putative involvement of a His residue in the active site for the deprotonation-protonation chemistry.

7.2.3 Reversibility of B. sublitis thiazole synthase ThiG

We were able to show that when thiazole tautomer phosphate was added back to ThiG, it disappeared, and preliminary evidence showed that a phosphorylated molecule is bound to the protein. This reversibility is yet to be explored further to assign the exact locus of attachment of the thiazole tautomer phosphate to the ThiG. We are predicting the reversible reaction to be happening at the Lys96 residue in ThiG as that is the active site residue that catalyzes thiazole biosynthesis. A Lys96Ala mutant has been used to show that the reversible reaction product, which is phosphorylated, may not be bound to this protein.

7.2.4 Thiamin pyrimidine biosynthesis

In the past, we were able to successfully reconstitute *in vitro,* the process of thiamin pyrimidine formation. With a systematic procedure in hand to produce pure and active protein, we established the fate of the C1 and C3 atoms of AIR in the complex rearrangement, the abstraction of 4'-H-atom and the 5'-H-atom from AIR by the 5'deoxyadenosyl radical, the order of abstraction of these two H-atoms and further the stereochemistry of abstraction of H from the 5'-position of the ribose ting of the AIR. We also investigated active site mutants of ThiC for interesting intermediates and activity and have established partial activity in two if the mutants.

7.3 Outlook

7.3.1 Kinetics of TenI

Due to the lack on large amounts of substrate, the kinetic characterization of TenI and its mutants have not been adequately done. Also, further mutational analysis of the protein to see whether any other residues seem to play a role in the enzymatic mechanism and change the rate of TenI could be done. The limiting factor for all further work on TenI is the synthesis of the thiazole tautomer phosphate. Attempts have been made for its synthesis, and they have been largely unsuccessful. In the future, if the synthesis is accomplished, we can have systematic and rigorous characterization of TenI and its mutants.

7.3.2 Reversibility of ThiG

The reversibility of the thiazole synthase ThiG of reacting with the unstable thiazole tautomer phosphate and producing a phosphate bound species has many interesting

consequences. Till date, a crystal structure of ThiG with any molecule bound in the active site has been elusive to us. If we can conclusively prove that the thiazole tautomer phosphate is reversibly being bound to Lys96, we may try to crystallize this protein bound with this intermediate, thus allowing us a glimpse into the active site of this protein. Also, being able to track a reversible reaction in a protein will be quite interesting in a complex 6-enzyme reconstitution. Again, obtaining the thiazle tautomer phosphate is the limiting step, but as the amount required is small, we can purify it out from *S. cerevisiae* THI4.

7.3.3 Role of the mononuclear metal center in ThiC.

We have seen the presence of a conserved mono-nuclear zinc binding site in the active site cavity in ThiC in the crystal structure. We have no clear role for this mono-nuclear Zn center or alternately, Fe-binding center in our understanding of ThiC as of now. Mutants of ThiC lacking the one or both of the two histidine residues, used to co-ordinate the metal, could be characterized to unveil the role of this metal binding site.

7.3.4 Kinetics of ThiC.

ThiC catalyzed reaction is very fast and is complete within ~30s. This makes it difficult to characterize the protein without rapid-kinetics set up. Anaerobic stopped flow and rapid quench experiments could allow kinetic characterization of this reaction, which will be valuable towards validating different mechanistic hypotheses.

7.3.5 Characterization of ThiC mutants.

7.3.5.1 Intermediates on the pathway: As discussed in Chapter 6, we have been able to identify two mutants C474S and Y277F which appear to be doing the partial reaction. We proved this 5'deoxyadenosine was not uncoupled and was relevant to the

reaction, by doing the reaction with labeled AIR substrates and showing label incorporation into the 5'deoxyadenosine. These mutants can be of high relevance in isolating an intermediate on the reaction pathway. A hypothesis we have for the intermediate that we may produce is that it may have a phosphate on it. If so, we can probe for the phosphate group by various techniques – neutral ion loss tracking of phosphate in the reaction mixture, phosphate NMR analysis, radioactively labeled AIR substrate etc., we may identify interesting intermediates on the reaction pathway for the protein.

7.3.5.2 Carbon monoxide formation in mutants: In addition, we will probe for the production of carbon monoxide for the mutants, to see the extent to which the reaction has occurred in both of the mutants.

7.3.5.3 H-abstraction from desamino-AIR: To check for the involvement of the amino group of the AIR in the rearrangement reaction, we will try the ThiC reconstitution reaction with desamino-AIR.