1-1-2012

Investigating the glutamine-trna (glutamine) synthesis appartus of the human pathogen helicobacter pylori

Nilesh Joshi
Wayne State University,

INVESTIGATING THE GLUTAMINE-TRNA (GLUTAMINE) SYNTHESIS APPARTUS OF THE HUMAN PATHOGEN *HELICOBACTER PYLORI*

by

NILESH JOSHI

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

2012

MAJOR: CHEMISTRY

Approved by:

_____

Advisor Date

ACKNOWLEDGMENTS

I came for graduate studies in United States with training in organic chemistry. Biological sciences offered a lot of promises then and an outlet to rediscover and invent much of the scientific activity of other fields. Excited by this prospect, I decided to pursue graduate studies in biochemistry. The next few years were educational beyond the bench and I am thankful to a number of people for it.

Firstly, my deepest thanks to my advisor Dr. Tamara Hendrickson. Under her direction, I went from being a biochemistry novice to developing a taste and ambition for deep research problems. I remember well her patience during my frustrations of developing an assay and the eventual joy of figuring it out. I could not have asked for a better guide for navigating the overwhelming and (sometimes distracting) jungle of biochemical research.

Thanks to Dr. Terry Cathopoulis, Dr. Keng-Ming Chang, Dr. Franklin John, Dr. Yug Verma and Megan Ehrenweth for being wonderful and welcoming seniors. Special thanks to Dr. Terry Cathopoulis and Dr. Keng-Ming Chang for their guidance in my initial stage of the lab. The support and discussions from Liangjun Zhao, Gayathri Silva, Sandamali Ekanayaka and Dilani Gigriwala have been invaluable. I enjoyed learning from all of you.

Thank you to Shirin Fatma for being a great friend, labmate and bench neighbor for many years. I will remember this time fondly. Thanks also to fellow graduate students and friends Chandra, Satish, Fatima, Rajesh, Geeta, Saptarashi, Yeajur, Elayaraja, Rama, Srinivas, Amit and Rebecca. You made my stay memorable.

Finally, none of my studies would have been possible without the continuous support from my parents and my brother. I hope to continue to make them proud as I move on to my next venture.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

Introduction

I. Introduction

Accurate protein biosynthesis is a vital process to all cellular life. Aminoacyl-tRNAs are at the heart of this process. A correctly aminoacylated tRNA fulfills two important roles in protein synthesis – First, it establishes a connection between information in the genetic code and its accurate representation as amino acids in proteins. Second, as a chemical entity, it serves as an activated monomer to facilitate peptide bond formation in the ribosome with ease.

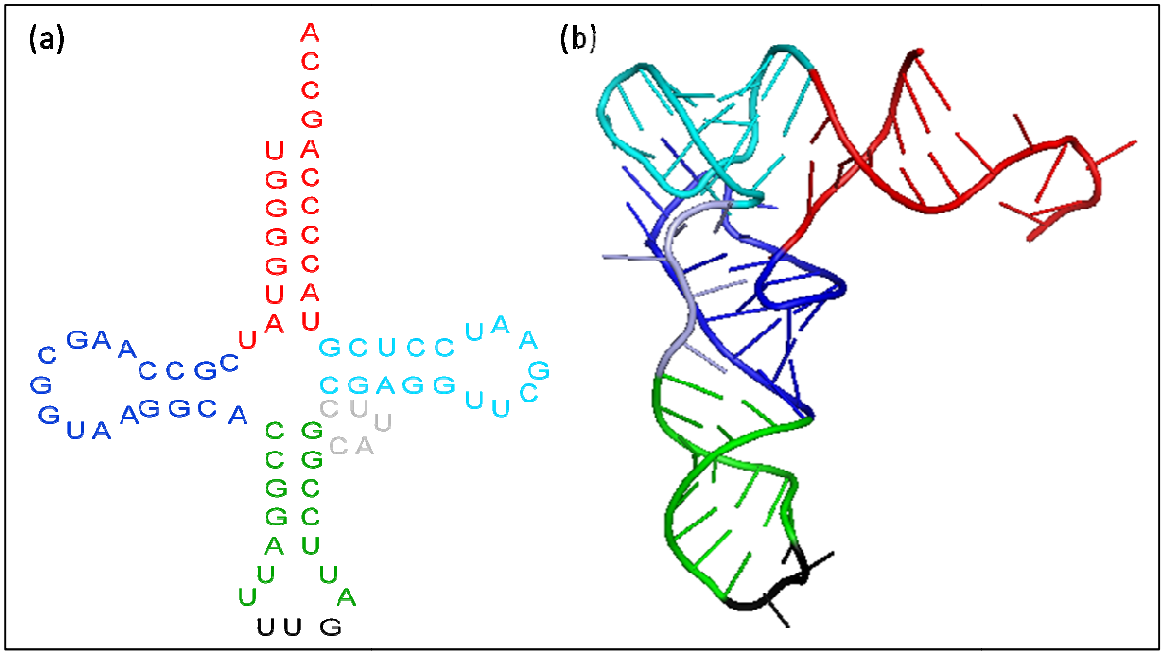
The correct synthesis of aminoacyl-tRNAs requires matching of the cognate tRNA to the cognate amino acid. There are two challenges for such a task. Organisms can have tRNA encoding genes ranging from 15(ref) to as high as 620(ref). Aminoacylation thus requires selecting the correct tRNA from a large number of tRNAs. Similarly, the cognate amino acid has to be selected from a cellular pool of 20 amino acids and many structurally similar secondary metabolites. This complex task is carried out by a family of enzymes called the aminoacyl-tRNA synthetases (aaRSs). Once correctly formed, the aminoacyl-tRNA pair is shuttled to the ribosome by EF-Tu. Mechanisms to maintain the fidelity of these processes are essential for the integrity of the genetic code. An understanding of the above machinery is thus important for understanding the protein biosynthesis pathway of organisms.

This chapter will provide a background on the agents involved in aminoacyl-tRNA synthesis and consumption. In particular, how *Helicobacter pylori*, a human pathogen, synthesizes its aminoacyl-tRNAs will be discussed. In conclusion, the chapter will

outline the dissertation research carried out on the indirect aminoacylation synthesis of Gln-tRNA^{Gln} in *H. pylori* using phylogenetic analyses and enzymatic assays.

II. Transfer RNAs

tRNAs have a distinctive secondary structure – the cloverleaf shown in Figure 1.1a, that consists of four regions. They are described below.



acceptor stem is

the anticodon in

TS (Courtesy: Dr.

The acceptor stem – This region contains both the 5' and 3' ends of the tRNA and is usually 7 bp long with four extra ribonucleotides on the 3'-end. The trinucleotides CCA are the last three for any tRNA molecule. They can be added post-transcriptionally or be a part of the encoded tRNA sequence. The identity of the fourth unpaired nucleotide varies for different tRNA molecules. This position is called the discriminator base and is important for substrate recognition by some aaRSs.[18, 19]

The D stem/loop – This region consists of 3-4 base pairs in the stem and about 8 unpaired ribonucleotides in the loop. Most uridine bases in the loop are post-transcriptionally reduced to dihydrouridine, abbreviated D. This region can also play a role in aaRS recognition depending on the length of the D stem.[14, 16] The D loop makes several contacts with the variable loop to form the stable tertiary structure of tRNA (see below).

The anticodon stem/loop – This region carries about 5 base pairs in its stem and a triplet of ribonucleotides (anticodon) in the loop that serves to decode the information in mRNA by Watson-Crick pairing with the corresponding codons. The anticodon loop is heavily modified post-transcriptionally and the nucleotides adjacent to the anticodon as well as the anticodon itself are often important identity elements for aaRSs. [6,9,18] The nucleotides in the anticodons can also be post-transcriptionally modified. For example, U34 in $tRNA^{Gln}$ is thiolated and this modification is critical for tRNA aminoacylation and specificity.[21]

The TψC stem/loop – This region has the conserved triplet UUC in its loop. Post-transcriptional modifications convert the first uridine to thymidine (T) by methylation while the second uridine is converted to psuedouridine (Ψ). Hence, this region is called the TψC region. Three base pairs in the stem are important for recognition by EF-Tu (see section 1.7 below).

Nearly all tRNAs fold into a characteristic L-shaped tertiary structure (Figure 1.1b). The anticodon is at one end in this structure while the free 3' – hydroxyl of the terminal ribose in the acceptor stem is on the other end. The two ends are separated by a distance of approximately 70 Å. The tRNA variable loop, which can have 4-20

nucleotides, forms tertiary contacts with positions in the D stem and D loop to form the hinge (or elbow region) of the tRNA. This structure leaves the anticodon triplet unpaired and available for Watson-Crick base pair formation with codons in mRNA.

A tRNA molecule typically encounters an average of 30 different enzymes after its transcription till its consumption. The selective recognition of a particular tRNA over others depends on the presence or absence of certain nucleotides in its primary sequence. The nucleotides that aid recognition are termed “determinants,” while the ones that disfavor it are called “antideterminants”. As noted above, the determinants and antideterminants for various enzymes (including aaRSs) are scattered throughout the body of tRNA. The presence of such an identity set for each protein that interacts with a particular tRNA thus constrains the primary sequence and the number of possible isoacceptors of tRNA for a particular anticodon.[18]

Considerations about identity sets for various tRNA interacting proteins play an important role in the goal of engineering microbes with desired properties. For example, constraints on the primary tRNA sequence were exploited in designing a tRNA molecule capable of introducing phosphoserine into the genetic code of *E Coli*.[22] With the accelerated pace of synthetic biology, such expanded genetic codes and their host microbes are expected to play a key role in uncovering and generating desired properties of living systems.[27]

III. Aminoacyl-tRNA synthetases (aaRSs)

Aminoacyl-tRNA synthetases synthesize aminoacyl-tRNAs by catalyzing two reactions. In the first reaction, the amino acid is activated by by ATP to form aminoacyl-

adenylate (aa-AMP, Figure 1.2). For certain aaRSs like LysRS I, GluRS, GlnRS, and ArgRS this step requires the presence of the corresponding tRNA.[1, 4] In the second step, the activated aa-AMP reacts with the 2' or 3' hydroxyl of the terminal ribose in the cognate tRNA to form the aminoacyl-tRNA (aa-tRNA).

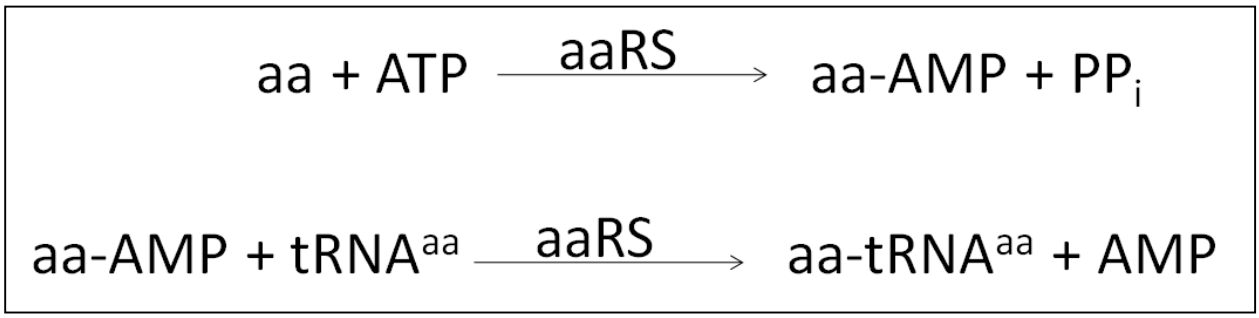


Figure 1.2 Aminoacyl-tRNA synthesis by the aaRSs. (a) The aaRS activates its cognate amino acid using ATP to form aa-AMP. (b) This aa-AMP is then reacted with the cognate tRNA to form the aa-tRNA.

Since all aaRSs catalyze the same two reactions, the sequence-structure-function dogma[11] leads to the expectation that these enzymes should share conserved features in their primary sequences, their overall structure, as well as their biochemical mode of catalysis. As will be discussed below, this is not the case. As a family of enzymes, aaRSs can thus be viewed as one of the counter-examples to the sequence-structure-function dogma, among many others.[12]

The family of aaRSs has been classified into two classes – Class I and Class II. Members of each class share sequence, structure, and mechanistic features. Subtle differences lead to a sub-division of classes; these will however not be discussed here.

Class I synthetases have conserved HIGH and KMSKS signature sequences.[1] These sequences are a part of the nucleotide-binding domain – the Rossmann fold[2] – that is part of the active site of these enzymes. The Rossman fold binds ATP in an extended form, while the HIGH and KMSKS motifs stabilize the transition state of the

first reaction – aa-AMP formation[1]. The binding of Class I aaRSs to the minor groove of the acceptor stem of the tRNA is followed by nucleophilic attack by the 2' hydroxyl of the 3' end of the tRNA towards the aa-AMP, generating the aa-tRNA[2]. The anticodon-binding domain of most Class I aaRSs is an alpha-helix bundle. The only known exception is GlnRS, here the anticodon-binding domain is made of β-sheets. Four Class I synthetases – GluRS, GlnRS, ArgRS and LysRS I – require cognate tRNA binding[1] before aa-AMP formation.

Class II aaRSs differ from their Class I counterparts on several counts. They have three conserved domains – domains I, II and III[2]. Domains II and III form the active site with an antiparallel β–fold[2]. Most Class II aaRSs are dimeric or multimeric in contrast to the monomeric Class I aaRSs. Domain I forms the dimer interface[2]. The rigid β-fold leads to ATP binding in a bent form with positioning of the amino acid for nucleophilic attack. The binding of Class II aaRSs to their cognate tRNAs is from the major groove leading to conformational changes in the anticodon loop[1]. The aminoacylation takes place on the 3' hydroxyl of the 3'end of tRNA.

IV. Expectation of ubiquity – and the revelation.

Protein biosynthesis primarily relies on twenty encoded amino acids. Consequently, it seems reasonable to expect that twenty aaRSs are required to generate the required twenty aa-tRNA pairs for protein synthesis. Thus, all life would utilize twenty aaRSs in their protein biosynthesis machinery. Contrary to this expectation, however, "*many organisms utilize a limited (<20) set of aaRSs or a non-standard set of aaRSs or tRNA aminoacylation mechanisms*". [6]

It is no longer surprising that there are exceptions to the twenty aaRSs per organism rule. Eukaryotes almost exclusively follow this rule. Many bacteria and almost all archaea, however, do not have a complete set of aaRSs. Thus, most of the life that we know does not obey this rule. The remainder of this chapter as well as the rest of this thesis will discuss the possibilities raised by this revelation.

The most ubiquitous of the missing synthetases are glutaminyl-tRNA synthetase (GlnRS) and asparaginyl-tRNA synthetase (AsnRS). These two aaRSs are missing in *Helicobacter pylori*,[7] an organism whose tRNA aminoacylation pathways are of interest to our lab. The different route taken by *H. pylori* to the products of these enzymes are discussed separately in the next section.

V. Gln-tRNA^{Gln} and Asn-tRNA^{Asn} biosynthesis in *H. pylori*

The complete genome of the human pathogen *Helicobacter pylori* revealed that genes encoding for both GlnRS and AsnRS are missing.[7] Like many other organisms that are missing these genes, *H. pylori* uses indirect pathways to synthesize Asn-tRNA^{Asn} and Gln-tRNA^{Gln}.

For Asn-tRNA^{Asn} production, a non-discriminating aspartyl-tRNA synthetase (ND-AspRS) catalyzes the formation of both Asp-tRNA^{Asp} and Asp-tRNA^{Asn} (Figure 1.3). The misacylated Asp-tRNA^{Asn} is then converted to the cognate Asn-tRNA^{Asn} by transamidation (see section VI, below).

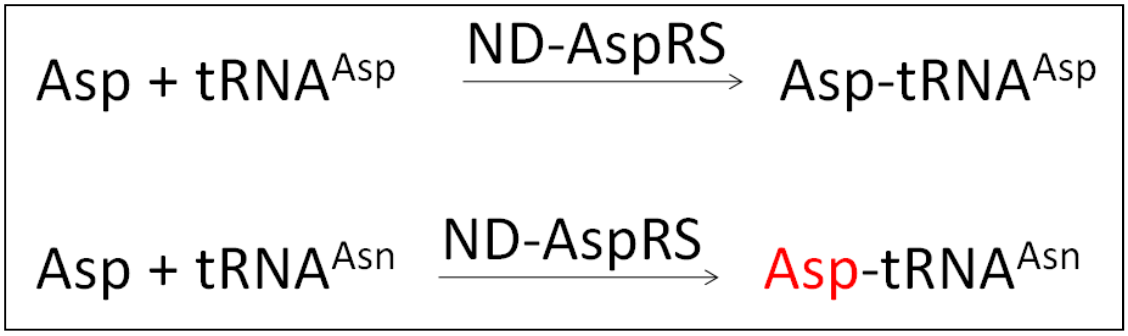


Figure 1.3 Reactions catalyzed by ND-AspRS: Both $tRNA^{Asp}$ and $tRNA^{Asn}$ are aspartylated by ND-AspRS to give the cognate Asp-$tRNA^{Asp}$ and the non-cognate Asp-$tRNA^{Asn}$ (colored in red).

Similar to ND-AspRS, organisms missing GlnRS usually utilize an analogous ND-GluRS for synthesis of Glu-$tRNA^{Glu}$ and Glu-$tRNA^{Gln}$ (Figure 1.4). Just like above, the misacylated Glu-$tRNA^{Gln}$ is converted to Gln-$tRNA^{Gln}$ by transamidation (see section VI)

Figure 1.4 Reactions catalyzed by ND-GluRS: Both $tRNA^{Glu}$ and $tRNA^{Gln}$ are glutamylated by ND-GluRS to give the cognate Glu-$tRNA^{Glu}$ and the non-cognate Glu-$tRNA^{Gln}$ (colored in red).

However, *H. pylori*, a representative of the ε-proteobacteria, takes a slightly different path to produce Gln-$tRNA^{Gln}$. *H. pylori* has two copies of the *gltX* gene, apparently generated by a gene duplication event.[9] GluRS1 catalyzes the direct aminoacylation of Glu-$tRNA^{Glu}$ and thus acts as a D-GluRS (Figure 1.5). GluRS2 (in contrast to ND-

GluRS) selectively aminoacylates $tRNA^{Gln}$ to form Glu-$tRNA^{Gln}$.[9] (Figure 1.5) The misacylated Glu-$tRNA^{Gln}$ is repaired by transamidation.

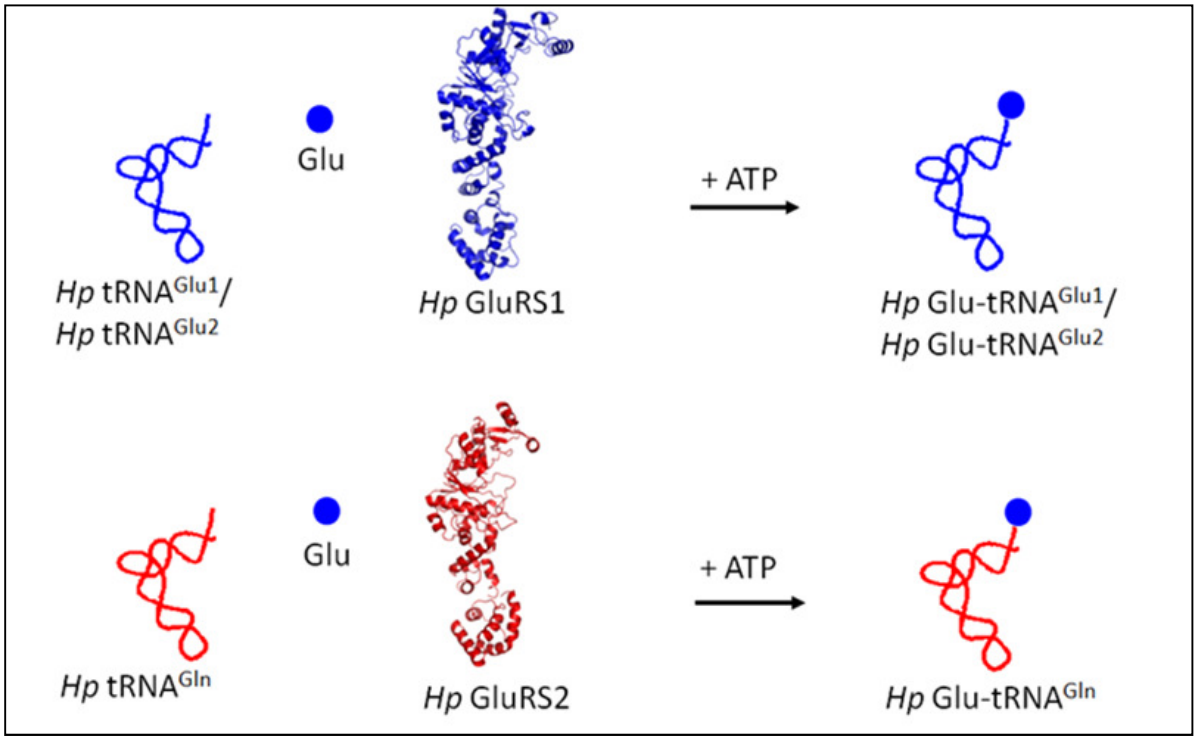


Figure 1.5 Aminoacylation of $tRNA^{Gln}$ and $tRNA^{Glu}$ in *H. pylori*: The two $tRNA^{Glu}$ isoacceptors are glutamylated by GluRS1 (blue) to form cognate Glu-$tRNA^{Glu}$. Transfer RNA^{Gln} is selectively glutamylated by GluRS2 to form the non-cognate Glu-$tRNA^{Gln}$. Protein structures are Rosetta models.[28]

As discussed above, synthetases with relaxed or non-cognate specificities generate misacylated tRNAs, like Asp-$tRNA^{Asn}$ and Glu-$tRNA^{Gln}$. In organisms like *H. pylori*, they are converted to the cognate Asn-$tRNA^{Asn}$ and Gln-$tRNA^{Gln}$ by transamidation. This reaction catalyzed by an amidotransferase is discussed below. For simplicity, only Gln-$tRNA^{Gln}$ synthesis is discussed. Similar discussion applies to Asn-$tRNA^{Asn}$ synthesis.

VI. The amidotransferase GatCAB in *H. pylori*

As discussed above, the misincorporation of Glu-tRNA^{Gln} is prevented by its conversion into Gln-tRNA^{Gln} by a heterotrimeric protein complex called GatCAB (Figure 1.6B).[8]

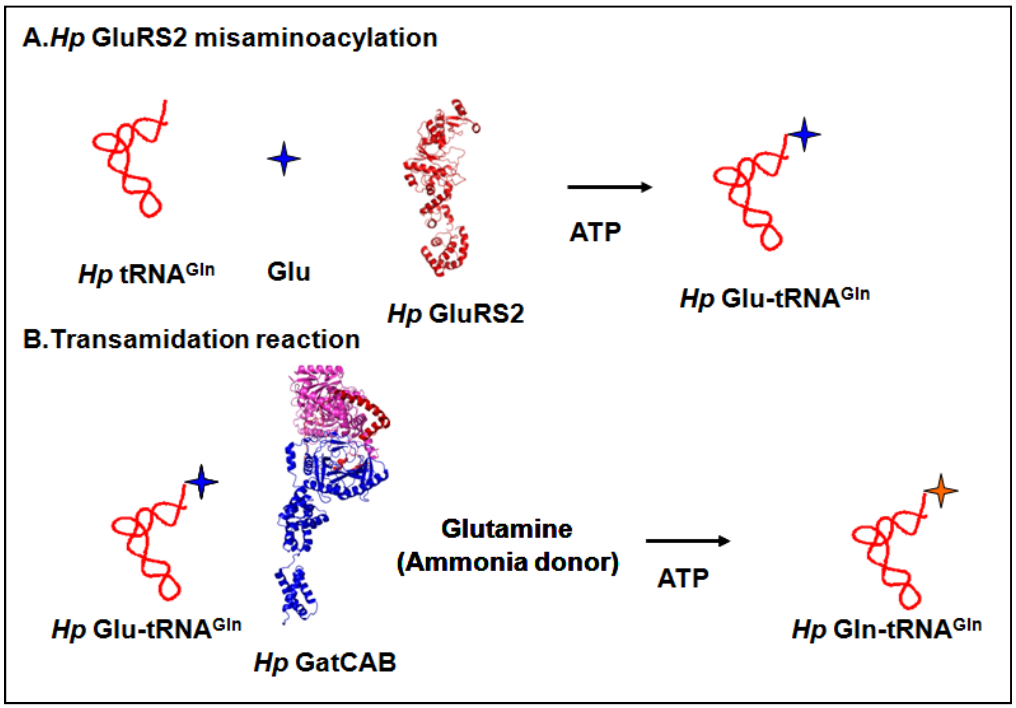


Figure 1.6: Indirect aminoacylation pathway for *H. pylori* Gln-tRNA^{Gln}. A: *Hp* GluRS2 glutamylates tRNA^{Gln} to generate Glu-tRNA^{Gln}. **B**. *H. pylori* GatCAB rescues this misacylation by a transamidation reaction, using glutamine as the ammonia donor, to convert Glu-tRNA^{Gln} into Gln-tRNA^{Gln}.

GatCAB is a heterotrimeric amidotransferase consisting of 3 subunits – GatC, GatA, and GatB.[10] GatCAB catalyzes three reactions: Glutaminase (glutamine hydrolysis, catalyzed by GatA); Phosphorylation of Glu-tRNA^{Gln} (catalyzed by GatB), and transamidation (transport of ammonia from GatA to GatB and its reaction with the phosphorylated intermediate to generate Gln-tRNA^{Gln}, catalyzed by GatB). The role of GatC is not known. The ammonia generated by glutamine hydrolysis in GatA remains

associated with the enzyme and is delivered through a tunnel to the active site of GatB. There, the ammonia nucleophilically attacks the phosphorylated amino acid to generate Gln-tRNA^{Gln} (Summarized in Figure 1.7).[6]

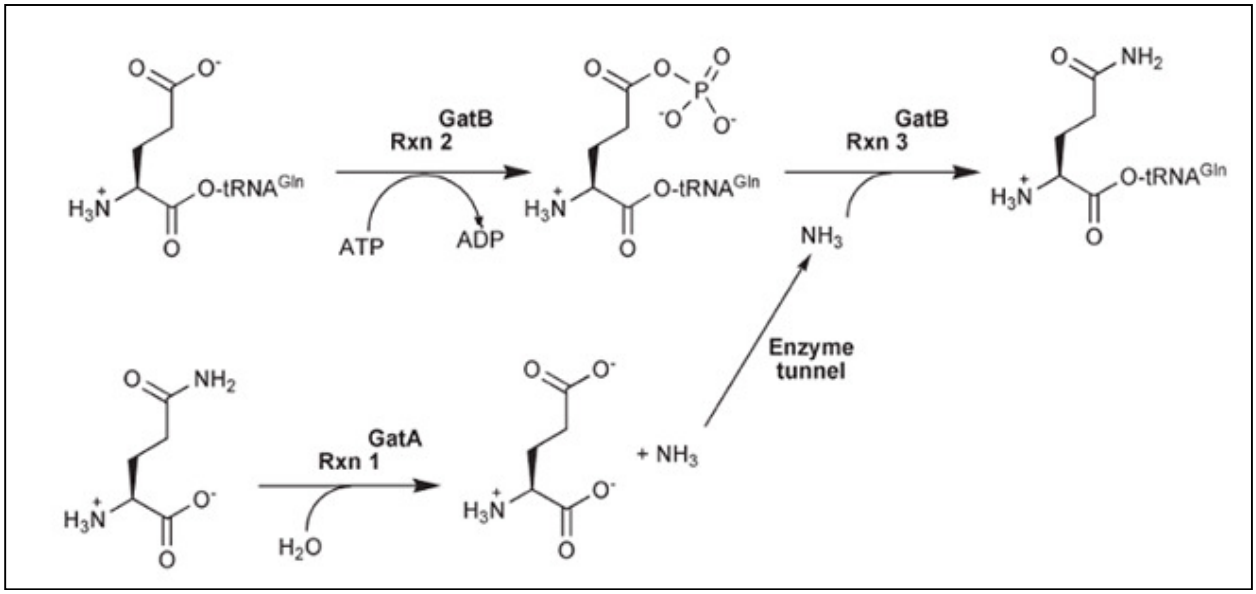


Figure 1.7: Reactions catalyzed by GatCAB:[6] **Rxn 1**: GatA catalyzes the hydrolysis of glutamine to generate ammonia. **Rxn 2**: GatB catalyzes the phosphorylation of Glu-tRNA^{Gln} to form a mixed anhydride. **Rxn 3**: The ammonia from GatA is transported to the active site of GatB where it acts as a nucleophile to generate Gln-tRNA^{Gln}. Figure from reference 6.

Such a rescue of misacylation is not unique to bacteria. In archaea, transamidation is performed by GatDE[23] while in organellar systems like mitochondria, etc (which are thought to be bacterial in origin) the heterotrimeric GatFAB performs the transamidation[24].

VII. Elongation factor (EF-Tu)

Generation of a correctly aminoacylated tRNA is closely followed by its consumption in the ribosome for protein synthesis. These two processes are bridged by elongation

factor Tu (EF-Tu). EF-Tu, in its GTP bound form, shuttles the correctly formed aa-tRNA to the ribosome. Here, hydrolysis of the bound GTP by EF-Tu releases the aa-tRNA for protein biosynthesis. [25]

Elegant studies have shown that EF-Tu recognizes the correct pair by "thermodynamic compensation."[26] Briefly, the binding affinity of a given aa-tRNA^{aa} to EF-Tu consists of the combined contributions of the amino acid and the first three base pairs in the TΨC stem of the tRNA. These contributions follow a bell-shaped distribution. An amino acid with a higher affinity for EF-Tu is matched with a tRNA with a lower affinity and vice-versa. The net effect is the uniform binding affinity of Ef-Tu to aa-tRNAs.

The Uhlenbeck model is thus another checkpoint in accurate protein biosynthesis. The power of such a model is that it allows engineering the three triplets so as to influence the binding of EF-Tu to a (cognate or non-cognate) tRNA pair.[22]

VIII. Dissertation research

This dissertation focuses on indirect aminoacylation and transamidation to produce Gln-tRNA^{Gln} in *H. pylori*. A combination of phylogenetic analyses and enzymatic assays were used that contribute to a picture of this process.

Chapter 2 discusses our examination of sequence conservation of *gltX2* (the gene that encodes GluRS2) across *different H. pylori* strains. The conclusions of this sequencing effort are compared with trends in sequences of genes important in the aminoacylation step of protein synthesis.

Chapter 3 discusses one aspect of Gln-tRNA^{Gln} production – the mechanism of transport of ammonia from the active site of one subunit (GatA) to the active site of another (GatB) through an intramolecular hydrophilic tunnel. Site-directed mutagenesis of key residues lining this tunnel and their transamidation assays were performed. With these studies, a preliminary picture of ammonia transport through the tunnel can be constructed.

IX. References

1. Ibba, M., and Söll D. (**2000**) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.***69**, 617-650.
2. Eriani, G., *et al.*(1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*. **347**(6289): 203-206.
3. Cusack, S. (**1997**) Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.***7**, 881-889.
4. Woese, C. R., Olsen, G. J., Ibba, M., and Söll, D. (**2000**) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Mol. Biol. Rev.***64**, 202-236.
5. Bult C.J., *et al.* (**1996**) The complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*. **273**, 1058-1073.
6. Cathopoulis, T., Chuawong, P., and Hendrickson, T. L. (**2007**) Novel tRNA aminoacylation mechanisms. *Mol. Biosyst.***3** (6):408-18.
7. Tomb, J.F., *et al*. (**1997**) The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* **388**(6642):539-47.
8. Tumbula, D. L., Becker, H. D., Chang, W., Söll, D. (**2000**) Domain-specific recruitment of amide amino acids for protein synthesis *Nature.***407**, 106.

9. Skouloubris, S., Ribas de Pouplana, L., De Reuse, H., Hendrickson, T.L. (**2003**) A noncognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proc. Natl. Acad. Sci. U S A.***100** (20):11297-302.

10. Nakamura, A., Yao, M., Chimnaronk, S., Sakai, N., Tanaka, I. (**2006**) Ammonia channel couples glutaminase with transamidase reactions in GatCAB. *Science.***312** (5782):1954-1958.

11. Anfinsen, C. B. (**1973**).Principles that govern the folding of protein chains. *Science* **181**,223–-230.

12. Alexander PA, *et al*. (**2007**). The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc Natl Acad Sci USA* **104**:11963–11968.

13. Zhao, L., *et al.* (**2012**) The kinase activity of *H. Pylori* Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase is sensitive to distal mutations in its putative ammonia tunnel *Biochemistry*. **51** (1):273-285.

14. Salazar, J.C, *et al*. (**2003**) Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proc. Natl. Acad. Sci. USA* **100**, 13863-13868.

15. Wu, J., Bu, W., Sheppard, K., Kitabatake, M., Kwon, S. T., Soll, D., Smith, J. L.(**2009**) Insights into tRNA-dependent amidotransferase evolution and catalysis from the structure of the *Aquifex aeolicus* enzyme. *J. Mol. Biol.***391**, 703-716.

16. Sekine,S., *et al*. (**1996**) Major identity determinants in the ‘‘augmented D helix’’ of tRNA(Glu) from *Escherichia coli. J. Mol. Biol*., **256**, 685–700.

17. Curnow, *et al*. (**1997**) Glu-tRNAGln amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation *Proc. Natl.Acad.Sci.USA* 94, 11819-11826.

18. Giege,R., Sissler,M. and Florentz,C. (**1998**) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**,5017–5035.

19. Crothers,D.M., Seno,T. and Soll,G. (**1972**) Is there a discriminator site in transfer RNA? *Proc. Natl Acad. Sci. USA*, **69**, 3063–3067.

20. Raushel FM, Thoden JB, Holden HM. (**2003**) Enzymes with molecular tunnels. *Acc Chem Res.***36** (7):539-4.

21. Madore,E., *et al*. (**1999**) Effect of modified nucleotides on *Escherichia coli* $tRNA^{Glu}$ structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. *Eur. J. Biochem.*, **266**, 1128–1135.

22. Hee-Sung Park, *et al*. (**2011**) Expanding the genetic code of *Escherichia coli* with phosphoserine *Science* **333**, 1151-1154.

23. Hiroyuki Oshikane, *et al.*(**2006**) Structural basis of RNA-dependent recruitment of glutamine to the genetic code *Science* **312**, 1950.

24. Mathieu Frechin, *et al.* (**2009**) Yeast mitochondrial Gln-tRNAGln is generated by a GatFAB-mediated transamidation pathway involving Arc1p-controlled subcellular sorting of cytosolic GluRS *Genes Dev* **23**, 1119-1130.

25. T. Dale & O. L. Uhlenbeck, (**2005)** Amino acid specificity in translation *Trends Biochem. Sci.* **30**, 659-665.

26. Frederick J. LaRiviere, *et al.* (**2001**) Uniform Binding of Aminoacyl-tRNAs to Elongation Factor Tu by Thermodynamic Compensation *Science* **294**, 165.

27. Ahmad S. Khalil, *et al.* (**2010**) Synthetic biology: Applications come of age *Nat. Rev. Gen.* **11**, 367-379.

28. http://robetta.bakerlab.org/ Protein structure prediction server.

29. Hatfield, D.L., *et al*. (**1992**) Transfer RNA in protein synthesis. *CRC press.*

Chapter II

Sequencing *gltX2* genes in Hp strains

I. Introduction

It is evident from genomic analyses that known glutaminyl-tRNA synthetases (GlnRS) originated in eukarya. Some bacteria do possess a GlnRS (*e.g.E.coli*), however these enzymes are products of horizontal gene transfer from eukarya.[1]To date, no GlnRS has been found that originated in archaea or in most bacteria.

The hypothesis, therefore, for the evolution of GlnRS is as follows: An ancient non-discriminating glutamyl-tRNA synthetase (ND-GluRS) gene underwent a gene duplication event to give two ND-GluRSs. After branching from archaea, the eukarya were able to evolve these into a D-GluRS and D-GlnRS. A few bacteria acquired this D-GlnRS through horizontal gene transfer, as mentioned above. Many archaea and bacteria continue to utilize a ND-GluRS for production of Glu-tRNA^{Glu} and Gln-tRNA^{Gln} by direct and indirect aminoacylation pathways, respectively.[2]

A few bacteria, including *Helicobacter pylori (Hp)*, have retained duplicate copies of *the gltX* gene. In *H. pylori,* GluRS1 is a discriminating GluRS, while GluRS2 has been proposed to be a missing evolutionary link between an ND-GluRS and a forthcoming D-GlnRS of bacterial origin.[3]

Because *Hp* is a highly adaptive organism, it has a high rate of evolutionary variation between strains. In fact, strains of *H. pylori* that have been isolated from different regions of the world differ substantially in infectivity and are genomically varied.[4] This fact provides us with a unique opportunity to study the possible evolution of GluRS2 into a bacterial GlnRS.

Previously, Dr. Terry Cathopoulis had cloned (into the TOPO pCR2.1 vector, Invitrogen) and sequenced 16 GluRS2 genes from genomic DNA of different *Hp* strains (These strains were chosen because they are representative of the geographic distribution of *Hp*).[5] Not surprisingly, most of these genes were highly homologous to the *gltX2* gene from Hp26695, the strain used for our original characterization of GluRS1 and GluRS2.[3] However, two sequences were potentially interesting. The first of these is called as the Cheetah strain, a strain of *Hp* that infects cheetahs. The Cheetah *gltX2* gene contained a premature stop codon leading to a predicted open reading frame thatencoded a truncated copy of GluRS2 (only the first 150 of GluRS2's 450 amino acids were encoded). The same gene in another *Hp* strain called R7 also contained a premature stop codon and was predicted to encode protein 394 amino acids in length. Finally, adequate sequencing data for the*gltX2* gene from the B3 strain was not obtained. Because these results were intriguing and were the result of only one evaluation, the goal of this aim was to revisit and hopefully confirm the Cheetah and R7 truncations and to complete the evaluation of the *gltX2* gene from strain B3.

II. Results and Discussion

The *gltX2* gene was amplified from the B3 *Hp* strain using primers NJ201 and NJ202 and the polymerase chain reaction (PCR). The products were cloned into the TOPO pCR2.1 vector and sequenced. The alignment of B3 with Hp 26695 is shown in Figure 1A. The two genes were highly homologous and so further evaluations of B3 were deemed unnecessary.

Figure 2.1: **Alignment of DNA sequences for the *gltX* genes from *Hp* strain 26695 *versus* B3.**

```
26695  ATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCACATAGGGAATTTAAGGGCAGCC 60
B3     ATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCACATAGGGAATTTAAGGGCAGCC 60
       ************************************************************

26695  ATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCCTTTCTCATTCGCATTGAAGAC 120
B3     ATTTTCAATTACATTGTGGCCAAACAGCAACACAAACCCTTTCTCATTCGCATTGAAGAC 120
       ******** *********** ********* * ***************************

26695  ACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAGATTTTAGAAATTTTAAAGCTT 180
B3     ACGGACAAAGAGCGCAACGTTGAAGGCAAAGACCAAGAGATTTTAGAAATTTTAAAGCTT 180
       ** *************** ****************************************

26695  ATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCATAATATAGATTACCACAGAGAA 240
B3     ATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCATAACATAGATTATCACAGAGAA 240
       ****************************************** ******** *********

26695  ATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTATTGTTATGCGAGTGCGGAGTTT 300
B3     ATGGCAGAAAAATTACTGAAAGAACATAAAGCGTTTTATTGTTATGCGAGCGCGGAGTTT 300
       ************************ ************************ *********

26695  TTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGCCCTTTCAGGTATTCAGACGAG 360
B3     TTAGAAAGAGAAAAAGAAAAAGCCAAAAACGAGAAACGCCCTTTCAGGTATTTAGACGAG 360
       ****************************** ** ******************** *******

26695  TGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTGGTGCGTTTAAAAGCCCCAAAT 420
B3     TGGGCCACTTTAGAAAAAGACAAGCATCATGCCCCTGTGGTGCGTTTAAAAGCCCCAAAT 420
       ************************** **********************************

26695  CATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTGAAATTTGAACCTGATGAATTG 480
B3     CATGCGGTGTCTTTCAATGATGCGATTAAGAAAGAAGTGGAATTTGAGCCTGATGAATTG 480
       ***************** ********** ********* ******* ************

26695  GATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACTTATAATTTCGCTTGCGCATGC 540
B3     GATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACTTATAATTTCGCTTGCGCATGC 540
       ************************************************************

26695  GATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGAAGATCATGTGAGTAACACC 600
B3     GATGATTTGCTCTATGAAATCAGTCTGATTATTAGAGGCGAAGATCATGTGAGTAACACC 600
       *********** *** ******************************************

26695  CCCAAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAATGATCCGATTGTTTATGCGCAT 660
B3     CCTAAACAAATCTTAATCCAGCAAGCTTTAGGCTCAAACGATCCGATTGTTTATGCGCAT 660
       ** ******************************** ** ********************

26695  TTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAGTAAAAGAGATGAAGCCTCC 720
B3     TTACCCATTATTTTAGATGAAACAAGCGGTAAAAAAATGAGCAAAAGAGACGAAGCCTCC 720
       ** ******************  ************ ***** ******** *********

26695  AGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGCGATTGCGAATTACCTCATC 780
B3     AGCGTGAAATGGCTTTTGAATCAAGGGTTTTTGCCGGTTGCGATTGCAAATTACCTCATC 780
       ******************************** ************** ************

26695  ACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGTTT 840
B3     ACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGTTT 840
       ************************************************************

26695  AGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTTAAAATATTTAAAACACTTA 900
B3     AGTTTAGAAAATCTTTCCAGTTCCCCGGCTCATTTTAATTTAAAATATTTAAAACACTTA 900
       ************************ ************************************



26695  AACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATTAGAACTCACTTCAATAAAA 960
B3     AACCACGAGCATTTAAAGCTTTTAGACGATGAAAAGTTATTAGAACTCACTTCAATAAAA 960
       ******************************** ****************************

26695  GATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGAATGCGGCACGCTTTTAGAA 1020
```

```
B3      GATAAAAACCTCCTAGGGCTTTTAAGATTGTTTGTAGAAGAATGCGGTACGCTTTTAGAA 1020
        ************ ******************** ************* ************

26695   TTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAAAAT 1080
B3      TTGAAAGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAGAAC 1080
        ****  ************************************************* **

26695   GAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCTAACAAGCATGGATTTTCAA 1140
B3      GAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCCCTAAAAGGCATGGATTTTCAA 1140
        ************************************ **** * **************

26695   GCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGCGATTAAGCCAGCTTAAGGGT 1200
B3      GCGTATAAGGATTTTGAAAGCTTTAAAAAAGAAGCCATGCGATTAAGTCAGCTTAAGGGT 1200
        ******************** ************************ ************

26695   AAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAACTCGCATGGCGTTGAATTG 1260
B3      AAGGATTTTTTCAAACCTTTGCGCATTCTTTTAACCGGGAACTCGCATGGCATTGAATTG 1260
        ************************** ********************** ********

26695   CCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGTTTTAAGGCTCAAAGCATGA 1320
B3      CCTTTGATTTTCCCCTATATCCAAAGCCACTATCAAGAAGTTTTAAGGCTCAAAGCATGA 1320
        ****************************  ****************************
```

The *gltX2* gene wasPCR amplified from the Cheetah *Hp* strain using primers NJ201 and NJ202. The products were cloned into the TOPO pCR2.1 vector and sequenced. Surprisingly, the sequences showed a considerable number of mismatches compared to the previous data reported by Dr. Cathopoulis. In order to get an unambiguous sequence, we repeated the above experiment using primers S3 and S4 (the ones originally used by Dr. Cathopoulis).

The sequences from the latter effort matched our own results and continued to show mismatches with the preliminary data. With one exception (below), our data aligned with very good agreement to the *gltX2* gene from*Hp* 26695.These results are shown in Figures 2.2 and 2.3. Based on a 3:1 occurrence, we have concluded that *gltX2* gene from Cheetah is not truncated as originally suspected.

Figure 2: **Alignment of sequencing data using forward primers for the strain Cheetah *gltx2*.** NJ; S3 and S4 – data from different sequencing trials. TC – Data from Dr. Terry Cathopulis. 26695 – Data from *H pylori* genome.

```
NJ      ---NGNGANTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGC
S3 & S4 NNGGGCGANTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGC
26695   ------------------------------------------------------------
TC      ------------------------------------------------------------
```

```
NJ       AGAATTCGCCCTTCGGGATCCATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCAC
S3 & S4  AGAATTCGCCCTTCGGGATCCATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCAC
26695    ----------------------ATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCAC
TC       ----------------------ATGCTTCGTTTTGCGCCTTCGCCTACTGGGGATATGCAC
                               ************************** ************

NJ       ATAGGGAATTTAAGGGCAGCCATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCC
S3 & S4  ATAGGGAATTTAAGGGCAGCCATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCC
26695    ATAGGGAATTTAAGGGCAGCCATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCC
TC       ATAGGGAATTTAAGGGCAGCCATTTTTAACTATATTGTGGCTAAACAGCAACATAAACCC
         ************************** ***** ****************** ********

NJ       TTTCTCATTCGCATTGAAGACACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAG
S3 & S4  TTTCTCATTCGCATTGAAGACACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAG
26695    TTTCTCATTCGCATTGAAGACACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAG
TC       TTTCTCATTCGCATTGAAGACACAGATAAAGAGCGCAATATTGAAGGCAAAGATCAGGAG
         ************************** *********** ************** ** ***

NJ       ATTTTAGAAATTTTAAAGCTTATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCAT
S3 & S4  ATTTTAGAAATTTTAAAGCTTATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCAT
26695    ATTTTAGAAATTTTAAAGCTTATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCAT
TC       ATTTTAGAGATTCTAAAGCTCATGGGAATGAACTGGGATAAACTCGTGTATCAAAGCCAT
         ******** *** ******* ***** ** * ****** ** ******************

NJ       AATATAGATTACCACAGAGAAATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTAT
S3 & S4  AATATAGATTACCACAGAGAAATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTAT
26695    AATATAGATTACCACAGAGAAATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTAT
TC       AACATAGATTACCATAGGGAAATGGCAGAAAAATTGCTTAAAGAAAATAAGGCGTTTTAT
         ** ************ ** ***************** ** *********** *********

NJ       TGTTATGCGAGTGCGGAGTTTTTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGC
S3 & S4  TGTTATGCGAGTGCGGAGTTTTTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGC
26695    TGTTATGCGAGTGCGGAGTTTTTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGC
TC       TGTTATGCGAGTGTGGGATTTTTAGAACAAGAAAAAGAAAAAGCCAAAAACGAAAAACGC
         ************** **  *********  ********************** *********

NJ       CCTTTCAGGTATTCAGACGAGTGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTG
S3 & S4  CCTTTCAGGTATTCAGACGAGTGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTG
26695    CCTTTCAGGTATTCAGACGAGTGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTG
TC       CCTTTCAGGTATTTAGATGAATGGGCGGCTTTAGAGAAAAACCAGCACAATACCCCTGTG
         ************** *** ** *****  ******* *** ** ***** ** ********

NJ       GTGCGTTTAAAAGCCCCAAATCATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTG
S3 & S4  GTGCGTTTAAAAGCCCCAAATCATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTG
26695    GTGCGTTTAAAAGCCCCAAATCATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTG
TC       GTGCGTTTAAA-GCCCCAAATCATGCGGTGTCTTTTAACGATGCGATTAAAAAAGAAGTG
         *********** ********************** ************************



NJ       AAATTTGAACCTGATGAATTGGATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACT
S3 & S4  AAATTTGAACCTGATGAATTGGATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACT
26695    AAATTTGAACCTGATGAATTGGATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACT
TC       AAATTTGAGCCTTATGAATTGGATTCTTTTGTGCTTTTAAGAAAGGATAAGAGCCCTACT
         ******** *** ************************ *** ******* *********

NJ       TATAATTTCGCTTGCGCATGCGATGATTTGCTTTATAAAATCAGTCTGATTATTTGCGGC
S3 & S4  TATAATTTCGCTTGCGCATGCGATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGC
26695    TATAATTTCGCTTGCGCATGCGATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGC
TC       TATAATTTCGCTTGTGCATGCGATGATTTGCTTTATGAAATCAGTCTTATTATTAGGGGC
         ************** ******************** ********** ****** * ***

NJ       GAAGATCATGTGAGTAACACCCCCAAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAAT
S3 & S4  GAAGATCATGTGAGTAACACCCCCAAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAAT
26695    GAAGATCATGTGAGTAACACCCCCAAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAAT
TC       GAAGATCATGTGAGTAACACCCCTAAACAAATCTTAATCCAACAAGCTTTAGGCTCCAAC
         ********************** ***************** *****************

NJ       GATCCGATTGTTTATGCGCATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATG
S3 & S4  GATCCGATTGTTTATGCGCATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATG
```

```
26695   GATCCGATTGTTTATGCGCATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATG
TC      AACCCTATTATTTATGCACATTTGCCCATTATCTTAGATGAAGCAAGCGGTAAAAAGATG
* ** *** ******* ************** ********** ****************

NJ      AGTAAAAGAGATGAAGCCTCCAGCGTGAAATGGCTTTTGAATCNANGGGTTTTTACCGGT
S3 & S4 ANTAAAAGAGATGAAGCCTCCAGCGTGAAATGGCTTTTGAATCNANGGGTTTTTACCGGT
26695   AGTAAAAGAGATGAAGCCTCCAGCGTGAAATGGCTTTTGAATCAAGGG-TTTTTACCGGT
TC      AGCAAAAGGGATGAAGCCTCTAGCGTGAAATGGCTTTTAAATCAAGGG-TTTTTGCCGGT
        *  ***** *********** ***************** **** * ** ***** *****

NJ      TGCGATTGCGAATTACCTCATCACTATCGGTAATAAAGTGCCNNANGGAAGTTTTTAGCC
S3 & S4 TGCGATTGCGAATTACCTCATCACTATCGGTAATAAAGTGCCTAANN-AAGTTTTTAGCC
26695   TGCGATTGCGAATTACCTCATCACTATCGGTAATAAAGTGCCTAAGG-AAGTTTTTAGCC
TC      TGCTATTGTGAATTACCTCATCACTATTGGTAATAAAGTGCCTAAAG-AAGTTTTTAGCC
        *** **** ****************** **************  *   ************

NJ      TTGATGAN-CGATAGAATGGNTTANTTTAGAAAATCTTTNCAGNNN--CNGNNTCNTTTN
S3 & S4 NTGATGAAGCGATAGAATGGTTTANTTTAGAAAATCTTTNCA------------------
26695   TTGATGAAGCGATAGAATGGTTTAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTA
TC      TTGATGAAGCGTTAGAATGGTTTAGTTTAGAAAACCTTTCTAATTCCCCAGCTCATTTTA
        ******  ** ******** *** ********* ****  *

NJ      ANTTAAANNNTT--AAANNCTTNANCNNCGAGCATTTNANGCTTTTAGACNANGA-----
S3 & S4 ------------------------------------------------------------
26695   ATTTAAAATATTTAAAACACTTAAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGT
TC      ATTTAAAATATTTAAAACACTTAAACCACCAGCATTTAAAGCGTTTAGATGATGAAAAAT


NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   TATTAGAACTCACTTCAATAAAAGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAG
TC      TATTAGAGCTTTCTCAAATAAAAGATAGGAATCTTTTAGGGCTTTTAAGATTATTCATAG


NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   AAGAATGCGGCACGCTTTTAGAATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGG
TC      AAGAATGCGATACGCTTTTAGAATTGAAAGAAAAAATTTCGTTGTTTTTAGAGCCAAAAG


NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   ATATTGTTAAAACTTATGAAAATGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACG
TC      ATATTGTTAAAACTTATGAAAACGAAGATTTTAAAGAGCGCTGCTCAATACTTTTTAACG


NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   CTCTAACAAGCATGGATTTTCAAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCA
TC      CCCTAAAAAGCATGGATTTTCAAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCA


NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   TGCGATTAAGCCAGCTTAAGGGTAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCG
TC      TGTGATTAAGCCAGCTTAAAGGTAAAGATTTTTTTAAACCTTTGCGCATTCTTTTAATTG


NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   GGAACTCGCATGGCGTTGAATTGCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAG
TC      GGGATTCGCATGGCGTTGAATTGCCTTTGATTTTCCCTTATATTCAAAGCCATTATCAAG


NJ      ------------------------
S3 & S4 ------------------------
26695   AAGTTTTAAGGCTCAAAGCATGA
TC      AAGTTTTAAGGCTCAAAGCATGA
```

Figure 2.3: **Alignment of sequencing data using reverse primers for the strain Cheetah *gltx2*.** NJ; S3 and S4 – data from different sequencing trials. TC – Data from Dr. Terry Cathopulis. 26695 – Data from *H pylori* genome

```
NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   ATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCACATAGGGAATTTAAGGGCAGCC
TC      ATGCTTCGTTTTGCGCCTTCGCCTACTGGGGATATGCACATAGGGAATTTAAGGGCAGCC

NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   ATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCCTTTCTCATTCGCATTGAAGAC
TC      ATTTTTAACTATATTGTGGCTAAACAGCAACATAAACCCTTTCTCATTCGCATTGAAGAC

NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   ACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAGATTTTAGAAATTTTAAAGCTT
TC      ACAGATAAAGAGCGCAATATTGAAGGCAAAGATCAGGAGATTTTAGAGATTCTAAAGCTC

NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   ATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCATAATATAGATTACCACAGAGAA
TC      ATGGGAATGAACTGGGATAAACTCGTGTATCAAAGCCATAACATAGATTACCATAGGGAA

NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   ATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTATTGTTATGCGAGTGCGGAGTTT
TC      ATGGCAGAAAAATTGCTTAAAGAAAATAAGGCGTTTTATTGTTATGCGAGTGTGGGATTT

NJ      ------------------------------------------------------------
S3 & S4 ------------------------------------------------------------
26695   TTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGCCCTTTCAGGTATTCAGACGAG
TC      TTAGAACAAGAAAAAGAAAAAGCCAAAAACGAAAAACGCCCTTTCAGGTATTTAGATGAA

NJ      ------------------------------------GNGN-GNGNTNAAAAGCCCCAAAT
S3 & S4 ------------------------NNCCATGCCCCNGNGN-GCGTTNAAAAGCCCCAAAT
26695   TGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTGGTGCGTTTAAAAGCCCCAAAT
TC      TGGGCGGCTTTAGAGAAAAACCAGCACAATACCCCTGTGGTGCGTTTAAA-GCCCCAAAT
        * *  * * * *** *********

NJ      CATGNNNNNTNTTT-NNNGATGCGATTAAAAAG-AAGTGAAATTTGAANCNGATGAATNG
S3 & S4 CNTGCGNNGTCTTC-AACGATGCGATTAAAAANGAAGTGAAATTTGANNNG-ATGAATNG
26695   CATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTGAAATTTGAACCTGATGAATTG
TC      CATGCGGTGTCTTTTAACGATGCGATTAAAAAAGAAGTGAAATTTGAGCCTTATGAATTG
        * **     * **     **************  *************     ****** *

NJ      GATTCTTTTGNGCTTTTGAGACAGGATAAAAGCCCTACTTATAATTTCGCTTGCGCATGC
S3 & S4 GATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCNNACTTATAATTTCGCTTGCGCATGC
26695   GATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACTTATAATTTCGCTTGCGCATGC
TC      GATTCTTTTGTGCTTTTAAGAAAGGATAAGAGCCCTACTTATAATTTCGCTTGTGCATGC
        ********** ****** *** ******* ****  ***************** ******

NJ      GATGATTTGCTTTATAAAATCAGTCTGATTATTTGCGGCGAAGATCATGTGAGTAACACC
S3 & S4 GATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGAAGATCATGTGAGTAACACC
26695   GATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGAAGATCATGTGAGTAACACC
TC      GATGATTTGCTTTATGAAATCAGTCTTATTATTAGGGGCGAAGATCATGTGAGTAACACC
        *************** ********** ****** * ************************

NJ      CCCCAAANCAANTCTTAATCCAGCAAGCTTTAGGCTCCAATGATCCGATTGTTTATGCGC
S3 & S4 CCCCAAAC-AANTCTTANTCCAGCAAGCTTTAGGCTCCAATGATCCGATTGTTTATGCGC
26695   CCC-AAAC-AAATCTTAATCCAGCAAGCTTTAGGCTCCAATGATCCGATTGTTTATGCGC
```

```
TC      CCT-AAAC-AAATCTTAATCCAACAAGCTTTAGGCTCCAACAACCCTATTATTTATGCAC
        **  ***  ** ***** **** *****************  * ** *** ******* *

NJ      ATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAGTAAAAGAGATGAAGCCT
S3 & S4 ATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAGTAAAAGAGATGAAGCCT
26695   ATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAGTAAAAGAGATGAAGCCT
TC      ATTTGCCCATTATCTTAGATGAAGCAAGCGGTAAAAAGATGAGCAAAAGGGATGAAGCCT
        ************** ********** ****************** ***** **********

NJ      CCAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGCGATTGCGAATTACCTCA
S3 & S4 CCAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGCGATTGCGAATTACCTCA
26695   CCAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGCGATTGCGAATTACCTCA
TC      CTAGCGTGAAATGGCTTTTAAATCAAGGGTTTTTGCCGGTTGCTATTGTGAATTACCTCA
        * ***************** ************** ******** **** ***********

NJ      TCACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGT
S3 & S4 TCACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGT
26695   TCACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGT
TC      TCACTATTGGTAATAAAGTGCCTAAAGAAGTTTTTAGCCTTGATGAAGCGTTAGAATGGT
        ******* ***************** ************************ *********

NJ      TTAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTTAAAATATTTAAAACACT
S3 & S4 TTAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTTAAAATATTTAAAACACT
26695   TTAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTTAAAATATTTAAAACACT
TC      TTAGTTTAGAAAACCTTTCTAATTCCCCAGCTCATTTTAATTTAAAATATTTAAAACACT
        ************** ***** * *** ** ********************************

NJ      TAAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATTAGAACTCACTTCAATAA
S3 & S4 TAAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATTAGAACTCACTTCAATAA
26695   TAAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATTAGAACTCACTTCAATAA
TC      TAAACCACCAGCATTTAAAGCGTTTAGATGATGAAAAATTATTAGAGCTTTCTCAAATAA
        ******** ************ ****** ***** ** ******** **  **  *****

NJ      AAGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGAATGCGGCACGCTTTTAG
S3 & S4 AAGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGAATGCGGCACGCTTTTAG
26695   AAGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGAATGCGGCACGCTTTTAG
TC      AAGATAGGAATCTTTTAGGGCTTTTAAGATTATTCATAGAAGAATGCGATACGCTTTTAG
        ******  ** ** ***************** ** *************  **********

NJ      AATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAAA
S3 & S4 AATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAAA
26695   AATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAAA
TC      AATTGAAAGAAAAAATTTCGTTGTTTTTAGAGCCAAAAGATATTGTTAAAACTTATGAAA
        ******  ***************************** ***********************

NJ      ATGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCTAACAAGCATGGATTTTC
S3 & S4 ATGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCTAACAAGCATGGATTTTC
26695   ATGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCTAACAAGCATGGATTTTC
TC      ACGAAGATTTTAAAGAGCGCTGCTCAATACTTTTTAACGCCCTAAAAAGCATGGATTTTC
        * ****************** ** * *   ************ **** ***************

NJ      AAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGCGATTAAGCCAGCTTAAGG
S3 & S4 AAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGCGATTAAGCCAGCTTAAGG
26695   AAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGCGATTAAGCCAGCTTAAGG
TC      AAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGTGATTAAGCCAGCTTAAAG
        ****************************************** **************** *

NJ      GTAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAACTCGCATGGCGTTGAAT
S3 & S4 GTAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAACTCGCATGGCGTTGAAT
26695   GTAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAACTCGCATGGCGTTGAAT
TC      GTAAAGATTTTTTTAAACCTTTGCGCATTCTTTTAATTGGGGATTCGCATGGCGTTGAAT
        **** ******** ************** *******  *** * ****************

NJ      TGCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGTTTTTAAGGCTCAAAGCA
S3 & S4 TGCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGTTTTTAAGGCTCAAAGCA
26695   TGCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGTTTT-AAGGCTCAAAGCA
TC      TGCCTTTGATTTTCCCTTATATTCAAAGCCATTATCAAGAAGTTTT-AAGGCTCAAAGCA
        ***************** ***** ********* ************* *************

NJ      TGACCCGGGGGAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTNNNNNTCCGAG-
S3 & S4 TGACCCGGGGGAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTNNNNNTCCGAGN
26695   TGA-------------------------------------------------------
```

```
TC      TGA----------------------------------------------------
        ***
```

One region of about 5 bp in the *gltX2* gene in the Cheetah straingave ambiguous signals in every sequencing attempt (Red box in Figure 2). As these ambiguous bases are silent in the translated protein product, they do not impact conclusions about the variability of *gltX2*.

The *gltX2* gene from R7 was PCR amplified using primers NJ201 and NJ202. The products were cloned into the TOPO pCR2.1 vector and sequenced. Similar to the Cheetah strain, the sequences showed a number of mismatches compared to our earlier results and the data aligned with very good agreement to the *gltX2* gene from *Hp* 26695.These results are shown in Figures 4 and 5. Based on these data, we have concluded that strain R7 *gltX2* is not truncated as originally suspected.

Figure 2.4: Alignment of sequencing data using M13forward(M13f) primer for the strain R7 *gltx2.* TC – Data from Dr. Terry Cathopoulis, 26695 – data from *H. pylori* genome

```
M13f    NGGCGANTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAG 60
26695   ------------------------------------------------------------
TC      ------------------------------------------------------------


M13f    AATTCGCCCTTCGGGATCCATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCACAT 120
26695   -------------------ATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCACAT 41
TC      -------------------ATGCTTCNTTTTGCGCCTTCGCCTACAGGGGATATGCACAT 41
                           ******* ********************************


M13f    AGGGAATTTAAGGGCAGCCATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCCTT 180
26695   AGGGAATTTAAGGGCAGCCATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCCTT 101
TC      AGGGAATTTAAGGGCAGCCATTTTTAACTACATTGTGGCCAAACAGCAACATAAACCCTT 101
        ************************ ************** ********* **********


M13f    TCTCATTCGCATTGAAGACACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAGAT 240
26695   TCTCATTCGCATTGAAGACACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAGAT 161
TC      TCTCATTCGCATTGAAGACACAGATAAAGAACGCAACATTGAAGGCAAAGATCAAGAGAT 161
        ************************ ***** ******************** ********


M13f    TTTAGAAATTTTAAAGCTTATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCATAA 300
26695   TTTAGAAATTTTAAAGCTTATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCATAA 221
TC      TTTAGAGATTTTAAAGCTTATGGGGATAAGTTGGGATAAACTCGTGTATCAAAGCCATAA 221
        ****** ************************ ***** ** ********************


M13f    TATAGATTACCACAGAGAAATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTATTG 360
26695   TATAGATTACCACAGAGAAATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTATTG 281
TC      CATAGATTACCACAGAGAAATGGCAGAAAAATTGCTTAAGGAAAATAAAGCGTTTTATTG 281
```

```
         ********************************* ** ** ********************

M13f     TTATGCGAGTGCGGAGTTTTTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGCCC 420
26695    TTATGCGAGTGCGGAGTTTTTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGCCC 341
TC       CTATGCGAGCGCGGAATTTTTAGAACAAGAAAAAGAAAAAGCCAAAAACGAAAAACGCCC 341
         ******** ***** *********  ********************** ***********

M13f     TTTCAGGTATTCAGACGAGTGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTGGT 480
26695    TTTCAGGTATTCAGACGAGTGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTGGT 401
TC       TTTCAGGTATTTAGACGAATGGGCGACTTTAGAAAAAGACAAGCACCATGATCCTGTGGT 401
         *********** ****** ***** *************************  ********

M13f     GCGTTTAAAAGCCCCAAATCATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTGAA 540
26695    GCGTTTAAAAGCCCCAAATCATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTGAA 461
TC       GCGTTTAAAAGCCCCAAATCATGCGGTTTCTTTTAATGATGCGATCAAAAAAGAAGTGAA 461
         *************************** ***** ** ******** **************

M13f     ATTTGAACCTGATGAATTGGATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACTTA 600
26695    ATTTGAACCTGATGAATTGGATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACTTA 521
TC       ATTTGAACCTTATGAATTGGATTCTTTTGTGCTTTTAAGAAAGGATAAAAGCCCGACTTA 521
         ********** ************************ *** ************* *****

M13f     TAATTTCGCTTGCGCATGCGATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGA 660
26695    TAATTTCGCTTGCGCATGCGATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGA 581
TC       TAATTTCGCTTGCGCATGCGATGATTTGCTTTATGAAATCAGTCTGATTATTAGAGGCGA 581
         ********************************** *************************

M13f     AGATCATGTGAGTAACACCCCCAAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAATGA 720
26695    AGATCATGTGAGTAACACCCCCAAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAATGA 641
TC       AGATCACGTGAGTAACACCCCTAAACAGATTTTAATCCAACAAGCTTTAGGCTCAAATGA 641
         ****** ************** ***** ** ******** ************** *****

M13f     TCCGATTGTTTATGCGCATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAG 780
26695    TCCGATTGTTTATGCGCATTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAG 701
TC       CCCTATTATTTATGCGCATTTACCCATTATCTTAGATGAAGCAAGCGGTAAAAAAATGAG 701
          ** *** ************** ******** ********** ************ *****

M13f     TAAAAGAGATGAAGCCTCCAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTANCGGTTGC 840
26695    TAAAAGAGATGAAGCCTCCAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGC 761
TC       CAAAAGAGACGAAGCCTCTAGCGTGAAATGGCTTTTAAATCAAGGGTTTTTGCCGGTTGC 761
         ******** ******** ***************** **************  *******

M13f     GATTGCGAATTACCTCATCACTATCGGTAATAAAGTGCCTAANGAAGTTTTTAGCCTTGA 900
26695    GATTGCGAATTACCTCATCACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGA 821
TC       GATCGTGAATTACCTCATCACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGA 821
         *** * ************************************ *****************

M13f     TGAAGCGATAGAATGGTTTAGTTTAGAAAATCTTTNCAGTNCTCNNNCTCNTTTTNATTT 960
26695    TGAAGCGATAGAATGGTTTAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTT 881
TC       TGAAGCGATAGAATGGTTCAGTTTGGAAAACCTTTCTAATTCCCCGGCTCATTTTAATTT 881
         ****************** ***** ***** ****  * * * *   *** **** ****

M13f     AAAATATTTAAAN--NCNNANNNNCNAGCATTTAA-GCTTTTNGACGATGANN------- 1010
26695    AAAATATTTAAAACACTTAAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATT 941
TC       AAAATACTTAAAACACTTAAACCACCAACATTTAAAGCGTTTAGACGATGAAAAATTATT 941
         ****** *****       *    * * ******* ** *** ********

M13f     ------------------------------------------------------------
26695    AGAACTCACTTCAATAAAAGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGA 1001
TC       AGAACTCGCCCCAACAAAAGATAAAAATCTTTTAGGGCTTTTAAGGTTATTCATAGAAGA 1001


M13f     ------------------------------------------------------------
```

```
26695   ATGCGGCACGCTTTTAGAATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATAT 1061
TC      ATGCGGCACGCTTTTAGAATTGAAAGAAAAAATTTCGTTGTTTTTAGAGCCAAAAGATAT 1061

M13f    ------------------------------------------------------------
26695   TGTTAAAACTTATGAAAATGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCT 1121
TC      TGTTAAAACTTACGAAAACGAAGATTTTAAAGAGCGTTGCTCAATTCTTTTTAACGCCCT 1121

M13f    ------------------------------------------------------------
26695   AACAAGCATGGATTTTCAAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGCG 1181
TC      AAAAAGCATGGATTTTCAAGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGTG 1181

M13f    ------------------------------------------------------------
26695   ATTAAGCCAGCTTAAGGGTAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAA 1241
TC      ATTGAGCCAGCTTAAGGGTAAGGATTTTTTCAAACCCTTGCGCATTCTTTTAACCGGAAA 1241

M13f    ------------------------------------------------------------
26695   CTCGCATGGCGTTGAATTGCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGT 1301
TC      CTCGCATGGCGTTGAATTGCCGTTGATTTTCCCTTATATTCAAAGCCATTATCAAGAAGT 1301

M13f    --------------------
26695   TTTAAGGCTCAAAGCATGA 1320
TC      TTTAAGGCTCAAAGCATGA 1320
```

Figure 2.5: Alignment of sequencing data using M13reverse (M13r) primer for the strain R7 *gltx2.* TC – Data from Dr. Terry Cathopoulis, 26695 – data from *H. pylori* genome

```
M13r    ------------------------------------------------------------
26695   ATGCTTCGTTTTGCGCCTTCGCCTACAGGGGATATGCACATAGGGAATTTAAGGGCAGCC 60
TC      ATGCTTCNTTTTGCGCCTTCGCCTACAGGGGATATGCACATAGGGAATTTAAGGGCAGCC 60

M13r    ------------------------------------------------------------
26695   ATTTTCAACTACATTGTGGCTAAACAGCAATATAAACCCTTTCTCATTCGCATTGAAGAC 120
TC      ATTTTTAACTACATTGTGGCCAAACAGCAACATAAACCCTTTCTCATTCGCATTGAAGAC 120

M13r    ------------------------------------------------------------
26695   ACAGACAAAGAGCGCAACATTGAAGGCAAAGACCAAGAGATTTTAGAAATTTTAAAGCTT 180
TC      ACAGATAAAGAACGCAACATTGAAGGCAAAGATCAAGAGATTTTAGAGATTTTAAAGCTT 180

M13r    ------------------------------------------------------------
26695   ATGGGGATAAGCTGGGACAAGCTCGTGTATCAAAGCCATAATATAGATTACCACAGAGAA 240
TC      ATGGGGATAAGTTGGGATAAACTCGTGTATCAAAGCCATAACATAGATTACCACAGAGAA 240

M13r    ------------------------------------------------------------
26695   ATGGCAGAAAAATTACTGAAAGAAAATAAAGCGTTTTATTGTTATGCGAGTGCGGAGTTT 300
TC      ATGGCAGAAAAATTGCTTAAGGAAAATAAAGCGTTTTATTGCTATGCGAGCGCGGAATTT 300

M13r    ------------------------------------------------------------
26695   TTAGAAAGAGAAAAAGAAAAAGCCAAAAATGAAAAACGCCCTTTCAGGTATTCAGACGAG 360
TC      TTAGAACAAGAAAAAGAAAAAGCCAAAAACGAAAAACGCCCTTTCAGGTATTTAGACGAA 360

M13r    ------------------------------------NNNNGNGCGTTTAAAAGCCCCAAAT 25
26695   TGGGCCACTTTAGAAAAAGACAAGCACCATGCCCCTGTGGTGCGTTTAAAAGCCCCAAAT 420
TC      TGGGCGACTTTAGAAAAAGACAAGCACCATGATCCTGTGGTGCGTTTAAAAGCCCCAAAT 420
                                           * ********************

M13r    CNTGNGNNNTCTT-CAACGATGCGATTAAAAAAGAAGTGAAATT--GANNNGATGAATTG 82
26695   CATGCGGTGTCTTTCAACGATGCGATTAAAAAAGAAGTGAAATTTGAACCTGATGAATTG 480
TC      CATGCGGTTTCTTTTAATGATGCGATCAAAAAAGAAGTGAAATTTGAACCTTATGAATTG 480
        * ** *   ****  ** ******** *****************   *    ********

M13r    GATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCNNACTTATAATTTCGCTTGCGCNTGC 142
26695   GATTCTTTTGTGCTTTTGAGACAGGATAAAAGCCCTACTTATAATTTCGCTTGCGCATGC 540
TC      GATTCTTTTGTGCTTTTAAGAAAGGATAAAAGCCCGACTTATAATTTCGCTTGCGCATGC 540
        ***************** *** ************  ******************** ***

M13r    GATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGAAGATCATGTGAGTAACACC 202
26695   GATGATTTGCTTTATAAAATCAGTCTGATTATTAGAGGCGAAGATCATGTGAGTAACACC 600
TC      GATGATTTGCTTTATGAAATCAGTCTGATTATTAGAGGCGAAGATCACGTGAGTAACACC 600
        *************** ******************************** ************

M13r    CCCCAAACAANTCTTAATCCAGCAAGCTTTAGGCTCCAATGATCCGATTGTTTATGCGCA 262
26695   CCC-AAACAAATCTTAATCCAGCAAGCTTTAGGCTCCAATGATCCGATTGTTTATGCGCA 659
TC      CCT-AAACAGATTTTAATCCAACAAGCTTTAGGCTCAAATGACCCTATTATTTATGCGCA 659
        **  *****  * ******** ************** ***** ** *** **********

M13r    TTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAGTAAAAGAGATGAAGCCTC 322
26695   TTTGCCCATTATTTTAGATGAAGTAAGCGGTAAAAAGATGAGTAAAAGAGATGAAGCCTC 719
TC      TTTACCCATTATCTTAGATGAAGCAAGCGGTAAAAAAATGAGCAAAAGAGACGAAGCCTC 719
```

```
        *** ******** ********** ************ ***** ******** ********

M13r    CAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGCGATTGCGAATTACCTCAT 382
26695   CAGCGTGAAATGGCTTTTGAATCAAGGGTTTTTACCGGTTGCGATTGCGAATTACCTCAT 779
TC      TAGCGTGAAATGGCTTTTAAATCAAGGGTTTTTGCCGGTTGCGATCGTGAATTACCTCAT 779
         ***************** ************** *********** * ************

M13r    CACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGTT 442
26695   CACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGTT 839
TC      CACTATCGGTAATAAAGTGCCTAAGGAAGTTTTTAGCCTTGATGAAGCGATAGAATGGTT 839
        ************************************************************

M13r    TAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTTAAAATATTTAAAACACTT 502
26695   TAGTTTAGAAAATCTTTCCAGTTCTCCGGCTCATTTTAATTTAAAATATTTAAAACACTT 899
TC      CAGTTTGGAAAACCTTTCTAATTCCCCGGCTCATTTTAATTTAAAATACTTAAAACACTT 899
         ***** ***** ***** * *** ********************** ***********

M13r    AAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATTAGAACTCACTTCAATAAA 562
26695   AAACCACGAGCATTTAAAGCTTTTAGACGATGACAAGTTATTAGAACTCACTTCAATAAA 959
TC      AAACCACCAACATTTAAAGCGTTTAGACGATGAAAAATTATTAGAACTCGCCCCAACAAA 959
        ******* * ********** ************ ** ************ *  *** ***

M13r    AGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGAATGCGGCACGCTTTTAGA 622
26695   AGATAAAAACCTCTTAGGGCTTTTAAGATTGTTTATAGAAGAATGCGGCACGCTTTTAGA 1019
TC      AGATAAAAATCTTTTAGGGCTTTTAAGGTTATTCATAGAAGAATGCGGCACGCTTTTAGA 1019
        ********* ** ************** ** ** ***************************

M13r    ATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAAAA 682
26695   ATTGAGGGAAAAAATTTCGTTGTTTTTAGAGCCAAAGGATATTGTTAAAACTTATGAAAA 1079
TC      ATTGAAAGAAAAAATTTCGTTGTTTTTAGAGCCAAAAGATATTGTTAAAACTTACGAAAA 1079
        *****  **************************** ***************** *****

M13r    TGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCTAACAAGCATGGATTTTCA 742
26695   TGAAGATTTTAAAGAGCGTTGTTTAGCGCTTTTTAACGCTCTAACAAGCATGGATTTTCA 1139
TC      CGAAGATTTTAAAGAGCGTTGCTCAATTCTTTTTAACGCCCTAAAAAGCATGGATTTTCA 1139
         ******************** * *   *********** **** ***************

M13r    AGCGTATAAGGATTTTGANAGTTTTAAAAAAGAAGCCATGCGATTAAGCCAGCTTAAGGG 802
26695   AGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGCGATTAAGCCAGCTTAAGGG 1199
TC      AGCGTATAAGGATTTTGAAAGTTTTAAAAAAGAAGCCATGTGATTGAGCCAGCTTAAGGG 1199
        ****************** ********************** **** **************

M13r    TAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAACTCGCATGGCGTTGAATT 862
26695   TAAGGATTTTTTCAAACCTTTGCGCATCCTTTTAACCGGGAACTCGCATGGCGTTGAATT 1259
TC      TAAGGATTTTTTCAAACCCTTGCGCATTCTTTTAACCGGAAACTCGCATGGCGTTGAATT 1259
        ****************** ******** ************ ********************

M13r    GCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGTTTTTAAGGCTCAAAGCAT 922
26695   GCCTTTGATTTTCCCCTATATCCAAAGCCATCATCAAGAAGTTTT-AAGGCTCAAAGCAT 1318
TC      GCCGTTGATTTTCCCTTATATTCAAAGCCATTATCAAGAAGTTTT-AAGGCTCAAAGCAT 1318
        *** ************ ***** ********* ************** **************

M13r    GACCCGGGGGAAAGGGCGAATTCCAGCACACNGGCGGCCGTTACTAGNGGNNCCGAGNNC 982
26695   GA------------------------------------------------------------ 1320
TC      GA------------------------------------------------------------ 1320
```

III. Conclusion

While Dr.Cathopoulis's original results were intriguing, upon further examination, they are unfortunately most likely the result of PCR errors.

The *gltX2* genes from strains B3, R7, and Cheetah were shown to be full-length and highly homologous to *gltX2* from *Hp*26695.

Thus, the 16 strains of *Hp* that we analyzed have highly homologous *gltX2* genes. Therefore, even though *Hp* has a high rate of evolution, *gltX2* seems to be under robust selection pressures to maintain its current primary sequence.

IV. Materials and Methods

Genes were PCR amplified using primers listed in Table 2.1 and Pfu polymerase according to the manufacturer's instructions. Cloning into the TOPO pCR2.1 vector was done according to the procedure in the TOPO manual (Invitrogen).[14]

Table 2.1: Primers used in this study.

Primer	Sequence
NJ201	GCTTGGCGTTAGCCAAGTGCTAATCTCTTAAATGATGCC
NJ202	CGTAATGAGCGAGCTTAAAATCACCGCTATCGC
S3	CGGGATCCATGCTTCGTTTTGCGCCTTCG
S4	TCCCCCGGGTCATGCTTTGAGCCTTAAAAACTTC
NJ205	GCGTTTTATTGTTATGCG
NJ206	GCCCCAAATCATGCGG
NJ207	GCCATTTCACGC
NJ208	CGCTTCATCAAGGC

All sequencing was done by the sequencing facility at Wayne State University, Medical School using the M13forward and M13reverse primers.

All alignments were performed using ClustalW2.[6]

V. Acknowledgement

The author would like to thank Dr. Douglas Berg of Washington University at Saint Louis for supplying the 16 different *Hp* genomic DNA.

VI. References

1. Brown, J.R., Doolittle, W.F. (**1999**) Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *J. Mol. Evol.***49**, 485-495.

2. Cathopoulis, T., Chuawong, P., and Hendrickson, T. L. (**2007**) Novel tRNA aminoacylation mechanisms. *Mol. Biosyst.***3**(6):408-18.

3. Skouloubris, S., Ribas de Pouplana, L., De Reuse, H., Hendrickson, T.L. (**2003**) A non cognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proc. Natl. Acad. Sci. U S A.***100**(20):11297-302.

4. Berg, D. *et al* (**2000**) Differences in genotypes of *Helicobacter Pylori* from different human populations. *J. Bacteriology.***182**, 3210-3218.

5. Dr. Terry Cathopoulis, Appendix A, PhD Thesis, Johns Hopkins University.

6. http://www.ebi.ac.uk/Tools/clustalw2/index.html

Chapter III

Investigating the mechanism of ammonia transport in the tunnel of *H. pylori* GatCAB

I. Introduction

The first crystal structure of GatCAB (from *Staphylococcus aureus*) was reported in 2006.[10] Recently, a second structure was reported, using the enzyme from *Aquifex aeolicus*.[15] These structures show that the two distal active sites in GatA and GatB are connected by a putative hydrophilic tunnel that is filled with several ordered water molecules. While other enzymes with ammonia tunnels are known, [22] the GatCAB enzyme is the first example of a *highly hydrophilic* tunnel.[10] This observation makes GatCAB an unusual enzyme with an apparently novel mechanism of ammonia delivery.

The ammonia molecule, generated by hydrolysis of glutamine in the active site of GatA, is transported through a putative 37 Å tunnel, to the active site of GatB. The presence of ordered waters, as well as conserved acidic and basic amino acids, throughout this tunnel seem to suggest transport of ammonia by successive protonation and deprotonation steps.[10] Glu125 (in GatB) lies close to the interface of GatA and GatB; this residue had been proposed to be a regulatory gate for the transport of ammonia.[10] The crystal structure of *Aquifex aeolicus* [15] however, suggests that the channel connects the two active sites of GatA and GatB *without* the presence of a gate. No consensus regarding the role of glutamate has emerged so far. Furthermore, the presence of an active site base near GatB should be essential for the delivery of neutral ammonia for its nucleophilic attack, however this putative base remains unknown.

The goal of this project was to develop an understanding of the mechanism of ammonia transport through the *H. pylori* GatCAB tunnel. An alignment reveals that all the acidic and basic residues in the tunnel are conserved amongst all known bacterial GatCAB orthologs.[10] Kinetic analyses will be used to quantify the impact of conservative and non-conservative mutations at these positions to gain insight into their roles in the mechanism of ammonia transport. The above goal will be accomplished in three steps:

Step 1: Optimize an assay for *H. pylori* GatCAB-catalyzed transamidation of Glu-tRNA^{Gln}.

Step 2: Apply this assay to an analysis of point mutations throughout the GatCAB ammonia tunnel.

Step 3: Combine kinetic results with molecular modeling to develop a mechanism for GatCAB-catalyzed ammonia transport.

II. Results and Discussion

A. Preparation of materials

E. coli BL21 (DE3) competent cells carrying the plasmid pSS003 (GatB) were transformed with pPTC032 (GatCA). (The plasmids introduce an N-terminal 6-His tag to GatB and GatC, respectively). The resultant cells were used to overexpress GatCAB and the enzyme was purified by Ni^{2+}- affinity chromatography.. In order to isolate only the complex from the eluates, size-exclusion chromatography (SEC) was used as a second purification step (Figure 3.2). The concentration of GatCAB, as determined by UV absorbance, was 93 μM.

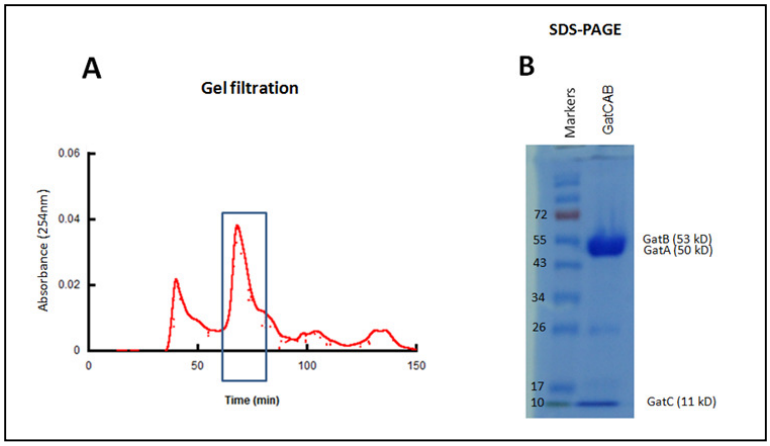


Figure 3.2: Purification of *Hp*GatCAB. A) SEC chromatogram of the eluate obtained from Ni^{2+} affinity chromatography. Five fractions were collected. Based on times of elution and molecular weight the boxed region was identified as GatCAB. B) SDS-PAGE of this region confirmed the presence of GatCAB. The impurity at 26 kD was attributed to a degradation product of GatB.

The plasmid encoding a six-histidine tagged variant of the CCA-adding enzyme was generously provided by Dr. Rebecca Alexander of Wake Forest University. The CCA-adding enzyme, required for the first step of the [^{32}P] tRNA/nuclease P1assay (Figure 3.5, also see the discussion below), was over expressed and purified by cobalt affinity chromatography (Figure 3.3). The concentration was determined by UV absorbance to be 55 µM.

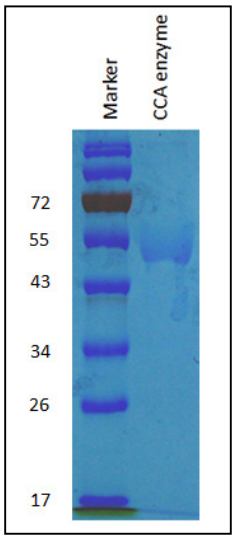


Figure 3.3: Purification of CCA-adding enzyme. The CCA-adding enzyme was overexpressed and purified by cobalt affinity chromatography. Analysis by SDS-PAGE gel shows that the enzyme is pure.

B. Optimization of an assay for *H. pylori* GatCAB-catalyzed transamidation of Glu-tRNA^{Gln}.

The net reaction catalyzed by GatCAB is the transamidation of Glu-tRNA^{Gln} to produce Gln-tRNA^{Gln}. This activity can be monitored by thin-layer electrophoresis (TLE) or two different thin-layer chromatographic (TLC) methods.[16] We initially sought to optimize our TLE assay in order to use it as a quantitative measure of GatCAB kinetics. This technique separates ^{14}C-labeled glutamate from glutamine (following hydrolysis from the tRNA) via electrophoresis on a TLC plate. However, several attempts (using a GatCAB concentration of 310 nM) failed to give clear resolution of time points over a reaction time course. A final assay, with 10 μM enzyme, was carried out over 120 min. In this case, only 12 – 25% conversion was observed and quantification by phosphorimager was inaccurate. These results led us to conclude that GatCAB was active, but that TLE/^{14}C-phosphorimaging was not sensitive enough for kinetic resolution.

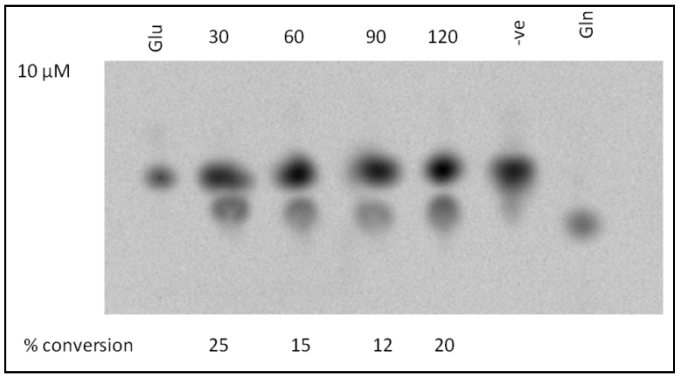


Figure 3.4: TLE assay of transamidation catalyzed by 10 μM *Hp*GatCAB. Time points of 30, 60, 90 and 120 min are shown. Control spots of ^{14}C-labeled Glu and Gln are on the left and right, respectively. A no enzyme control (-ve) is also included. Conversion of Glu to Gln is observed but is non-linear.

Next, we decided to use the [^{32}P] tRNA/nuclease P1 assay for our system. This assay is based on the incorporation of ^{32}P-ATP into the 3'-end of the tRNA, using the

CCA-adding enzyme, and subsequent treatment of the reaction with P1 nuclease.[17] Figure 3.5 shows a summary of this assay: the 3' AMP of tRNA is exchanged with ^{32}P-ATP by the CCA adding enzyme, incorporating a ^{32}P label between the last two nucleotides of the tRNA (Figure 3.5A). Aminoacylation and transamidation reactions take place on the radioactive terminal nucleotide of $tRNA^{Gln}$ (Figures 3.5B and 3.5C). Digestion of $tRNA^{Gln}$ with P1 nuclease cleaves the terminal nucleotide with the attached amino acids (Figure 3.5D). TLC and phosphorimaging are used to resolve, visualize, and quantify the starting material (Glu-AMP) and products (Gln-AMP) (Figure 3.5E).

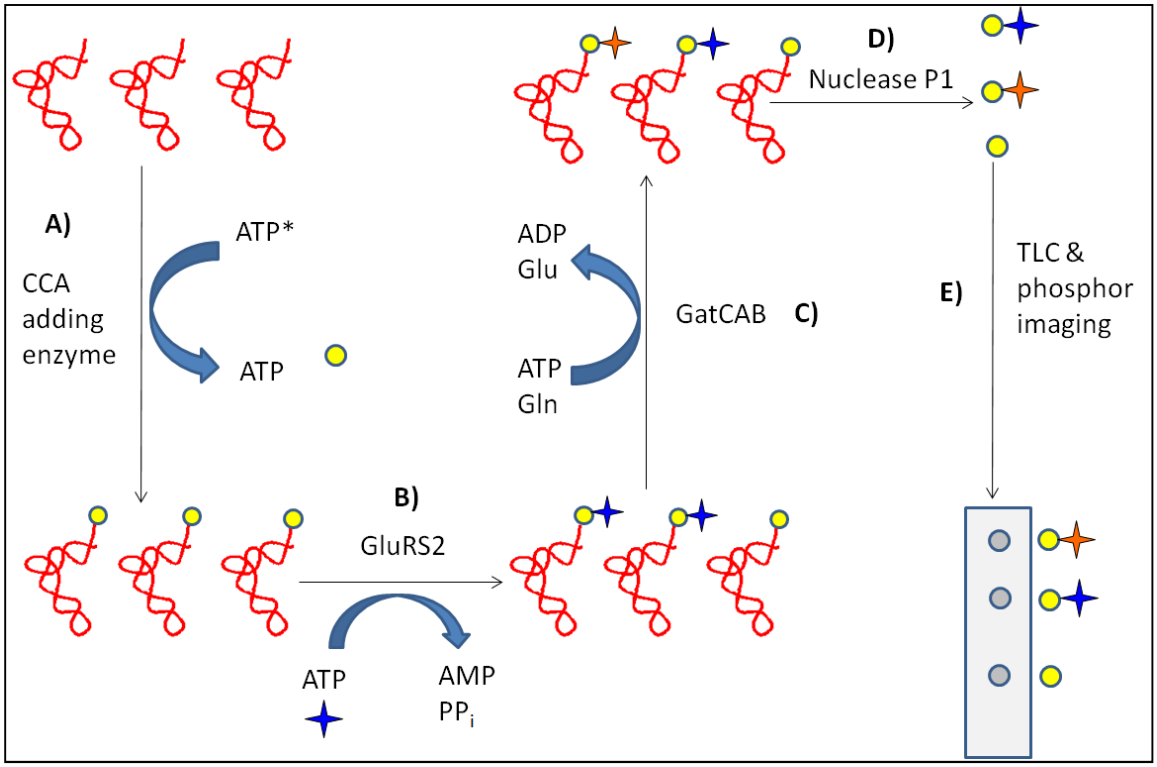


Figure 3.5:[17] [^{32}P] tRNA/nuclease P1 assay for transamidation catalyzed by *H. pylori* GatCAB. A) Transfer RNA^{Gln} is labeled with ^{32}P by treatment with the CCA adding enzyme and α–^{32}P-ATP. **B**) The labeled $tRNA^{Gln}$ is aminoacylated with Glu by GluRS2. **C**) Glu-$tRNA^{Gln}$ is converted to Gln-$tRNA^{Gln}$ by GatCAB. **D**) Digestion of by nuclease P1 gives three possible products in the form of labeled AMPs. **E**) These products are separated and visualized by TLC and phosphorimaging, respectively.

This assay required optimization before kinetic analyses were possible. The result of an initial attempt is shown in Figure 3.6. A spot above Glu-AMP did appear over time, suggesting that it is Gln-AMP. However, the two spots were not well resolved.

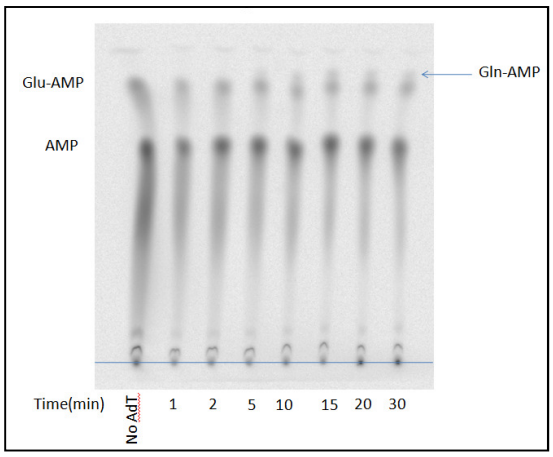


Figure 3.6: Initial attempt at a ^{32}P-based transamidation assay. Time points (5 µL) were quenched in a P1 nuclease mix and spotted 1 cm from the base of a PEI – cellulose plate. A no enzyme control was also conducted. The plate was run in 100 mM ammonium acetate in 5% acetic acid. Gln-AMP grows in over time, but with poor resolution

Based on a report on the detection of nucleotides by mass spectrometry, [20] pre-treatment of the TLC plates with ammonium hydroxide was tested to improve resolution. While this method afforded good resolution, ammonium hydroxide corroded the plates, causing the PEI-Cellulose matrix to disintegrate. Another report suggested pre-treatment with water.[21] This procedure yielded good resolution while maintaining plate integrity (Figure 3.7).

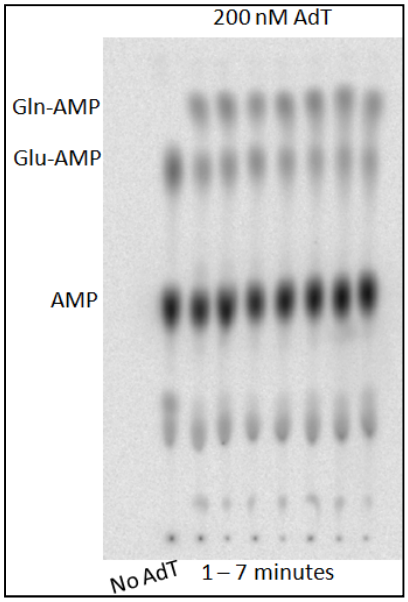


Figure 3.7: Improved resolution. Pre-treatment of the TLC plate with water improves resolution. The assay was performed with 200 nM of AdT. Time points were taken at one minute intervals.

Next, we set out to identify conditions for measuring initial rates (Figure 3.8). Enzyme concentration and time points were varied. GatCAB concentrations below 10 nM gave less than 10% conversion of Glu to Gln (ideal for measuring initial rate kinetics). The last optimization of the assay was focused on the observation of an extra spot in the no enzyme control lanes (Figure 3.9). This spot has the same R_f value as Gln-AMP and its intensity varied from 30% to 60% (normalized with respect to Glu-AMP and Gln-AMP). Such a high percentage was not desirable for kinetic analyses and efforts were sought to resolve this impurity. A number of steps summarized in Section III brought the intensity down to 3%.

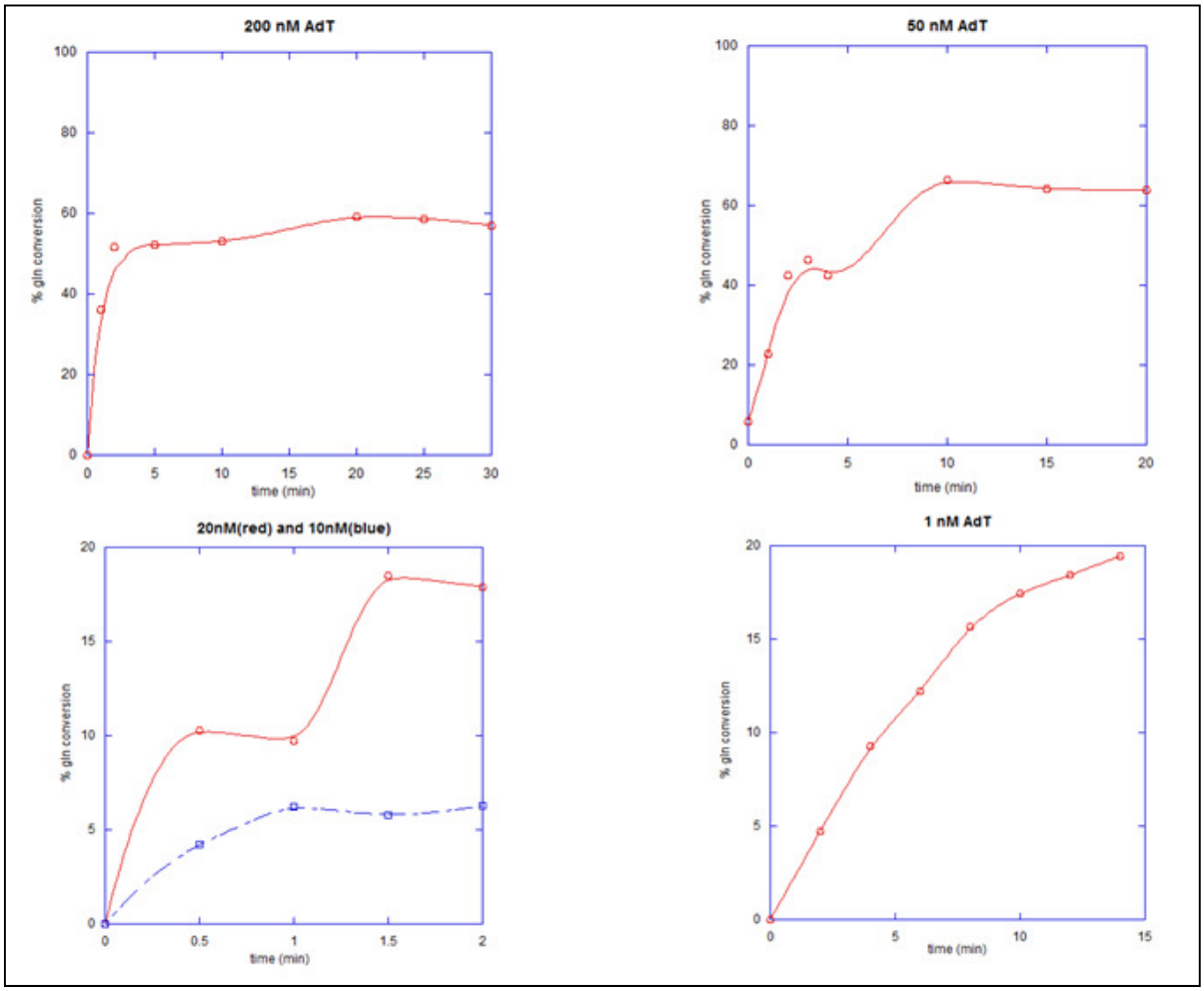


Figure 3.8: Conditions for measuring initial rates. High concentrations of AdT (200 nM and 50 nM) show greater than 10% conversion of the substrate to the product in less than 1 min. Lowering the enzyme concentration to 20 nM or 10 nM leads to a slower increase (bottom left graph). The best results were observed with less than 10 nM GatCAB (1 nM GatCAB shown here).

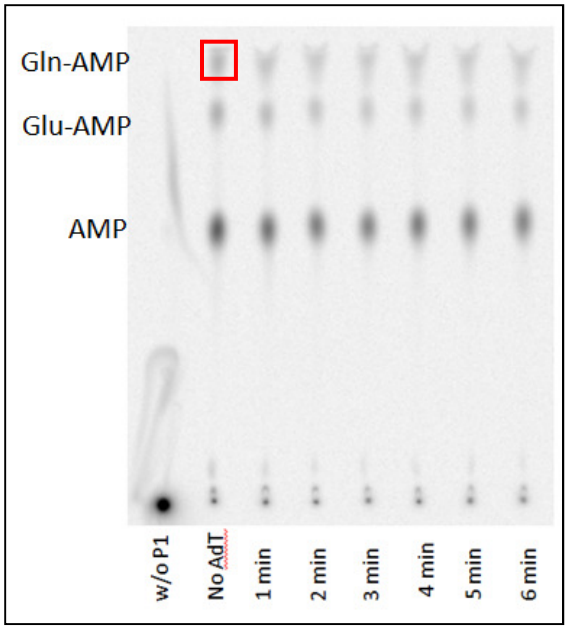


Figure 3.9: High intensity of the negative control spot. A spot of comparable intensity (red box) with the time-points was observed for the no AdT control. The exact origin of this spot is unclear. A no P1 nuclease treatment shows an absence of the spot indicating that this is not a small molecule contamination from the preparation of the Glu-tRNA^{Gln}.

A final assay with 5 nM GatCAB and all the conditions described above is shown in Figure 3.10.

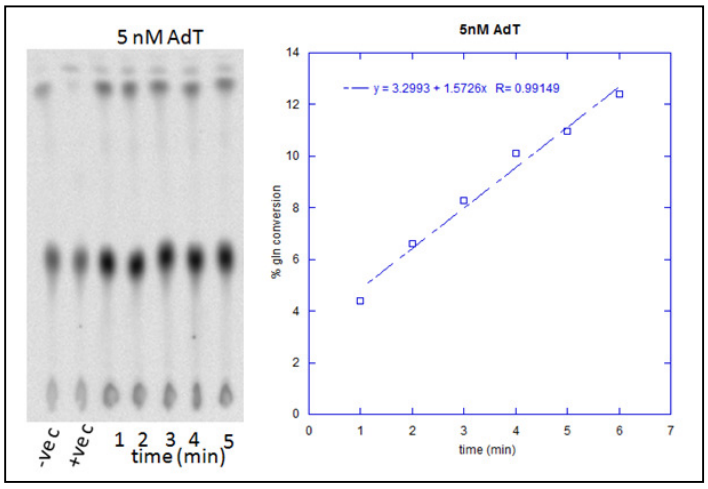


Figure 3.10: Final conditions for the assay. The extra spot in the negative control lane is now reduced to 3%. Also, a positive control (prepared by using 400 nM GatCAB over 1 hr) confirms the identity of Gln-AMP. The increase is linear over the time range tested.

The conditions shown in Figure 3.10 were repeated in triplicate, however, large variations were observed from trial to trial. A representative triplicate assay is shown in Figure 3.11.

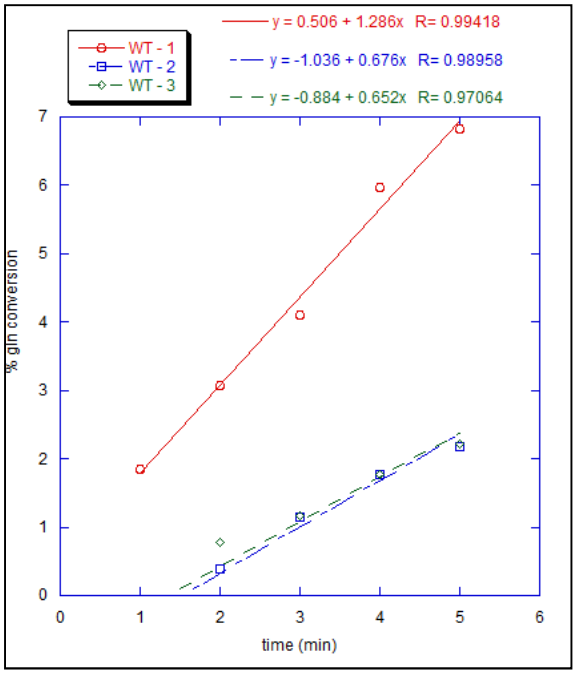


Figure 3.11: Representative triplicate assay. Each line represents one trial. Large deviations are seen from trial to trial.

We attributed these deviations in slope (two-fold in Figure 3.11) to errors in pipetting small volumes (1 µL) of enzyme stored in 50% glycerol. Enzyme dilution led to higher reproducibility but much higher levels of conversion (Figure 3.12). Lowering the enzyme concentration to 1 nM resolved this issue (Figure 3.13).

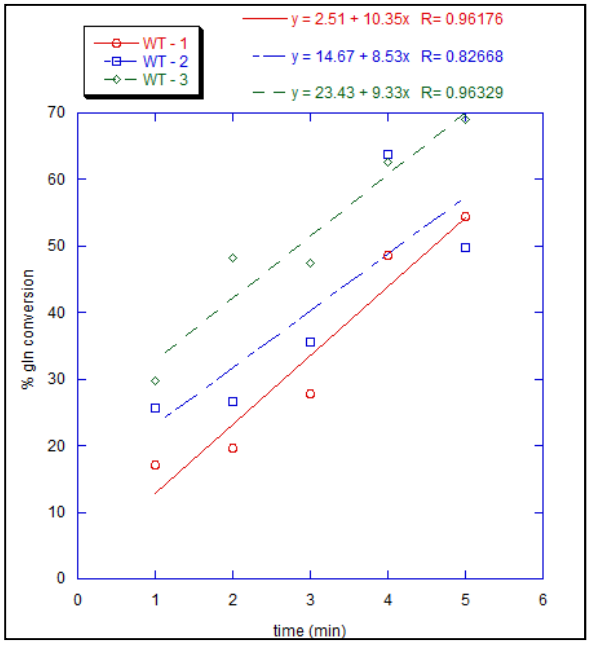


Figure 3.12: Eliminating effects of glycerol. Dilution of the enzyme in water leads to better agreement between various trials.

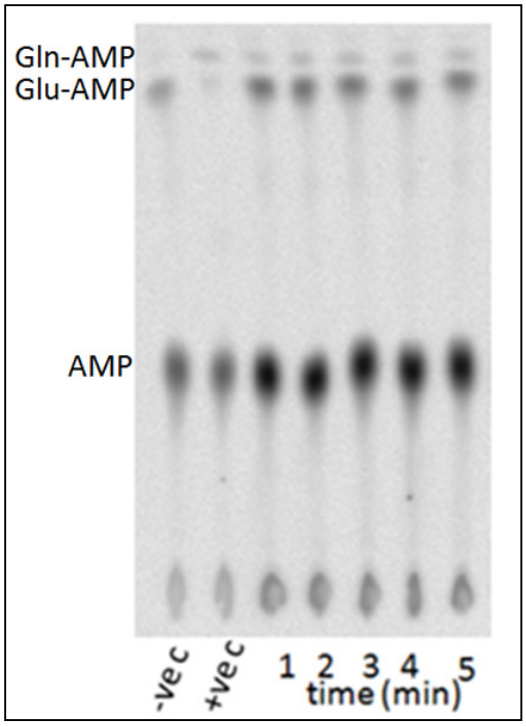


Figure 3.13: A representative image of a transamidation assay: A phosphorimage of transamidation with 1 nM AdT is shown. The negative control is of Glu-tRNA^{Gln} treated with P1 nuclease (no AdT). The positive control is Gln-tRNA^{Gln} formed by aminoacylation of *Hp* tRNA^{Gln} by *E. coli* GlnRS.

C. *Hp* GatCAB shows a biphasic response

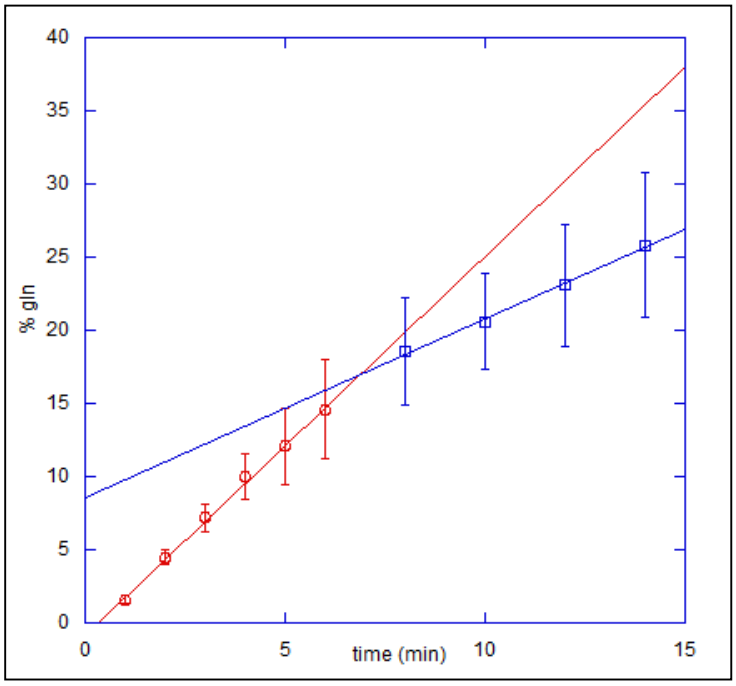


Figure 3.14: Biphasic response of GatCAB: An initial rate profile of transamidation by 1 nM GatCAB. Transamidation shows a biphasic response, probably due to the accumulation of the misacylated product.

An initial rate profile of transamidation by GatCAB shows a biphasic response towards transamidation. This is most likely due to accumulation of the correctly acylated product. As Gln-tRNA^{Gln} accumulates over time, it is possible that it becomes a competing substrate to Glu-tRNA^{Gln} leading to a reduced rate of transamidation.

D. An analysis of point mutations throughout the GatCAB ammonia tunnel.

QuikChange site-directed mutagenesis was used to construct two mutations in GatA (D185A and D185N) and four mutations in GatB (K89A, K89R, Y91A, and Y91F); the entire open reading frame of each clone was sequenced in its entirety. The remaining mutations in Table 3.1 were constructed by other members of the lab.

GatA	GatB
T149A,T149S,T149V	R79A,R79K
R174A,R174K	K80A,K80R
S182A,S182T	Y82A,Y82S
D185A,D185N	K89A,K89R
R323A,R323K	Y91A,Y91F
	Q92A,Q92E
	E125A,E125Q,E125D
	D278A,D278N

Table 3.1: Proposed mutations in GatCAB.

The four mutants in the GatB subunit, K89A, K89R, Y91A and Y91F, were co-expressed with wild-type GatCA and purified as a complex by cobalt affinity chromatography. We found that using cobalt resin instead of Nickel resin gave higher

purity and SEC was not required to further purify these mutants (Figure 3.15). The remaining mutants in Table 3.1 were purified by other members of the lab.

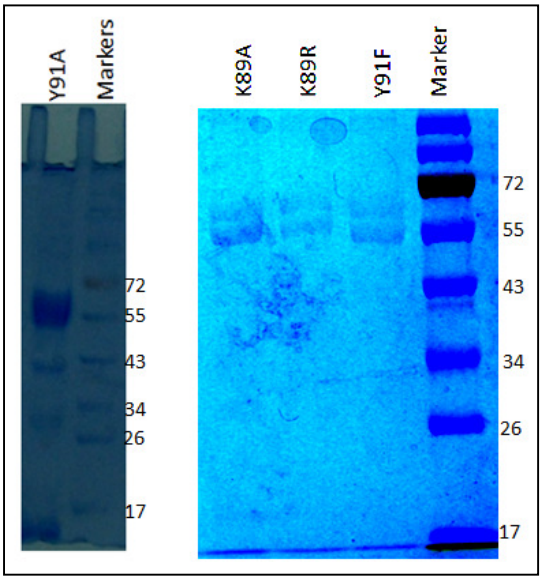


Figure 3.15: Purification of GatB mutants. Each of the GatB mutants was co-expressed with wild-type GatCA. They were purified as a complex by cobalt affinity chromatography. SDS-PAGE of the purified fractions is shown here.

The impact of some of these mutants on initial rates of transamidation is shown in Figure 3.16. A general conclusion that can be drawn is that the ammonia tunnel is sensitive to mutagenesis.

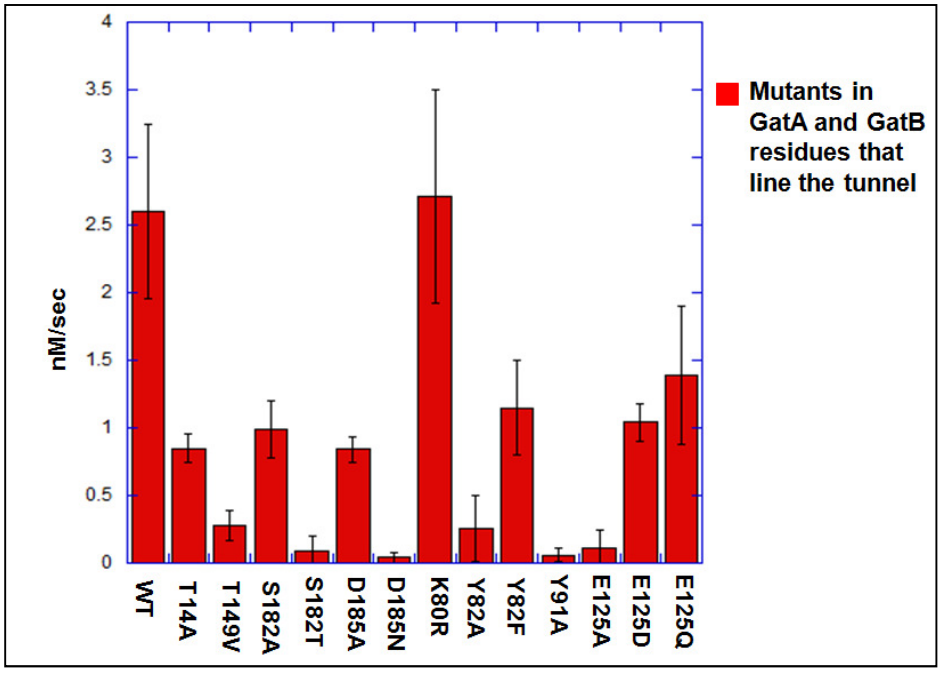


Figure 3.16 Effect of mutants on GatCAB's initial rate of transamidation: Most mutations are deleterious to enzyme activity. For a detailed discussion on each mutant, see text.

Residues in GatA: Conservative mutations in GatA (T149V, S182T and D185N) completely abolish transamidation activity. Alanine mutations, however, tend to have ~40% activity. It remains to be seen if this is a general feature of the residues in GatA that line the tunnel.

Residues in GatB: Out of all the mutants screened so far, K80R is the only mutant that shows wild-type activity. This result implies that a positive charge near the GatB active site is important for efficient transamidation. For the remaining residues, alanine mutations completely abolished transamidation activity implying an important role for these residues. Glu125, a residue at the interface of GatA and GatB, shows 50% activity when it is mutated to E125Q and E125D, implying that the charged side chain as well as its size play important roles. Tyr82, a potential base near the GatB active site that could deprotonate ammonium to the nucleophilic ammonia necessary for transamidation, shows 50% activity when mutated to phenylalanine. This result suggests that Tyr82 is not serving as a base, however, the role of its aromatic side chain seems important. Tyr91, another residue near the interface of GatA and GatB, has no transamidation activity upon alanine mutation. This residue has been implicated in domain-domain communication.[13]

III. Conclusions

Two mutations in GatA (D185A and D185N) and four mutations in GatB (K89A, K89R, Y91A, and Y91F) have been constructed. The GatB mutants were co-expressed with wild-type GatCA and purified by cobalt affinity chromatography.

The [^{32}P] tRNA/nuclease P1 assay was optimized for our system by making a number of changes:

- Dilution (10X) after synthesis of labeled $tRNA^{Gln}$ and labeled Glu-$tRNA^{Gln}$ followed by concentration by passing the mixture through a 3k spin column. This removes small molecule impurities as well as unreacted radioactive ATP.
- Addition of 18 µM unlabeled Glu-$tRNA^{Gln}$ to the GatCAB reaction mixture.
- Pre-treatment of the TLC plate with water
- Solvent system – Water:1M NH_4Cl:Acetic acid (85:10:5)
- Initiation of the reaction by adding Gln and ATP
- Diluting AdT from a 50% glycerol stock to water and using a higher volume (5 µL) to avoid errors in pipeting solution containing glycerol

Preliminary results suggest that T149, S182 and D185 in GatA while K80, Y82, Y91 and E125 in GatB are important for transamidation.

IV. Materials and Methods

Unless otherwise stated, all materials were purchased from Sigma-Aldrich or Fisher Scientific.

Site-directed mutagenesis

QuikChange site-directed mutagenesis was performed on the plasmids pPTC032 (containing N-terminally tagged *gatCA* genes in an operon)[11] and pSS003 (N-terminally tagged *gatB* gene).[12] The primers used for each mutation are listed in Table 2. Typically, a polymerase chain reaction (50 µL) was conducted under the following conditions: 95 ℃$^{10.00}$; [95 ℃$^{1.30}$;65 ℃$^{1.00}$;72 ℃$^{3.00}$]$_{30}$; 72 ℃$^{10.00}$. DNA (1 µL, OD=0.5), 4 mM dNTPs, 1 mM primers and 1 µM Pfu polymerase(New England Biolabs) were used. The products were loaded on an agarose gel (treated with ethidium bromide). Agarose gel chromatography was used to confirm the expected products. The open reading

frames of each plasmid were confirmed by DNA sequencing at the sequencing facility at Wayne State University Medical School.

Table 3.2: Primers used for GatA and GatB mutagenesis. The capital letters in the primer sequence indicate the positions of the nucleotides that were mutated.

Primer	Mutation	Sequence
NJ07	K89A	ggaaaaattatttttaccctgatttgcctGCggcttatcaaatttcgc
NJ08		gcgaaatttgataagccGCaggcaaatcagggtaaaaataatttttcc
NJ26	K89R	ggaaaaattatttttaccctgatttgcctCGggcttatcaaatttcgc
NJ27		gcgaaatttgataagccCGaggcaaatcagggtaaaaataatttttcc
NJ28	Y91A	gatttgcctaaggctGCtcaaatttcgcagtttgaag
NJ29		cttcaaactgcgaaatttgaGCagccttaggcaaatc
NJ30	Y91F	gatttgcctaaggcttTtcaaatttcgcagtttgaag
NJ31		cttcaaactgcgaaatttgaAaagccttaggcaaatc
NJ32	D185A	cgcgtattgctctagttttgCtcaaatcgggcc
NJ33		ggcccgatttgaGcaaaactagagcaatacgcg
NJ34	D185N	cgcgtattgctctagttttAatcaaatcgggcc
NJ35		ggcccgatttgatTaaaactagagcaatacgcg

Purification of wild-type and mutant GatCAB variants

E. coli BL21 (DE3) competent cells carrying the plasmid pSS003 encoding GatB[12] were transformed with pPTC032 encoding GatCA.[11] (These plasmids introduce N-terminal 6-His tags onto GatB and GatC, respectively). These two plasmids contain two different antibiotic genes and two different origins of replication. Due to this, it is

possible to introduce two plasmids in one cell. The resultant cells were used to overexpress GatCAB and the enzyme complex was purified by Ni^{2+}-affinity chromatography. Eluted fractions were diluted to 1 mL in SEC buffer (50 mM Hepes, pH 7.2, 30 mM KCl, 6 mM $MgCl_2$, 0.1 mM EDTA, 5 mM β-mercaptoethanol). The sample was injected onto a Superdex 200 gel filtration column (Amersham). The fractions were analyzed by SDS-PAGE. The fractions containing pure GatCAB were concentrated using a YM-10 filter and stored in 50% glycerol.

The GatCAB mutants were transformed similar to wild-type above. The resultant cells were used to overexpress the GatCAB mutants. The resulting complexes were purified using cobalt affinity chromatography, since concentration of the purified complex was found to be higher compared to nickel affinity chromatography. Also the complex thus obtained did not contain significant impurities, thus avoiding size exclusion chromatography. The purified fractions were stored in 50% glycerol.

Thin layer electrophoresis (TLE) assay

The TLE assay was performed as described previously.[16]

[^{32}P]-tRNA/nuclease P1 assay

Polyethyleneimine-cellulose plates were purchased from EMD chemicals and P1 nuclease was purchased from Sigma-Aldrich. [α-^{32}P] ATP was purchased from American Radiolabeled Chemicals.

Transfer RNA^{Gln} was labeled with ^{32}P as previously reported with the following changes;[17] after phenol/chloroform extraction, the aqueous layer was diluted 10-fold and

passed through a 3k spin column until the volume was reduced to 100 µL. Transfer RNA^{Gln} was precipitated by isopropanol.

Glu-$tRNA^{Gln}$ was prepared by mixing the ^{32}P-labeled $tRNA^{Gln}$ with 37 µM unlabelled $tRNA^{Gln}$. This mixture was incubated with 2 µM GluRS2, as described previously.[9] The resulting Glu-$tRNA^{Gln}$ was purified and precipitated similar to $tRNA^{Gln}$ above.

The efficiency of both tRNA labeling and Glu-$tRNA^{Gln}$ synthesis was monitored by liquid scintillation counting.

The assay was performed with the key steps listed in section III.Briefly, the procedure was as follows. 18 µM unlabeled Glu-$tRNA^{Gln}$ was added to the GatCAB reaction mixture. The reaction mixture also contained 1nM of GatCAB, this was prepared by diluting from a 50% glycerol stock to water and using a higher volume (5 µL) to avoid errors in pipeting solution containing glycerol.. Other components of the reaction were as previously described.[17] The reaction was initiated by adding glutamine and ATP. The reaction was quenched as described.[17] The TLC plate was pretreated with water and dried. A solvent system of Water:1M NH_4Cl:Acetic acid (85:10:5) was used to develop the plates.

TLC plates were dried and exposed to a Kodak imaging screen for 16 hours. The screens were imaged using a Typhoon phosphoimager. ImageQuant software was used for analysis of the data. Kaleidagraph was used to plot and analyze the graphs shown in the above figures.

V. References

1. Ibba, M., and Söll D. (**2000**) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.***69**, 617-650.
2. Arnez, J.G., Moras, D. (**1997**) Structural and functional consideration of the

aminoacylation reaction. *Trends. Biochem. Sci.***22**, 211-216.

3. Cusack, S. (**1997**) Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.***7**, 881-889.
4. Woese, C. R., Olsen, G. J., Ibba, M., and Söll, D. (**2000**) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Mol. Biol. Rev.***64**, 202-236.
5. Bult C.J., *et al.* (**1996**) The complete genome sequence of the methanogenic archaeon, *Methanococcusjannaschii.Science*. **273**, 1058-1073.
6. Cathopoulis, T., Chuawong, P., and Hendrickson, T. L. (**2007**) Novel tRNA aminoacylation mechanisms. *Mol. Biosyst.***3** (6):408-18.
7. Tomb, J.F., *et al*. (**1997**) The complete genome sequence of the gastric pathogen *Helicobacter Pylori. Nature* **388**(6642):539-47.
8. Tumbula, D. L., Becker, H. D., Chang, W., Söll, D. (**2000**) Domain-specific recruitment of amide amino acids for protein synthesis *Nature.***407**, 106.
9. Skouloubris, S., Ribas de Pouplana, L., De Reuse, H., Hendrickson, T.L. (**2003**) A noncognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proc. Natl. Acad. Sci. U S A.***100** (20):11297-302.
10. Nakamura, A., Yao, M., Chimnaronk, S., Sakai, N., Tanaka, I. (**2006**) Ammonia channel couples glutaminase with transamidase reactions in GatCAB. *Science.***312** (5782):1954-1958.
11. Courtesy, Dr. Pitak Chuawong.
12. Courtesy, Dr. Stephane Skouloubris.
13. Zhao, L., *et al.* (**2012**) The kinase activity of *H. Pylori* Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase is sensitive to distal mutations in its putative

ammonia tunnel *Bichemistry*.**51** (1):273-285.

14. Invitrogen (**2006**) TOPO TA Cloning: User Manual.
15. Wu, J., Bu, W., Sheppard, K., Kitabatake, M., Kwon, S. T., Soll, D., Smith, J. L.(**2009**) Insights into tRNA-dependent amidotransferase evolution and catalysis from the structure of the *Aquifex aeolicus* enzyme. *J. Mol. Biol.***391**, 703-716.
16. Cathopoulis, T., Chuawong, P., and Hendrickson, T. L. (**2007**) A Thin-Layer Electrophoresis Assay for Asp-tRNAAsn/Glu-tRNAGln Amidotransferase, *Anal. Biochem.***360**, 151-153.
17. Sheppard, K., Akochy, P.M., Soll, D. (**2008**) Assays for transfer-RNA dependent amino acid biosynthesis. *Methods.***44**, 139-145.
18. Curnow, *et al*. (**1997**) Glu-tRNAGln amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation *Proc. Natl.Acad.Sci.USA* 94, 11819-11826.
19. Stratagene. QuikChange site-directed mutagenesis kit instruction manual.
20. Isbell, D.T., *et al*. (**1999**) Analysis of nucleotides directly from TLC plates using MALDI-MS detection *J. Anal. Chem.***365**: 625–630.
21. Wolfson, A. D, *et al*. (**2002**) Modulation of $tRNA^{Ala}$ identity by inorganic pyrophosphatase *Proc. Natl. Acad. Sci. USA***99**, 5965-5970.
22. Raushel FM, Thoden JB, Holden HM. Enzymes with molecular tunnels.(**2003**) *Acc ChemRes.***36**(7):539-4.

ABSTRACT

INVESTIGATING THE GLUTAMINE-TRNA (GLUTAMINE) SYNTHESIS APPARTUS OF THE HUMAN PATHOGEN *HELICOBACTER PYLORI*

by

NILESH JOSHI

August 2012

Advisor: Tamara L. Hendrickson

Major: Chemistry

Degree: Master of Science

Accurate protein biosynthesis is a vital process to all cellular life. Aminoacyl-tRNAs are at the heart of this process: A correctly formed aminoacyl-tRNA is critical for protein biosynthesis. Organisms have evolved many mechanisms to repair misacylated tRNAs before they cause errors in protein biosynthesis, thus maintaining the integrity of the genetic code. The human pathogen *Helicobacter pylori* (*H. pylori*) synthesizes Glu-tRNA^{Gln} as an intermediate to producing Gln-tRNA^{Gln}. This misacylated intermediate could cause lethal errors if used by the ribosome for protein synthesis. *H. pylori* repairs this intermediate by the amidotransferase GatCAB.

This dissertation focuses on indirect aminoacylation and transamidation to produce Gln-tRNA^{Gln} in *H. pylori*. A combination of phylogenetic analyses and enzymatic assays were used that contribute to a picture of this process.

Chapter 2 discusses our examination of sequence conservation of *gltX2* (the gene that encodes GluRS2) across *different H. pylori* strains. The conclusions of this sequencing effort are compared with trends in sequences of genes important in the aminoacylation step of protein synthesis.

Chapter 3 discusses one aspect of Gln-tRNA^{Gln} production – the mechanism of transport of ammonia from the active site of one subunit (GatA) to the active site of another (GatB) through an intramolecular hydrophilic tunnel. Site-directed mutagenesis of key residues lining this tunnel and their transamidation assays were performed. With these studies, a preliminary picture of ammonia transport through the tunnel can be constructed.

AUTOBIOGRAPHICAL STATEMENT

Nilesh Vasudeo Joshi

EDUCATION

MS, Chemistry, Wayne State University Detroit, MI – 2012
MSc, Chemistry, Indian Institute of Technology, Delhi, India – 2007
BSc, Chemistry, University of Mumbai, India – 2005

Relevant courses: Molecular Biology, Molecular Biophysics, Metabolism, Biochemistry, Quantum Chemistry, Organic synthesis, Chemistry of natural products, Co-ordination chemistry

RESEARCH

Publication, accepted in *Biochemistry*, 2012: Zhao, L, *et al*. "The kinase activity of the *Helicobacter pylori* Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase is sensitive to distal mutations in its putative ammonia tunnel"
Poster, 2011: "The mechanism of ammonia transport in GatCAB" at the 2011 International Symposium on Aminoacyl-tRNA Synthetases in Salt Lake City, Utah.

WORK EXPERIENCE

National Chemical Laboratory, Pune, India (2005 and 2006) Summer KVPY fellow
Performed computational studies (Hartree-Fock calculations) on the ethane molecule, Researched analyzed and prepared a report on the use of programming techniques in computational chemistry.
Indian Institute of Science, Bangalore, India (2004) Summer KVPY fellow
Synthesized a conjugated polymer poly-(dimethoxy p-phenylene xylylene) using the halo precursor route
Wayne State University, Detroit, USA (2008 – 2011) Graduate Teaching Assistant

- Responsibilities included conducting lectures; performing laboratory demonstrations of key experimental procedures; grading exams, lab reports, assignments, homework of undergraduates at above institutions
- Courses taught: General Organic Chemistry, Survey of Organic chemistry and Biochemistry, Experimental techniques in Physical Chemistry, Laboratory techniques in Organic chemistry, General and Inorganic Chemistry

SCHOLARSHIPS AND AWARDS

- Recipient of **KVPY fellowship** funded by **Dept. of Science and Technology**, Government of India. 2003-2007
- Recipient of **'Special Appreciation Award (Academics)'** from Vice Chancellor, University of Mumbai. 2005
- Selected to attend the "Nurture Camp" under the **National Initiative for Undergraduate Science.** 2004-2005
- **National Top 1 %(** out of 3000 candidates) in the **National Graduate Examination in Physics.** 2004