Importance of single molecular determinant in bacterial tryptophanyl - tRNA synthetase fidelity in expanded genetic code

By Zuzana Šimová

Ph.D. dissertation

School of Chemistry

Thesis supervisor: Dr. Eric M. Tippmann

Cardiff 2011

UMI Number: U585509

UMI U585509

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed .... (candidate)   Date ....07/12/11.........

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of
...........PhD............(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed .... (candidate)   Date ......07/12/11.........

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed .... (candidate)   Date ......07/12/11.........

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)   Date ....07/12/11.........

Acknowledgements

In the first place, I would like to thank to my supervisor Dr Eric M. Tippmann for giving me the opportunity of undertaking my Ph.D. in his laboratory, his help, encouragement and guidance, as well as the new ideas he brought to my project.

My appreciation also goes to all members of the Physical Organic Chemistry (POC) group in Cardiff University, for their help, suggestions, support and an unforgettable working atmosphere. Special thanks to Alicja Antonczak, Josephine Morris, Dr Amy Baldwin, Dr Rebbeca Fulleton, Dr Carol Goss and Dr Oliver Ladebeck. My great gratitude goes to Dr Niklaas J. Buurma and Lavinia Onell who took care of me when learning about ITC. I also appreciate their vast help with my ITC experiments and their input to the discussion. Further thanks must go to Dr Robert Richardson, Dr Larry Goldman, Dr Mihaela Dorin, Dr Mazin Othman and Stefania Narduolo.

I would like to express my gratitude to Prof Matthias Bochtler who has introduced me to protein crystallography and helped me towards a successful solution of protein structure. I also would love to say many thanks to members of his laboratory, remarkably Dr Honorata Czapinska (for her help with the crystal structure refinement), Dr Ania Piasecka, Dr Renata Filipek and Marek Wojciechovski, who gave me much advice on protein purification and crystallization. I very much appreciated the opportunity to disscuss some obstacles in my project with them.

I am very greatful to my colleagues from Prof Rudolf K. Allemann's and Dr Dafydd Jones' laboratories for the useful dialogues that helped to improve the quality of my research. I very much appreciated Dr Veronica Gonzalez's help and support with the ÄKTA system at the beginning of my Ph.D.

My thanks also goes to Gin Jhen Goh and Christian Jensen who have helped me with the language revision of this thesis.

Finally, I want to give a very big thanks to my family and friends for their continuous support, love and patience.

Table of Contents

Summary

Nonnatural amino acid incorporation is a valuable method for introducing novel chemical functional groups into proteins. For this method, an orthogonal aminoacyl-tRNA synthetase (AARS) and a cognate tRNA that suppress an encoded stop codon are introduced into the cell (these components are required to be orthogonal). Nonnatural amino acids (NAAs) are usually incorporated efficiently by using *Methanocaldococcus jannaschii* tyrosyl-AARS/tyrosyl-tRNA pair (*Mj* TyrRS/*Mj* tRNA^{Tyr}) in *Escherichia coli*. High translation fidelity of a synthetase is achieved by site-directed mutagenesis of the competent active site residues. The active site of the mutant *Mj* TyrRS displays two crucial mutations of residues that interact with the tyrosine hydroxyl group (-OH). We demonstrated that the fidelity of the synthetase would be affected if only one of these residues is restored and does not undergo mutagenesis. We found a similar situation in the case of tryptophanyl-AARS (TrpRS) from *Bacillus subtilis*. TrpRSs are structurally similar to TyrRSs, but there is one crucial residue of the substrate specificity. We uncovered that a NAA system developed to incorporate 5-hydroxytryptophan (5-OH Trp) in mammalian cells does not contain this crucial residue mutation in the TrpRS active site. Even though this mutant TrpRS was designated as a high fidelity enzyme, our results challenge this conclusion.

Expanded genetic codes have a similar capacity to impact science as has standard mutagenesis. Only the full impact of the method will be achieved if the technology functions in all cell types. Therefore, our reinvestigation of the first report of expanded genetic code in mammalian system is critical to ensuring that the field is on the optimum path to realising the full potential of the method.

List of publications

1. A.K.Antonczak (1), **Z.Simova** (1), I.Yonemoto, M.Bochtler, A.Piasecka, H.Czapinska, A.Brancale, E.M.Tippmann (2011) The importance of single molecular determinants in the fidelity of expanded genetic code. PNAS. 108, 1320-1325.

[(1) these authors contributed equally to this work]

2. Antonczak AK, **Simova Z**, Tippmann EM (2009) A critical examination of Escherichia coli esterase activity. J Biol Chem. 284, 28795-28800.

Abbreviations

Å	Ångström (10^{-10} m)
AA	Amino acid
AARS	Aminoacyl-tRNA synthetase
APS	Ammonium persulfate
ATP	Adenosine-5'-triphosphate
bp	Base pair(s)
BSA	Bovine Serum Albumine
Da	Dalton (M.W.)
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetraacetic acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_a	Association constant
K_d	Dissociation constant
K_M	Michaelis constant
k_{cat}	Turnover number
NAA	Nonnatural amino acid
Ni-NTA	Nickel-nitrilotriacetic acid
nt	Nucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
TAE	Tris-Acetate-EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
wt	Wild-type
5-OH Trp	5-Hydroxytryptophan

Nucleotide DNA bases (one letter code):

A adenine

T thymine

G guanine

C cytosine

Amino acids (one and three letter codes):

A **Ala** Alanine

N **Asn** Asparagine

D **Asp** Aspartic acid

C **Cys** Cysteine

Q **Gln** Glutamine

E **Glu** Glutamic acid

G **Gly** Glycine

H **His** Histidine

I **Ile** Isoleucine

K **Lys** Lysine

L **Leu** Leucine

M **Met** Methionine

F **Phe** Phenylalanine

P **Pro** Proline

R **Arg** Arginine

S **Ser** Serine

T **Thr** Threonine

W **Trp** Tryptophan

Y **Tyr** Tyrosine

U **Sec** Selenocysteine

O **Pyl** Pyrollysine

List of Figures and Tables

15

Introduction

The expanded genetic code field is based upon well-known textbook ideas such as site-directed mutagenesis, translation, enzyme function and kinetics, structure function. In order to introduce our work, a partial review of some concepts is provided here.

1. The Amino Acids of Proteins

Amino acids (AAs) are the building blocks of proteins.[1] There are only 20 common amino acids assigned by 61 coding triplet combinations plus 3 stop codons. All known organisms share these 20 common amino acids and although their chemical properties are quite limited they are responsible for protein diversity and their distinct properties. Two other, rather rare natural amino acids, selenocysteine (Sec)[2] and pyrollysine (Pyl)[3] were discovered later.[4]

In the structure of a single amino acid, there is a variable side chain, which is typically classified as R. In each amino acid, there is a central tetrahedral C_α covalently joined to the amino (-NH_2) and the carboxyl (–COOH) group. C_α also links to a hydrogen atom. Depending on pH, an amino acid gains a charge. At pH 7, the carboxyl group exists as -COO^- and the amino group as –NH_3^+ so the overall molecule is neutral and it is called a zwitterion (Figure 1). Amino and carboxyl groups can react together after losing a water molecule to form a covalent amide bond (peptide bond). Polymerization of amino acids results in peptide chain assembling. In this manner, a new polypeptide or protein is formed.[1]
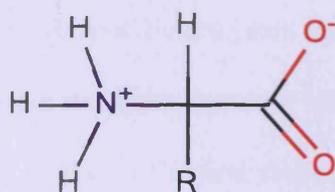
Figure 1: Amino acid general structure. –NH_3^+ and -COO^- are linked to the C_α that is also substituted with a hydrogen atom and a R-group.

Much effort has been made to study enzymes, their structures, functions and their involvement in biological processes.[5] For each enzyme's unique features, there has always been an attempt to be able to

manipulate with their structure and thereby generate some new properties. As mentioned, there are only 20 common amino acids and their chemical properties do not vary widely. Despite this inherent limitation, life has flourished with a diverse array of life forms. To help unlock the secrets of life, scientists developed methods to mutate one amino acid into one of the other 19 common ones and in this way, they have shone light on a number of life's processes.

As early as 1962, researchers became interested in mutating a common amino acid into a nonnatural one.[6] Since then, scientists have investigated a number of ways to introduce nonnatural amino acids into proteins. The desire to do so stems from the fact that nonatural amino acids (NAAs) are unique and differ from the classical 20 amino acids by their side group modification.[7] For instance, some of these amino acids are metal chelators, photoaffinity labels or have distinctive spectroscopic properties. NAAs that we do not normally find in living organisms have been used as a target for incorporation into proteins. In this manner, protein properties (acidity, nucleophilicity, hydrogen-bonding potential, etc.) can be enhanced. Proteins containing a NAA have been used for biophysical, chemical and structural studies.[8-10] The development of the NAA incorporation method is the main motivation for this work.

In the cell, proteins are synthesized during translation.[11] In this process, a messenger RNA (mRNA) molecule is used as a template, in which an order of nucleotides gives the information about how a new amino acid chain will assemble. There are specific enzymes required during translation. There would be no chance for the cell to survive if these catalysts were not involved in translation.[12] Aminoacyl-tRNA synthetase (AARS) is an enzyme that catalyzes the first steps of protein biosynthesis.[13] This enzyme is the focus of our work therefore, its function and structure will be described more in the next chapter.

2. Study of Aminoacyl-tRNA synthetases (AARSs)

Central to this work are a class of enzymes called AARSs. They are key components in translation and contribute to the fidelity of the genetic code by a number of well studied mechanisms. Our motivation for studying them stems from our desire to incorporate nonnatural amino acids in vivo and, as such, having an intimate understanding of these key players in translation[14] is essential to understanding how these enzymes can be best used in expanded genetic codes.

To begin with, AARSs perform key reactions. In these reactions, the AA, adenosine-5'-triphoshate (ATP) and transfer RNA (tRNA) are engaged. In a summary of this process, AARS is involved in an ester bond formation between AA and tRNA terminal hydroxyl group of adenosine (either 2'-hydroxyl, 2'OH or 3'-hydroxyl, 3'OH group[15]):[13]

$$AA + ATP + tRNA \rightarrow AA\text{-}tRNA + AMP + PP_i$$

The products of this reaction are the AA bound to tRNA (AA-tRNA), adenosine-5'-monophosphate (AMP) and pyrophosphate (PP$_i$). Without the formation of AA-tRNA (Figure 2), AA would not be transfered to the ribosome and the peptide chain would not form. There are at least two known mechanisms of aminoacylation in the cell. In a direct acylation there are 20 different AA-tRNA complexes that are products of direct charging by one of 20 different AARS's. Thus, one AARS uses one amino acid of the genetic code.[11, 16] An indirect pathway of acylation (tRNA-dependent amino acid modification) is the incorporation of Sec (21st amino acid).[17] Here, tRNA (serine-tRNA) is aminoacylated by an amino acid "precursor" (serine, Ser) by a non-discriminating AARS (serine-tRNA synthetase, SerRS) and further converted to Sec by another enzyme (selenocysteine synthase).[18] This was first found when an archaeal genome was fully mapped.[19, 20] Since the genome only contains 16 (out of 20) AARS sequences, the rest of AA-tRNA complexes are produced by non-specific AARS charging.

That includes charging of asparagine (Asn), cysteine (Cys), glutamine (Gln) and lysine (Lys). For example, glutaminyl-tRNA synthetase (GlnRS) is not found in archaebacteria but also in Gram-positive eubacteria, or eukaryotic organelles.[11] For tRNA[Gln] to be charged by Gln, glutamic acid (Glu) is first attached to form Glu-tRNA[Gln]. This product is taken by GluRS amidotransferase that catalyzes the removal of the ammonia group from glutamine and transfers it to a substrate that forms a new carbon-nitrogen group. Resulting tRNA[Gln] is charged with Gln (Gln-tRNA[Gln]).[18] It is not clear why there are organisms with alternative pathways of tRNA charging. The question is why we find some examples that lack certain types of AARS in nature and therefore possibly require energy or extra molecules in order to complete protein synthesis.
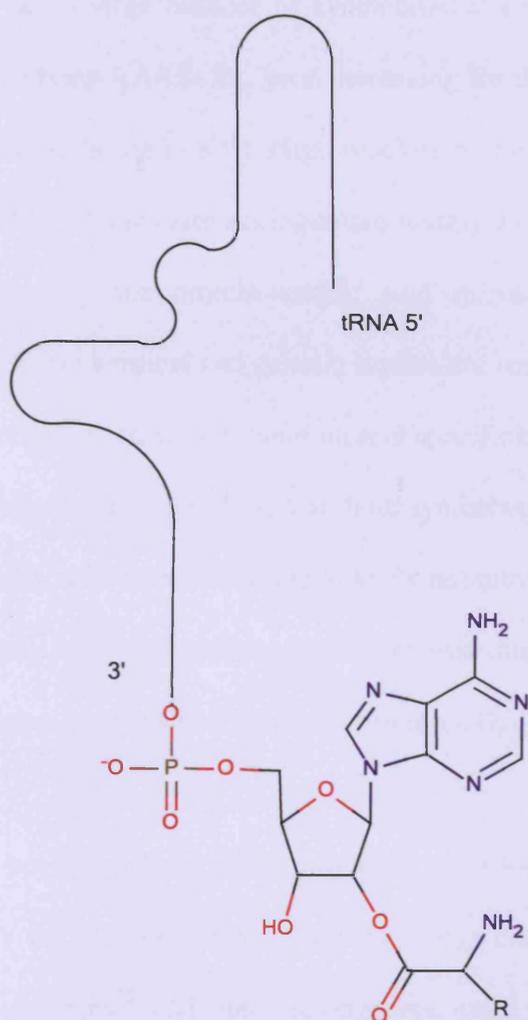
Figure 2: AA-tRNA. Amino acid is attached to the
3' tRNA terminal site at 2'OH group of the ribose.

AARSs are not only essential for protein biosynthesis but also provide high selectivity, so that they are able to join a specific amino acid with a specific tRNA molecule. This very precise system depends mostly on the AARS' ability to distinguish the right substrate from the other competing noncognate substrates. An error rate for AARSs was determined to be about 1 in 10,000.[11] This high specificity is due to a large number of contacts between AARS and amino acid or tRNA. Therefore, removing a part of the key synthetase structure or only a crucial residue mutation, could result in a high level of aminoacylation errors, which is catastrophic for any biological system.[11] The mechanisms by which AARS achieves high fidelity are at the centre of this project.

There are a large number of synthetases available for analysis and comparison.[13] Therefore, the interest in studying AARSs has been increasing for decades. That is obvious from the amount of research conducted since the early 80's. High-resolution crystal structures of AARS coupled with tRNA, amino acid or a different substrate analogue are widely available now.[21-24] Exploring these structures helps in the understanding the protein-nucleic acid interactions. Improvement of crystallization techniques together with biochemical and genetic studies are reasons why we can now understand much more about AARSs expression, structure, function and specificity. From this understanding, one can hardly think of how a biological world could exist without synthetase enzymes.[11]

Scientists have been analyzing AARSs not only to expand their structures and functions but so they might understand the regulation of gene expression. Synthetases are the link between the nucleic acid code and amino acid chain. It has been observed that AARS and tRNA directly or indirectly control gene expression.[13]

On a profound level, large numbers of AARS's can be investigated, compared and used for evolutionary analysis since they represent a large class of enzymes with similar functions.[13] Synthetases are ancient proteins[25] and their phylogenetic trees can be compared.[18] Even though their functional similarities are obvious, these do not reflect their structural framework. For instance, one question is:

how two distinctive synthetase groups have evolved?[18, 25, 26] By exploring of the standard molecular level of enzymes we can discover in depth the cell organization, how does the whole cell machinery functions as well as how it has evolved.[18]

2.1   Mechanism of aminoacylation reaction

In order to understand how a synthetase functions, the aminoacylation reaction will be described in more detail here. In Figure 4 the general mechanism of aminoacylation reaction is demonstrated with the incorporation of tryptophan (Trp) used as an example.[11] Prior to the actual incorporation, the tryptophanyl-tRNA synthetase (TrpRS) is essential for the condensation reaction of Trp with ATP while forming a tryptophan adenylate (Trp ~ AMP) and releasing PPi.[27,28]

TrpRS + Trp + ATP → TrpRS:Trp ~ AMP + PPi
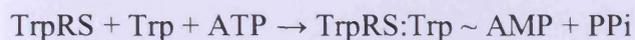
In activated Trp ~ AMP, carboxyl group of Trp is linked to the phosporyl group of AMP (Figure 3).
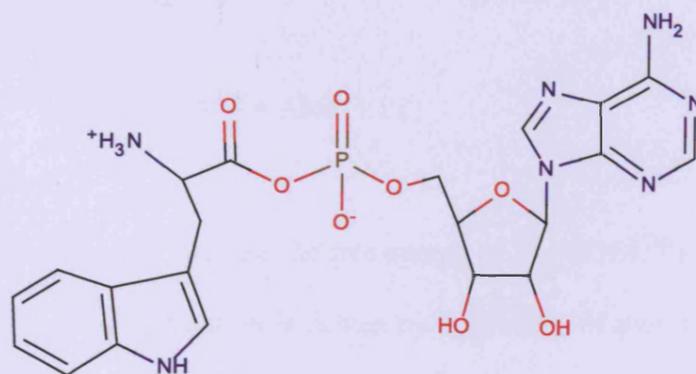
Figure 3: Structure of Trp~AMP. Trp~AMP is a mixed anhydride that is formed after condensation of Trp with ATP.

After Trp is activated, incorporation of this amino acid is accomplished through a two-step reaction. In the first step, TrpRS which is bound to tryptophan adenylate (TrpRS:Trp ~ AMP) charges tryptophanyl-tRNA (tRNA^{Trp}) with Trp to generate a Trp-tRNA^{Trp}.[28]

$$TrpRS:Trp \sim AMP + tRNA^{Trp} \rightarrow Trp\text{-}tRNA^{Trp} + AMP + TrpRS$$
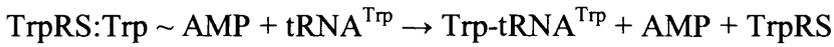
Trp-tRNA^{Trp} is in the second step brought to the aminoacyl (A) site of the ribosome by a translation elongation factor Tu.[11] The A site is a place of peptide bond formation.
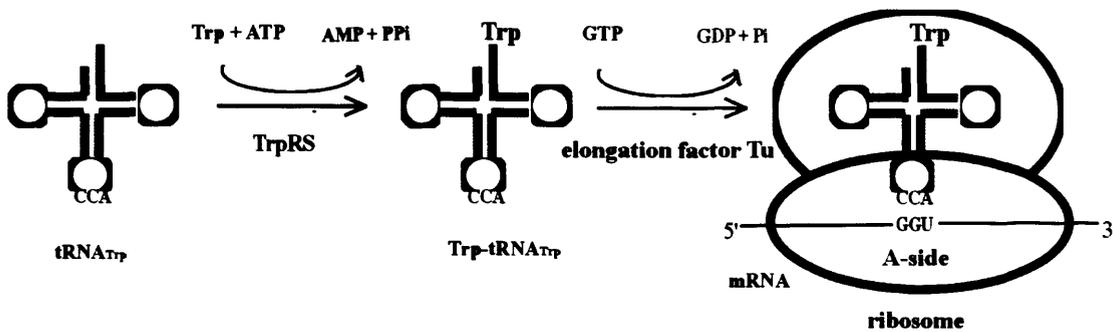
Figure 4: Mechanism of Trp incorporation.[11]

The overall aminoacylation reaction is:

$$Trp + ATP + tRNA^{Trp} \rightarrow Trp\text{-}tRNA^{Trp} + AMP + PP_i$$
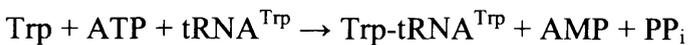
$\Delta G°$ of this reaction is close to 0 because the free energy of Trp-tRNA^{Trp} hydrolysis is similar to the free energy of ATP hydrolysis. The reaction is driven by hydrolysis of pyrophosphate and therefore is exergonic. In total, equivalent of two ATP molecules are consumed to form Trp-tRNA^{Trp}. 1 ATP is used to form the ester linkage between Trp-tRNA and 1 ATP is used as a driving force of the reaction.[29]

2.2   General features of aminoacyl-tRNA synthetases

It is important to note that even though AARSs provide the same aminoacylation reaction and use the same amino acid-like substrates, they display a high degree of diversity[13] and dissimilarities in size, amino acid sequence, and their three dimensional structures.[30] Using *Escherichia coli* (*E. coli*) as a model organism, it was found that the quaternary structure of synthetases follows up from monomers *e.g.* isoleucyl-tRNA synthetase (IleRS) to tetramers alanyl-tRNA synthetase (AlaRS). In terms of primary structure, the length can range from as short as 334 amino acids (TrpRS) to as long as 1112 amino acids for the phenylalanyl-tRNA synthetase (PheRS). In native molecular mass we can find synthetases from 51,000 Da cysteinyl-tRNA synthetase (CysRS) to 384,000 Da (AlaRS).[25] Another variability relates to the number and structure of subunits in the synthetase structures. In this manner, a synthetase could form a monomer α, homodimer α_2, homotetramer α_4, or heterotetramer $\alpha_2\beta_2$. Based on that, synthetases can be separated into four groups.[13]

Sequence data based searching techniques[31] have shown that 20 different AARSs are related to each other. Based on X-ray crystal structure,[30] those were divided into two classes of 10 enzymes each: class I and class II (Table 1). The class distinction was firstly derived from crystal structure similarities of SerRS[32] with aspartyl-tRNA synthetase/aspartic acid tRNA complex (AspRS:tRNA^{Asp})[33] and homology of GlnRS[34] to methionyl-tRNA synthetase (MetRS)[35] and tyrosyl-tRNA synthetase (TyrRS).[36,37] Class determination is based mainly on the active-site architecture shared by all members of the same class,[38] but also on additional domains attached to the central active-site domain with functions other than catalytic function. The type of an amino acid which is to be attached by synthetase to tRNA is crucial.[26] AARSs of two classes undergo conformational changes of the active site in different manners while binding to a nucleotide substrate. In case of synthetase and tRNA recognition, the main identity determinants are distributed along tRNA, mainly the anticodon loop and the amino acid acceptor stem.

Each of the classes is further divided into three subclasses based on their tRNA acceptor stem recognition. These subclasses are summarized and named alphabetically as Ia, b, c, and IIa, b, c.[39] There

is a similar manner of rotation and orientation of a synthetase around the tRNA acceptor stem axis between Ia and IIa, or Ib and IIb. TyrRS (Ic) and PheRS (IIc) are unique in the orientation of tRNA binding and differ remarkably from all other synthetases of their classes. That also indicates how synthetases couple according to the stereospecificity of an amino acid they need to recognize. In case of lysyl-tRNA synthetase (LysRS), two distinct types were found: first one is a typical class I synthetase but the second has characteristic II class synthetase features.[40]

Class I synthetase members are monomers (Ia and Ib) or homodiamers (Ic). It was shown earlier that only the MetRS has originally four subunits but is converted into a monomer active form by limited proteolysis.[41] Class I also displays two short common consensus sequences 'HIGH' histidine-isoleucine-glycine-histidine (His-Ile-Gly-His)[42] and 'KMSKS' lysine-methionine-serine-lysine-serine (Lys-Met-Ser-Lys-Ser)[43], which indicate the presence of a structural domain, the nucleotide-binding Rossman fold (a parallel β-sheet nucleotide-binding fold[26]) of the active site.[30] HIGH and KMSKS are positioned right in the place of ATP binding.[26]

Class II synthetases (dimers or tetramers) form usually an antiparallel β-sheet surrounded by α-helices and are characteristic for their three conserved motifs. Motif 1, which always involves a proline residue (Pro) in the intersubunit contacts, is formed by a long α-helix linked to a β-strand and is involved in a dimer interface. Motif 2 contains two antiparallel β-strands linked by a long loop and in case of Motif 3 a β-strand is directly associated with an α-helix. Motifs 2 and 3 form a part of the active site and contain an arginine residue (Arg).[26, 30]

On a molecular level, all amino acids structurally display an $-NH_2$ group followed by C_α and a -COOH group as a central core. But depending on their side chain we find one subgroup of the class I AARS capable of the activation only in presence of tRNA (Glu, Gln, Arg)[44] and another subgroup specific only for aromatic amino acids (tyrosine Tyr, Trp). Whereas, the rest of I AARS charge aliphatic or sulphur containing amino acids (valine Val, Ile, leucine Leu, Cys, Met). In one of the subgroups of class II AARS, the consensus sequence is significant at their amino termini and they activate a tRNA

with charged amino acids (aspartic acid Asp, Asn, Lys). On the other hand for some of the II class AARS', sequence similarities are at their carboxy termini and specificity is towards small and polar amino acids[15] (Gly, Pro, threonine Thr, His).[26]

Two classes were separated based on how each of the synthetases charges a tRNA with an amino acid. Class I aaRS would be selective towards 2'OH of a terminal adenosine of tRNA while class II uses 3'OH (except of PheRS which attaches Phe on 2'OH).[26]

Table 1: AARSs (class I and class II). Structural and functional differences of two classes.[11, 30]

Characteristics	Class I	Class II
Members (subunit composition)	*Subclass Ia* ArgRS (α), ValRS (α), IleRS (α), MetRS (α, α₂), CysRS (α), LeuRS (α, αβ)	*Subclass IIa* SerRS (α₂), ThrRS (α₂), ProRS (α₂), GlyRS (α₂), HisRS (α₂), AlaRS (α)
	Subclass Ib GlnRS (α), GluRS (α), LysRS1 (α)	*Subclass IIb* AspRS (α₂), AsnRS (α₂), LysRS2 (α₂)
	Subclass Ic TrpRS (α₂), TyrRS (α₂)	*Subclass IIc* PheRS [(αβ)₂], AlaRS (α₄), GlyRS [(αβ)₂]
Conservative motifs*	φHφGH, TIGN (TrpRS) KMSKS parallel β-sheet (Rossmann fold)	(1) Motif 1 (2) Motif 2 (3) Motif 3
Active site topology	parallel β-sheet (Rossmann fold)	seven-stranded antiparallel β-sheet
Position specificity	2'-OH group	3'-OH group (PheRS is exception)

* Conservative residues are indicated with capital letters, φ is any hydrophobic residue.

In chapter 3, our focus is on two Ic synthetases; TyrRS and its close structural homologue, TrpRS. Primarily, we want to compare them in order to relate their substrate specificity mechanisms.

2.3   Key features of aminoacyl-tRNA synthetase fidelity

Since the study of synthetase fidelity is our main objective, here we explore the synthetase active site and the mechanism by which this fidelity is accomplished.

Although synthetases have large structural differences in their subunit sizes and quaternary structures, the aminoacyl adenylate is recognized by sequences located in the N-terminal half of the proteins. On the other hand, tRNA recognition relies on both, the N-terminal half and the C-terminal half of the sequences.[27] In terms of the aminoacyl adenylate binding specificity for all synthetases, homologous residues are in charge of recognition of ATP, $-NH_2$ and -COOH groups arround the C_α. Different unique residues would recognize specific features, such as the side chain. For instance, class I synthetases are characteristic for their HIGH[42] and KMSKS[43] motifs, which specifically recognize ATP. We know from MetRS and TyrRS crystal structure complexes with ATP that both histidine residues from HIGH play an important role in substrate activation by stabilizing the ATP phosphoryl groups in the transition state of the activation.[45,46] Also in TyrRS, two lysine residues from KMSKS are stabilizing transition state, during tyrosine adenylate (Tyr-AMP) formation by strong interaction with PP_i moiety, similarly as in GlnRS.[34, 46, 47]

All synthetases have developed high specificities towards their natural substrates. Therefore, protein expression can be accurate.[11,48] The selection of one correct amino acid from a reaction of D- and L- amino acid optical isomers, precursors of amino acid biosynthesis or products of amino acid degradation is a challenge for a synthetase.[49] Charging of an incorrect amino acid and its incorporation would greatly affect cell survival. Different mechanisms providing correct aminoacylation have been uncovered while studying synthetases.[50-52] That does not include only recognition itself, but also editing (or proofreading), gene duplication and the use of alternative biosynthetic pathways. The degree of synthetase fidelity for a natural amino acid depends on charge and on how much the amino acid differs from its competitors in size so that binding energy is higher and binding tighter.[48] For example, although Asp and Glu share some similarities with Asn and Gln, their corresponding synthetases are able to discriminate them based on additional restrictions. Further, the discrimination mechanism for some aliphatic amino acids such as Val, Ile and Leu is different. Since these display large similarities between each other and there are only small differences in their potential binding energies, an additional

synthetase editing mechanism (double-sieving mechanism) is essential.[53] For valyl-tRNA synthetase (ValRS), IleRS, leucyl-tRNA synthetase (LeuRS), an extra CP1 domain insertion has been found[54] to recognize and hydrolyze AA-AMP and misacylated tRNAs.[55-58]

The active site hydrogen bonding is crucial for synthetase fidelity. This intermolecular noncovalent interaction occurs between two electronegative atoms that share a proton.[59] For instance, in $C - H \cdots O$, H is attached to an electronegative C atom (hydrogen–bond donor) and a hydrogen bond is formed between the positive charge of H and an electon lone pair of the O atom (hydrogen-bond acceptor). This well-established[60] dipole-dipole interaction is commonly present in protein and nucleic acid structures. Hydrogen bonding is the main structure determinant and is the key element of protein folding and stability.[61] On the other hand, an important hydrogen bonding interaction does stabilize the substrate transition state and influences the rate of the enzymatic reaction. According to a classical view, it has been well studied that complementary hydrogen bonds and salt bridges as well as hydrophobic interactions[53] are the main elements of synthetase specificity.[62]

2.4 Common ways to study AARS's fidelity

The main theme to this work is investigating the fidelity of AARSs. Thus, it is important to understand first how this property has been defined and studied for more than 40 years. The fidelity of a synthetase has historically been determined by various *in vitro* techniques.[63] Although there are a number of high resolution crystal structures that allow us to observe, on the molecular level, how a number of AARSs achieve fidelity, standard mutagenesis and biochemical assays were previously able to piece together the important AARS residues taking part in an amino acid and tRNA discrimination.

One standard technique to determine AARS' fidelity involves standard steady-state pyrophosphate exchange and aminoacylation assays to determine the enzyme's kinetics.[63] Specifically, they measure substrate binding and product release, which helps to characterize enzyme activity.[64] These methods

allow taking a small amount of samples for easy and fast manipulation. Furthermore, large numbers of synthetase-tRNA samples can be tested and compared.[65] Pyrophosphate exchange assay measures catalytic activity of synthetase-like enzymes forming enzyme-adenylate intermediates.[66] In this assay, we can follow the rate of exchange of radioactive PP_i into ATP. In such an assay, one typically reports two parameters, turnover number (k_{cat}) and Michaelis constant (K_M). The important feature of such an assay is that it allows one to easily compare one AARS to another although it may be a mutant or even from a different organism.

For example, reaction rates of AARSs, in mixture with their cognate amino acids, $[\gamma\text{-}^{32}P]ATP$ (gamma-32P ATP) and inorganic phosphate, were measured[67] by pyrophosphate exchange method and their stoichiometries determined (etc. TyrRS from *Bacillus stearothermophilus, Bst* TyrRS, binds only one aminoacyl adenylate per dimer). On the other hand, the aminoacylation assay evaluates the rate of AA-tRNA^{AA} formation. This assay was used to determine the efficiency of a wild type (wt) and mutant TyrRS's from *Methanococcus jannaschii* (*Mj* TyrRS) in tRNA^{Tyr} recognition.[68] Moreover, this enabled finding the important tRNA recognition residue (Asp^{286}) in *Mj* TyrRS structure. In another example,[69] the aminoacylation assay helped to determine the main Trp recognition residue (Asp^{132}) that also discriminates against Tyr, in *Bacillus stearothermophilus* TrpRS (*Bst* TrpRS) active site. Different studies have found K_M for TrpRSs from various organisms. The Michaelis constant of the aminoacylation reaction catalyzed by human TrpRS-6×His showed a $K_M = 7.4$ μM with L-Trp.[70] Aminoacylation assays were also performed in order to determine the kinetic parameters for wt *Bst* TrpRS and its variants. $K_M = 1.6 \pm 0.1$ μM for the wt synthetase.[69]

Furthermore, adenylation and aminoacylation have been studied using fluorescence approaches.[63] For the fluorescence studies, the presence of an intrinsic or extrinsic fluorescence probe, as Trp, is required. Binding AARS with its substrate is usually accompanied by a change in fluorescence of the synthetase. In this way, one can study synthetase interactions, structure and folding. In this manner, 5-

OH Trp is incorporated into oncomoduline[71] and changes in 5-OH Trp fluorescence enables one to study interactions with anti-oncomodulin antibodies.

Pre-steady-state kinetic methods measure the formation and consumtion of enzyme-substrate intermediate in the first milliseconds, until the steady-state is reached. These methods have been used for more detailed analysis of synthetase and tRNA single residues in specific steps of a reaction.[63] Out of these methods, the most common is the rapid chemical quench assay. This assay measures rapid reactions that cannot be monitored by absorbance or fluorescence spectroscopy. The rapid quench method directly measures radioactively labelled reaction product. For example, by using radioactivate probes, a cognate synthetase misacylation rate between Val and Ile was estimated to be as high as 1 in 5. Nevertheless, this rate was found to improve to 1:3000 due to a synthetase editing activity.[72, 73]

Other methods have been used to characterize AARS specificity. In our work, we use isothermal titration calorimetry (ITC). This is a good alternative to determine binding parameters[74] and therefore specificity of a synthetase towards its substrate. ITC enables us to measure a released heat of reaction between protein and ligand. This gives the information about the binding equilibrium. Greater binding responds to lower value of dissociation constant (K_d). ITC was used for binding studies of the human T2 TrpRS with two tryptophans, which are part of the VE-cadherin.[75] This experiment helped to understand more about expanded functionality of the synthetase active site.

The *in vivo* complementation assay is another method we use in this project to investigate the specificity of synthetase. In this approach, a bacterial strain struggles to grow on a media deficient for an amino acid. Using such *E. coli* strain that required added Trp to grow well, Söll and coworkers[69] studied a number of TrpRSs. This strain contains a mutation of TrpRS gene that destabilizes the synthetase structure and affects synthetase activity. Synthetase can only function in the presence of excess substrate. In this manner, different synthetase mutants expressed in the strain cell can be tested for their fidelity if they are able to complement the auxotrophic phenotype under Trp-starved conditions. Specificity of

several TrpRS mutants were tested by complementation assay.[69] Growth of the Trp auxotrophic strain

was only restored if an active mutant TrpRS was provided.

3. Introducing TyrRS and TrpRS

Although there are several classes of AARSs, each with its own specific nuances of mechanism, here we focus soley on the TryRS and TrpRSs. This is because these two AARSs have been used to incorporate the vast majority of all the NAAs incorporated to-date. Furthermore, we can review their background together because it is likely that these two enzymes share a common ancient ancestor, and as such concepts about the mechanism and fidelity of one AARS are often applicable and insightful for the other AARS.

From extensive literature, we know a lot about many TyrRS structures and their active sites.[46] It is the most used synthetase for nonnatural amino acid incorporation.[4] On the other hand, TrpRS displays high similarity to TyrRS structurally, therefore we can learn much from TyrRSs and apply it for the study of TrpRS. This is important because the vast majority of nonnatural amino acids have been incorporated using TyrRS mutants, and it is our contention that if we better understand TrpRSs, then it may be possible to incorporate NAAs with TrpRSs that might not be possible with the more common TyrRS systems.[76]

Thus, one of the main objectives of this study is to explore the usage of a *Bacillus subtilis* TrpRS (*Bs* TrpRS) wt and one mutant in expanded genetic codes.

3.1   Exploring of TyrRS active site

The active site substrate binding has been well studied on TyrRS.[46] This synthetase is also the best explored example in synthetase/amino acid discrimination.[53] In 1982, very little was known about the catalytic mechanism and the energetics of TyrRS or any other synthetases but within five years, the situation changed thanks to site directed mutagenesis, crystal structures and powerful kinetic tools. At this point, the major focus was the importance of binding energies in enzyme

catalysis reaction. In order to understand this reaction completely, we need to know all interaction energies between the enzyme and its substrate. Starting from binding energies, we can get an insight into how the activation energy is lowered and how equilibrium constants fit in, as well as learning how the specificity is determined by the enzyme. Binding energies were defined as "differences between ligand and receptor bound to water and ligand and receptor bound to each other".[46] Early work done on TyrRS helped to investigate methods for quantifying the strengths of enzymatic interactions and how they participate in catalysis and specific determination.[46]

Fersht and co-workers,[77] analyzed how the hydrogen bond is important in enzyme substrate specific binding. As an experimental system, they used the TyrRS from *Bacillus stearothermophilus*, the structure of which was at that time already known. In an earlier work, they indicated eleven possible hydrogen bonds bound to the aminoacyl adenylate in the active site, from which eight amino-acid side chains underwent a site directed mutagenesis (Asp^{176}, Tyr^{34}, Asp^{38}, Tyr^{169}, Gln^{195}, His^{48}, Thr^{51}, Cys^{35}).[36] Additionally, some other important hydrogen bonding residues were indicated, and divided into subsites (Figure 5).[46] Some of these are hydrogen bond acceptors and others hydrogen bond donors.
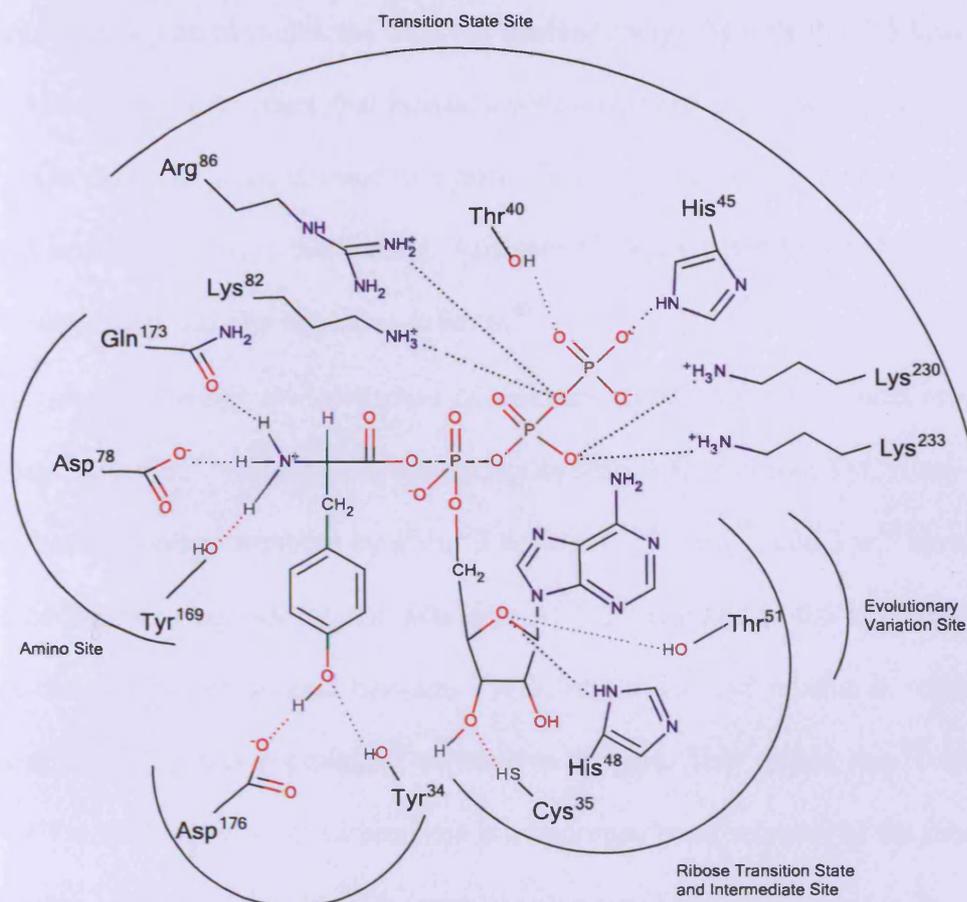
Figure 5: TyrRS active site residues divided into five subsites. Tyr-AMP complex is shown in its transition state and hydrogen bonds interacts between the substrate and important amino acid chains.[36]

With a choice of target residues for mutagenesis, one has to consider many factors. These mutations must not disrupt, or largely change, the protein structure. Ideally, if we introduce a mutation, a side chain is replaced so the size of the cavity is altered but the structure of an enzyme is able to tolerate this change.[46] This "nondisruptive mutation" effect[78] was proven on crystal structures of some mutants[79-81] as well as from kinetic data.[82, 83] Every mutation resulted in modification of a hydrogen bond, which helped to define its importance. The effect of the active site mutations was energetically measured in the activation reaction of Tyr and the mutant enzyme was compared to the wt enzyme. Such comparisons give direct information about a hydrogen bond importance in specificity and catalysis.[83, 84] For instance, deletion of a side chain that forms a hydrogen bond with

an uncharged group on the substrate weakens binding energy by only 0.5-1.5 kcal/mol. Nevertheless, deletion of a residue side chain that interacts with a charged group weakens binding energy by 3-6 kcal/mol. On the other hand, deletion of a group forming long hydrogen bond appends to an opposite effect and actually improves the binding. Analysis of the mutation effects helped to describe TyrRS subsites, interacting with the ligand, as follows.[46]

The tyrosine binding site comprises of residues specific for amino acid binding. Distinctively, Tyr^{169}, Asp^{78} and Gln^{173} form their hydrogen bonds with $-NH_2$ group of Tyr. If any of these residues is mutated, binding energy weakens by about 3 kcal/mol.[77, 85] Asp^{176} and Tyr^{34} have bseen indicated as residues of Tyr site chain specificity. Mutation of Tyr^{34} resulted in 0.5 kcal/mol energy loss, which indicates that the hydrogen bond between Tyr substrate and Tyr residue is relatively weak. On the other hand, Asp^{176} mutation produced an inactive enzyme. That makes Asp^{176} an extraordinary key residue of Tyr specificity, as its carboxylate is a hydrogen bond acceptor of the substrate -OH group.[86] Fersht further observed[46] that Tyr^{34}Phe mutation decreased discrimination for Phe substrate. Although only one -OH group is removed because of this mutation, the effect on the specificity of Tyr/Phe recognition is significant. As was mentioned before, the energy of the hydrogen bond linking -OH of Tyr^{34} and a lone electron pair of the substrate Tyr oxygen is quite low. This is because Tyr ligand with its -OH group is a poor hydrogen bond acceptor. Therefore, K_M for the Tyr^{34}Phe mutant in activation of Tyr increases only by a factor of two and the specificity for Tyr decreases by 15 fold. The importance of two residues Tyr^{34} (hydrogen bond donor) and Asp^{176} (hydrogen bond acceptor) was demonstrated on the wt TyrRS in an activation reaction of Phe.[77] Usually, if the wt enzyme is not associated with a ligand, these two residues are hydrogen bonded to a water molecule. These bonds are broken when Tyr comes to the active site and new hydrogen bonds form between its -OH group and Tyr^{34} and Asp^{176}. If Phe accommodates the active site, the binding energy is reduced by about 7 kcal/mol.[48] The reason is probably that the water molecule remains in between Tyr^{34} and Asp^{176}, and this is causing some unfavourable interactions with the substrate for lack of a space. In summary, for

the TyrRS to switch its specificity from Tyr to Phe substrate, two key residues Tyr[34] and Asp[176] need to be mutated.[77] Normally, because of the great difference in initial binding energies of Tyr and Phe, TyrRS does not need any editing mechanism, as Phe is activated 1.5×10^5 times less efficiently than Tyr.[48] This makes TyrRS highly specific towards its substrate Tyr.[77]

The TyrRS binds not only the amino acid chain, but also has a special region to bind the adenylate molecule. *The ribose binding site* has three important residues, Cys[35], Thr[51] and His[48]. These three contribute to stabilization of the enzyme:Tyr-AMP transition state. The result is that the binding energy of these relatively distant groups is used in catalysis to lower the energy difference between the ground and transition state.[62, 87-89]

The ATP phosphates and pyrophosphate amino acid site require residues His[45] and Thr[40] which clearly bind and stabilize the gamma-(λ)-phosphate of ATP in the transition state and PP$_i$ in the enzyme:Tyr-AMP·PP$_i$ complex.[86, 87, 90, 91] Different residues, Arg[86] and Lys[230] also bind to the ATP in the transition state and to the PP$_i$, while Lys[82] and Lys[233] interact with the enzyme:Tyr·ATP complex.[92]

Thr[51] is a nonconserved residue and therefore represents *the evolutionary variation site* of *Bst*TyrRS.

3.2 Similarities between bacterial TyrRS and TrpRS

More than 15 years ago, it was observed that TyrRS and TrpRS I class synthetases are more linked to one another than any members of the same class.[15, 37, 93] Because of the close homology, two synthetases were assigned to the Ic subclass. Carter and co-workers (1995) solved the X-ray crystal structure of *Bst* TrpRS in complex with tryptophanyl-5'AMP (Trp-5'AMP) to 2.86 Å resolution and compared it to the structure of *Bst* TyrRS:tyrosyl-5'AMP (Tyr-5'AMP).[85] The sequence identity between TyrRSs and TrpRSs is about 10-20%. Similarly, we can find the case of two TyrRS

orthologs from eukaryotic and bacterial organism or for two TrpRS orthologs.[94] The crystal structure alignment of both structures indicates a high identity of tertiary structure and when comparing two of the sequences, the highest significance is found in their amino acid activation region (Figure 6).[95] In particular, the TrpRS and TyrRS active sites interact similarly with adenine, ribose, -NH_2 moieties of ATP as well as with the side chain of Trp or Tyr. In order to understand the mechanism of TrpRS specificity and catalysis, this similarity with TyrRS enables us to couple these two structures.

3.3   TrpRS molecular determinants compared to those from TyrRS

After having so much structural and kinetic information about *Bst* TyrRS, the question arises: How does the relation between TyrRS and TrpRS helps us understand how TrpRS gains its fidelity toward its substrate?

Before the crystal structure of TrpRS was known, only three other Class I synthetase structures were identified, and it was more difficult to find connections between synthetases, their mechanism of specificity and catalysis from data that was available. This was due to significant structural differences. Specifically, it was the structure of GlnRS with ATP and tRNA^{Gln},[34] MetRS with ATP,[96] and TyrRS with tyrosyl adenylate and Tyr[85] known at the time. If we think about similarities between them, we can find a few important features. All three synthetases are united by the presence of their amino terminus Rossmann (dinucleotide) domain, with two characteristic consensus sequences, HIGH and KMSKS. HIGH motif contains Gly, which is a strictly conserved residue within all Class I synthetases and plays an important role in ATP adenine ring interaction.[95] In addition, the Rossmann domain contains an insertion in between its first and second symmetrical halves. However, this insertion differs in size for TyrRS and GlnRS.[34] Although TyrRS and GlnRS bind their substrates in a similar way, there are a number of active site unrelated interactions found for one enzyme but not for the other one.[34] Furthermore, some similarities were found in the C_α backbone of the TyrRS and

MetRS active site region but mostly TyrRS binds ATP in a different orientation than MetRS.[96] The variability of synthetases within the same class and the structure of their domains indicate diversity in the recognition of their cognate tRNA respectively.[97]

From Carter's comparison of *Bst* TrpRS and *Bst* TyrRS active sites (Figure 6),[95] interactions with the activated amino acid Tyr-5'AMP and Trp-5'AMP, were marked as almost identical. There is number of conserved residues that do not vary for both structures at all. Furthermore, TyrRSs and TrpRSs lack any sort of editing domain.
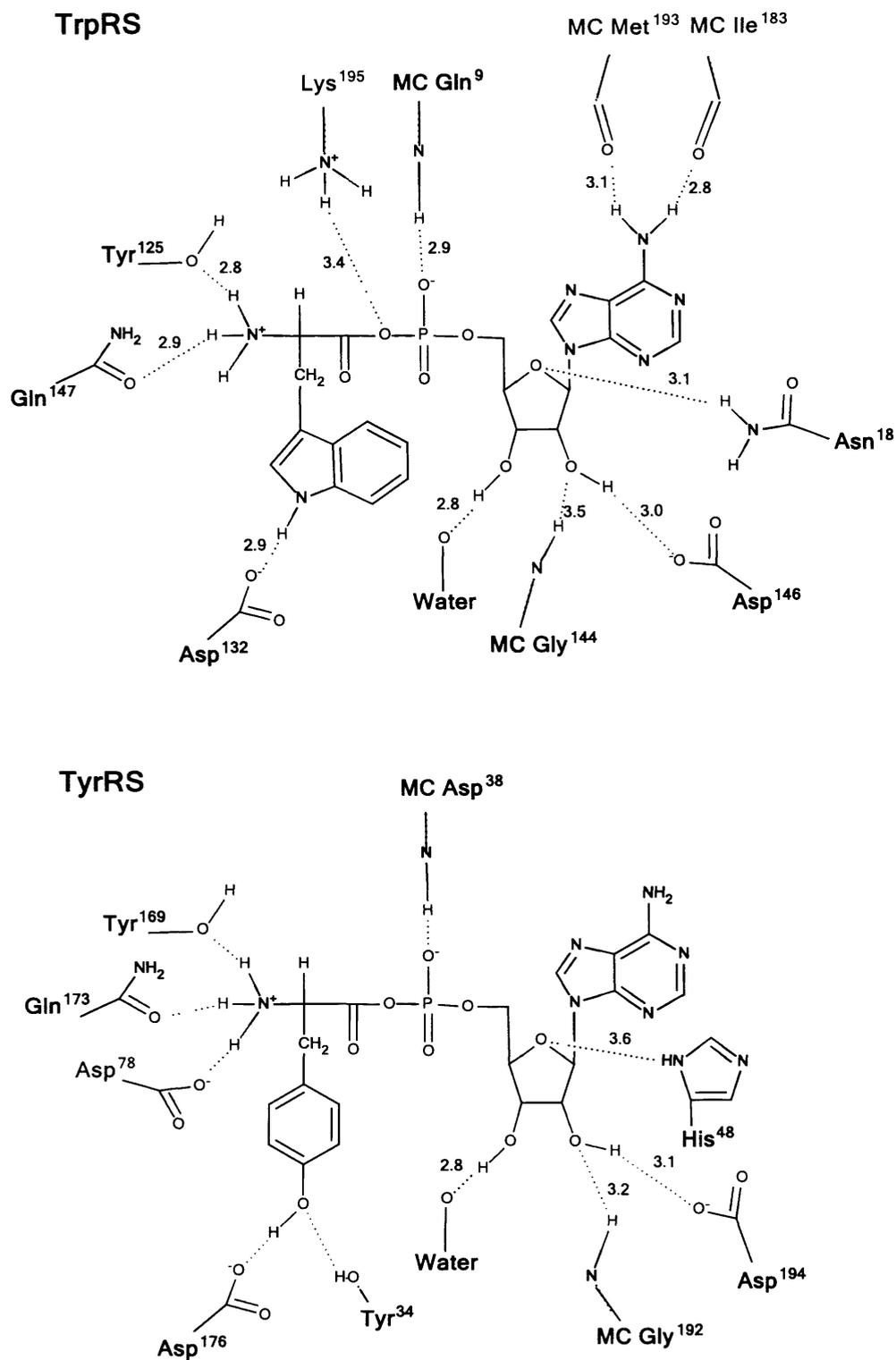
Figure 6: *Bacillus stearothermophilus* **TrpRS and TyrRS active site models.** Amino acid analogues are in purple and unique residues in green. MC = main chain. Distances are given in angstroms.[95]

Trp-5'AMP binds in the place of the Rossmann fold. The Rossmann fold, instead of a HIGH consensus sequence in case of TyrRS, comprises TIGN sequence in TrpRS. Gly[17] from TIGN is a conserved residue and plays a role in ATP adenine ring recognition. As was mentioned earlier, this Gly has analogues in all other I class synthetases in similar positions. Other conserved interactions with adenosine phosphate were indicated as follows. Gln[9] in TrpRS, which binds to the phosphate by its backbone amide, has its Asp[38] analogue in TyrRS. 2'-OH and 3'-OH groups of the ribose in TrpRS are attached to Asp[146], bound water, and the Gly[144] backbone amide consistently as Asp[194], water and Gly[192] in TyrRS. Asn[18] from the TIGN sequence of TrpRS interacts with ribose similar as His[48] in HIGH of TyrRS. Other than adenosine phosphate, the $-NH_2$ group of Trp also comprises interactions with Gln[147] and Tyr[125] identically to Gln[173] and Tyr[169] in TyrRS.

Despite these, there were some dissimilarities observed in the active site interactions. In TyrRS, Asp[78] hydrogen binds to the $-NH_2$ of Tyr. Nothing similar is obvious in TrpRS. On the other hand, Met[193] and Ile[183] are interacting with the adenine 6-amino group with their carbonyl oxygen's in the active site of TrpRS but are not found in TyrRS. Futhermore, in the TrpRS, Lys[195] is hydrogen bonded to the α-phosphate of the adenylate. This Lys is part of the KMSKS loop, and needs to move ~ 7.5 Å in order to interact with the active site.[95] On the other hand, distance residues in TyrRS: Lys[230], Lys[233] on one side of the pocket, together with Lys[82] and Arg[86] on the other side, are involved in tyrosyl adenylate formation.[47]

However, active site of two TyrRS and TrpRS is highly similar to each other, especially residues that recognize and specifically bind the amino acid. It was shown that residues Tyr[125], Met[129] and Asp[132] form a specificity-determining helix in TrpRS, and are related to similar residues in an analogous TyrRS helix of Tyr[169], Gln[173] and Asp[176]. For both synthetases, Asp[132]/Asp[176] is a key residue of the side chain. The hydrogen-bond length of Asp with the side chain of the substrate is ~ 2.8 Å for both synthetases. If Trp binds to the TyrRS, it leaves an unfavourable gap of 4 Å. On the other hand, Tyr binding to the TrpRS would result an unfavourable Van der Waals distance of ~ 1 Å.

The superposition of two TyrRS and TrpRS active sites indicated a slight, but important difference in Asp^{176}/Asp^{132} residues position. Asp^{132} is placed closer to the Trp indole nitrogen with its -COOH group. Pro^{126} and Pro^{127} are important to provide this Asp^{132} recognition helix bend in the TrpRS active site. In terms of the side chain accommodation, Gln^{189} in TyrRS is replaced by Val^{141} in TrpRS so that there is an extra space for Trp indole ring. As mentioned, in TyrRS the Tyr^{34} is participating with its hydrogen bond to the substrate Tyr -OH group. One of the facts Carter and co-workers showed in detail was that for TrpRS/TyrRS, the main substrate identity determinants are identical: Asp^{132}/Asp^{176}. For the TrpRS to switch its specificity towards another amino acid, Asp^{132} or two prolines holding the Asp^{132} in a right position (Pro^{126} and Pro^{127}) need to be replaced.[95]

The importance of the hydrogen bond interaction between Asp^{132} and Trp nitrogen hydrogen of indole in TrpRS was further tested.[98] Replacement of the nitrogen by other heteroatoms (selenium, sulphur, of oxygen) resulted in translation inactivation. At the same time, the introduction of extra nitrogen close to the protonated nitrogen of indole did not result in a substrate, which was recognized (2-Aza-Trp). When position seven of Trp is replaced by nitrogen (7-Aza-Trp), this substrate is recognized and used by TrpRS.

3.4 Analysis of bacterial TrpRS: structure and mechanism of binding its substrates

For our purposes, it is best to focus on specific bacterial TrpRS structure and mechanism. Whilst there are of course human TrpRSs, the way they achieve fidelity is different from bacterial TrpRSs. Furthermore, expanded genetic codes are almost exclusively based on engineering or evolving bacterial TrpRSs and as such, they form the focus of our analysis.

Bs TrpRS as a member of the Ic subclass has a primary function to esterify 2'OH of tRNA ribose by Trp and produce Trp-tRNA. It is a homodiameric α_2 enzyme that contains the smallest subunit chains among known AARSs (330 amino acids long),[99] it has been considered as an ideal candidate

for studying structure function by means of site-directed mutagenesis.[95] By comparing amino acid sequences, *Bs* TrpRS is 55.8% homologous to that of *E. coli* TrpRS (*Ec* TrpRS) and 78.1% to that of *Bst* TrpRS.[99] TrpRSs share some similar features and a precise description was offered on *Bst* TrpRS homodiamer crystal structure.[95] One monomer contains two unequal domains. A first central big domain comprises already mentioned the Rossmann dinucleotide-binding fold (1-200 residues). The Rossmann fold is made of parallel ß-strands with α-helix crossover connections, which form a binding pocket for Trp and adenosine. Figure 7 shows a position of ATP in between ßD and ßE strands and αA and αE helices.[24] Binding of the adenosine is afforded by TIGN and KMSKS motifs brought together by parallel ß-sheet twist.[37] Ribose is interacting with another identity determinant G×DQ (glycine×aspartic acid-glutamine), characteristic for Ic subclass. G×DQ is located at the N-terminus of αE.[24] The Rossmann region also contains an amino acid sequence specifically binding Trp-tRNA^{Trp} acceptor stem.[99] At the edge of the synthetase centre, a second small domain (SD) is built from four helixes (207- 280 residues), called a helix bundle. This is the region of high motion[95] but also an anticodon Trp-tRNA^{Trp} binding site.[99] One of the four helices expands along the whole monomer size and terminates at the dimer axis (265-326 residues).[95] C-terminal Met and Met,[92] that contribute in network of hydrophobic interactions, have the important roles in Trp pocket stabilization. Furthermore, Carter describes TrpRS domain movement while binding its substrates.[100] When adenylate binds[24] the SD together with the αA helix, TIGN, KMSKS and 175-182 residues move towards the active site in the Rossmann fold. After that, four chain segments (ßD, ßE, αA, αE) associate with the ligand. In a ligand free structure,[100] αA helix moves about 13° away from the centre (αE and G×DQ) interacting with the helix bundle to form SD. Moreover, two consensus sequences and 175-182 residues move together with SD helixes by about 4-6 Å and separates from the active site. In this manner, ATP binding site splits into two. Therefore, ligand free TrpRS structure is extended and larger. On the other hand, when Trp binds in the active site, four sides are involved: ßA, the specificity determining helix; αD; the mobile loop/helix 106-117; and the C-terminal helix of the

Rossmann fold. In a Trp free TrpRS, these four groups are slightly opened to allow Trp to enter. ATP

and Trp can access a TrpRS active site independently and therefore either of them can be the first one

to bind.
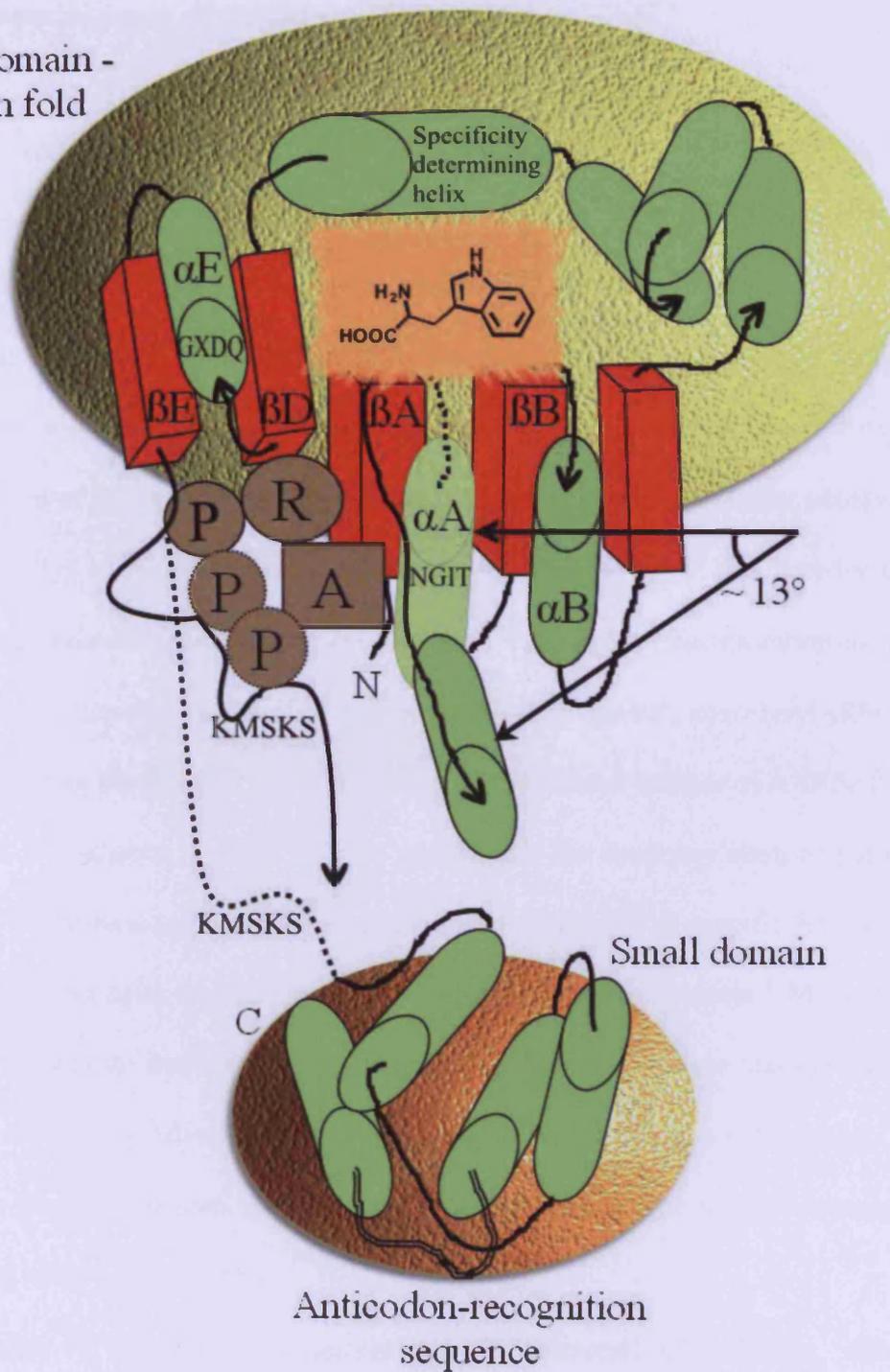
Figure 7: Carter's *Bacillus stearothermophilus* TrpRS monomer crystal structure model.[24]

3.5   Special cases of amino acid incorporation

As mentioned, there are now two well known, but rare instances of organisms using more that the common 20 amino acids. It is interesting and insructive for our work to note how these amino acids are incorporated into proteins by the host organism.

Incorporation of Sec and Pyl differs from the mechanism of the common 20 amino acids. Sec indirect pathway of acylation has already been descibed in chapter 1. Nevertheless, it was discovered that the incorporation of Sec was directed by the unique UGA codon.[17,2] A similar situation was found in the case of Pyl. Pyl is a 22nd amino acid, found in bacterial species[3] that was for the first time discovered in the *Methanosarcina barkeri (M. barkeri)*.[101] Since the incorporation of Sec,[18] as 21st amino acid, does not involve a specific AARS rather than SerRS, pyrrolysyl-tRNA synthetase (PylRS) was assigned as the 21st AARS.[102] Therefore, 21 is the total number of AARSs that have been discovered. PylRS was shown recently[102] to be specific for the aminoacylation of pyrrolysyl-tRNA (tRNA^{Pyl}) by Pyl in *Methanosarcinaceae*. Such charged Pyl-tRNA^{Pyl} is specific for the unique UAG codon that is part of an open reading frame of a methyltransferase protein.[3] Methyltranferases in general facilitate growth on methylamines as energy supplies. It's thought that the incorporation of Pyl gives some evolutionary advantage to the organism.[101] Furthermore, the *M. barkeri* Pyl-tRNA^{Pyl} pair was found to be an orthogonal pair in *E. coli*.[103] Interestingly, Sec and Pyl incorporations take advantage of using unique stop codons.

Additionally,[104] in some *Candida* species, the universal CUG codon, which normally encodes Leu, is reassigned to encode serine instead. Similar to these examples, using one of the unique codons (usually a stop codon) for a nonnatural amino acid incorporation *in vivo*, has become successful and popular in the sense of the expanded genetic code.[105,5] This indicates that nature had already been using benefits of the unique codon for an amino acid incorporation long ago (etc. 3 billion years ago for PylRS)[21] before the engineered *in vivo* approach came to attention. In this *in vivo*

approach, high specificity of a synthetase towards its substrates is achieved by a directed evolution of an 'orthogonal' tRNA synthetase/tRNA (AARS/tRNA) pair.[106] NAAs in the expanded genetic code will be described in the next chapter, as it is the main subject of this thesis.

4. Expanded genetic codes

Several methods of nonnatural amino acid incorporation have been developed in order to enhance properties of proteins by including additional NAAs in their structures. If such an amino acid is part of the peptide chain, it allows new ways to study the structure and function of proteins.[105] These methods have been divided as followed; *in vitro* or *semi in vitro* and *in vivo* methods (Table 2). For each of the methods, several advantages and disadvantages are defined.[107, 108]

For *in vitro and semi-in-vitro* manipulation, three common methods were assigned; chemical protein synthesis, chemical misacylation of tRNA and the import of AA-tRNA into eukaryotic cells. Within *in vivo* manipulations, we classify the use of amino acid auxotrophs and the use of orthogonal and new AARS/tRNA pairs. However, the primary focus of this work is on the usage of the last method (generation of an orthogonal and new AARS/tRNA pair) in the expanded genetic code. More specifically, we want to explore how the previously evolved[76] mutant *Bs*TrpRS is orthogonal and specific for the use of nonnatural amino acid incorporation.

Table 2: Methods of NAAs incorporation. Advantages and disadvantages of *in vitro, semi-in-vitro* and *in vivo* approaches.[107, 108]

Manipulation	Method	Advantages	Disadvantages
in vitro and semi-in-vitro	Chemical protein synthesis[109]	# incorporation of NAAs that are toxic to a cell or incompatible with translation # insertion of isotopic labeled amino acids	# complex experiments # low protein yields # limitation in size of a peptide # problems with solubility
	Chemical misacylation of tRNA[110]	# amino acid incorporation at single predetermined positions # cell free system	# complex misacylation experiment # poor yield of aminoacylation reaction # competition with host amino acids # limited capacity of a misacylated tRNA to fully decode a nonsence codon # stability of mRNA and a new protein
	Import of AA-tRNA into Eukaryotic cells[111]	# avoiding toxicity of a NAA # two different NAAs can be initroduced into the same protein # applicable for sensitive assays where a low amounts of modified protein is required	# complex misacylation experiment and further manipulations # poor yield of aminoacylation reaction
in vivo	Use of Amino Acid Auxotrophs[112]	# small number of genome-wide mutations # doesn't involve complicated manipulations # high yield and fidelity	# high level of substitution and toxicity # lack of control over natural substrate # requires using of a very similar nonnatural analogue
	Orthogonal and new AARS/tRNA pairs[106]	# high fidelity and efficiency # high protein yields as a result of controlled heterologous expression systems # position-specific incorporation # minimal suppresion-associated toxic effect	# bad acceptance of some "exotic" amino acids for the cellular permease system # complicated mutagenesis and selection cycles # metabolic toxicity

Chemical protein synthesis is one of the methods that enables a peptide to be chemically synthesised with a NAA in the structure. Two main approaches have been widely used in this method: classic solution synthetic chemistry and solid-phase peptide synthesis (SPPS)[113, 114] In both cases the size of the peptide that is synthesised is limited to 30 - 50 amino acids.[115] Other issues were described; mainly that purity and yield of the peptide were decreased. For synthesis of longer constructs, protected segments of a peptide can be coupled or a chemical ligation of nonprotected purified sequences is applied (Figure 8).[116] This method is more efficient and provides a synthesis of proteins $\leq$ 30 kDa. Since the chemical ligation uses the Cys residue at the peptide bond formation site, at least one Cys must be present at the ligation site.
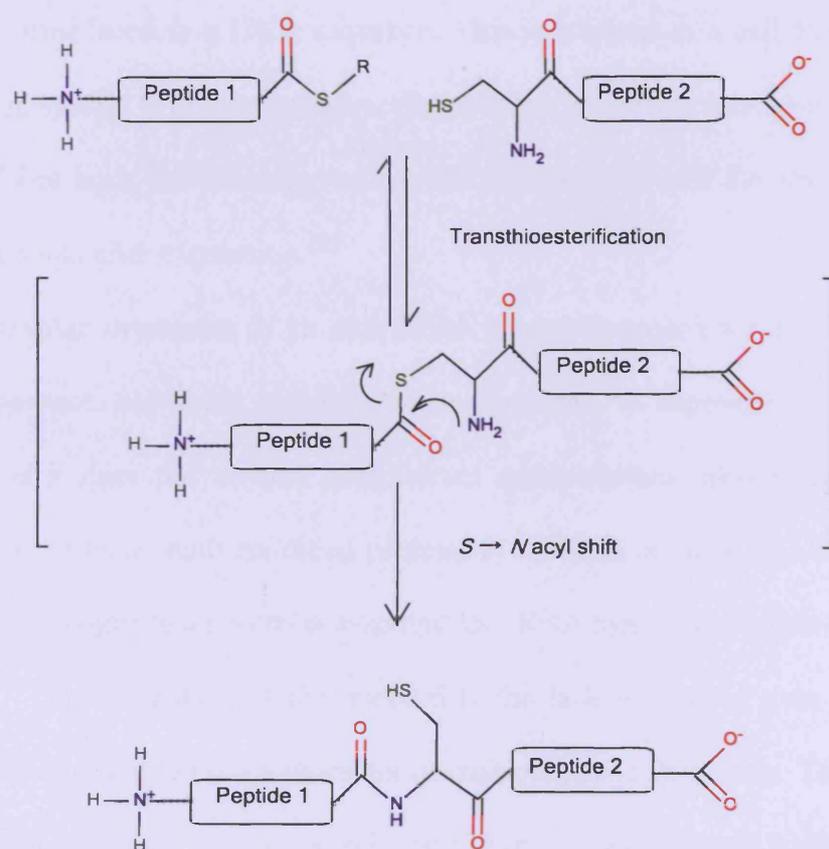
Figure 8: Chemical peptide ligation.[116] This method requires two unprotected peptides. Petpide 1 displays a C-terminal α-thioester and Peptide 2 an N-terminal cystein residue. In the first reaction – transthioesterification, peptide bond is formed between the N-terminal thiol (Peptide 2) and the α-thioester (Peptide 1) groups by nucleophylic attack of the thiol group. This reaction is followed by nucleophylic attack of the primary amine in $S \rightarrow N$ acyl shift. In this fashion, Peptide 1 and Peptide 2 are joined together.

In another approach, a chemically synthesised peptide containing a NAA is coupled to a biosynthetically produced protein. In this fashion, the position of substituent is limited only to N or C terminus.[117] The outstanding issues of chemical protein synthesis are in general limited by the size of the peptide and the extracellular environment. Additionally, the peptide synthesis by ligation reaction is restrained by a Cys residue of the peptide site.[105]

Different methods of the nonnatural amino acid incorporation are based on the process of translation. This is when tRNA molecules associate with the genetic code in mRNA and the amino acid chain forms. Through a general *in vitro* biosynthetic method, a NAA can be successfully incorporated into a peptide. In principal, tRNA is "chemically misacylated"[118] by a NAA and specifically recognizes one of the stop codons misplaced in a DNA sequence. This is attained in a cell free system.[7, 119] In a different approach, a cell system is used when misacylated tRNA is injected into *Xenopus* oocytes (*semi-in-vitro* method).[111, 120] For both, the formation of the tRNA associated with the amino acid is complex and there is low protein yield after expression.[105]

The use of auxotrophic organisms is an alternative *in vivo* approach for nonnatural amino acid incorporation. This approach has some benefits, when compared to chemical synthesis or *in vitro* biosynthetic methods as it does not involve complicated manipulations, also yields and fidelity are higher. Moreover, it enables us to study modified proteins *in vivo* and *in vitro*. One condition is that the NAA must be a close homologue to a natural one so that the tRNA synthetase is able to recognize it and misacylate the tRNA.[121] The limitation of this method is the lack of control over competing natural substrates. In addition, such a method has a character of multiple sites substitution. Therefore, the use of an auxotrophic strain helps to inhibit the natural substrate incorporation, because such strain is unable to synthesise it.[122] However, this method is limited for the requirement of a very similar nonnatural analogue.[105] Some analogues were incorporated after relaxation of tRNA synthetase fidelity through active site mutations[123] or by decreasing synthetase editing activity.[124]

The methods that are most relavent to the work in this thesis are the methods that rely on suppression by an orthogonal and new AARS/tRNA pair with nonsense, rare or 4 base pair codons has brought success for nonnatural amino acid incorporation into proteins. This method is able to provide high fidelity and efficiency and likewise more than 70 NAAs have been site-specifically incorporated into proteins *in vivo.*[119,125-129] That includes amino acids with novel functional group,[130-132] photocrosslinkers,[133] heavy atoms and redox active moieties. Three requirements of the method are: a unique (an orthogonal) tRNA-codon pair, corresponding aminoacyl-tRNA synthetase (an orthogonal synthetase), and a NAA. In principle, an orthogonal synthetase recognizes and aminoacylates a specific orthogonal tRNA with the NAA. Afterwards, this tRNA is brought to the ribosome and the NAA is inserted in frame with the growing peptide in response to the rarely used codon (Figure 9). The orthogonal tRNA must not be charged by any of the endogenous synthetases, just as the orthogonal synthetase must be precise in a specific recognition of only the orthogonal tRNA molecule. Besides, the orthogonal tRNA must be only charged by the NAA to the unique stop codon, and this codon must not be recognized by any other endogenous tRNA's. Furthermore, the orthogonal synthetase must only charge the tRNA with the NAA and not any of the 20 common amino acids. In the same way, the rare amino acid must not be a substrate for cellular synthetases. At last, the transport of the NAA needs to be cell permeable and must not be toxic or a target for degradation by cellular enzymes.[107, 134]
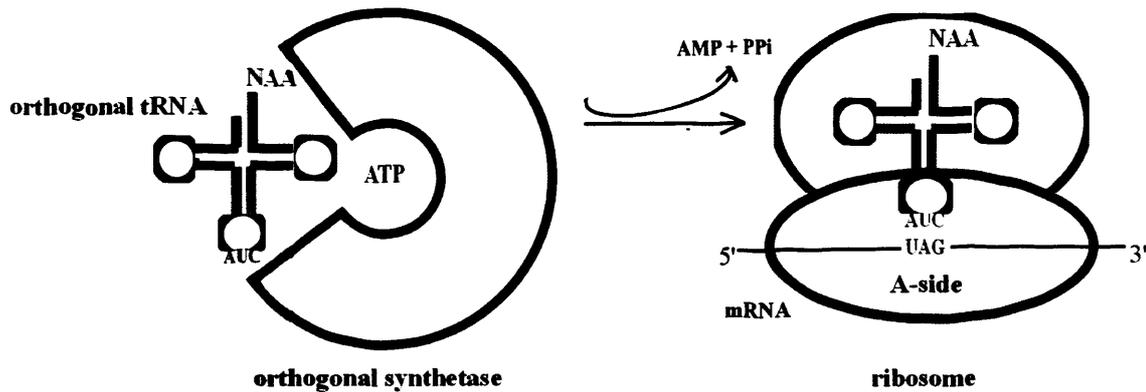
Figure 9: Nonnatural amino acid incorporation *in vivo*. The orthogonal synthetase specifically aminoacyletes the orthogonal tRNA with the NAA. tRNA with NAA are brought to the A-side unique stop codon of the ribosome. In this manner, NAA is incorporated into a peptide chain.

The generation of an orthogonal and new AARS/tRNA pair in the expanded genetic code has brought some challenges. Initially, *E. coli* and its AARS/tRNA (*Ec* AARS/*Ec* tRNA) pair were used as a target for nonnatural amino acid incorporation. However, this approach did not find complete success due to endogenous tRNA's misaminoacylations by designed *Ec* AARS.[135] Nonetheless, the same year a paper was published by Furter with a more effective approach.[136] In this work a yeast PheRS/tRNA[Phe] pair was used to incorporate *p*-fluoro-phenylalanine (*p*-F-Phe) in dihydrofolate reductase (DHFR) in *E. coli*. In principal, *p*-F-Phe as an external added amino acid, is activated by yeast PheRS in a *p*-F-Phe-resistant-*E. coli* strain and incorporated in place of a "blank" amber stop codon. The *p*-F-Phe-resistant-*E. coli* strain consists of *Ec* PheRS mutants that are able to exclude *p*-F-Phe substrate from the substrate binding. This incorporation was preferred (11-21 fold) to the endogenous Phe or Lys incorporations. The PheRS/tRNA[Phe] orthogonality, high specificity and a good protein yield after expression were a good start of the *in vivo* incorporation methods enhancement.

An important contribution that would allow many of the shortcomings in these early orthogonal systems to be overcome came from Schimmel and co-workers (1999). These workers carried out some studies of the archeal *Mj* TyrRS and found that due to the absence of a second domain, including the anticodon recognition site, the synthetase only had a minimal anticodon recognition ability.[137] This is

important because most synthetases, especially the ones in Schultz and Further's early studies, would have been detrimentally affected as a consequence of recoding the anticodon loop of the tRNA from the standard anticodon to an anticodon for a stop codon. This is because the tRNA anticodon loop is an important identifying element for most AARSs. Thus, Schimmel's results showed that this specific synthetase lacked any interaction with the anticodon loop, and this had a lot to do with the fact that *Mj* TyrRS are now the most widely used AARSs in the expanded genetic code. Another important discriminatory factor is the first base pair of the acceptor stem. In the case of the bacterial tRNA, it is G1:C72 base pair, whereas eukaryotic and archeal organisms have C1:G72 pair in the same place. For that reason, *Mj* TyrRS is able to recognize and charge eukaryotic tRNA^{Tyr} but not *Ec* tRNA^{Tyr}. Besides, it was shown that in the structure of tRNA the variable arm is another key discriminator.[138] As a result, Furter's and Schimmel's work strongly supported the idea of orthogonality as a powerful tool for the expanded genetic code.

A significant change in the method efficiency was brought by introducing an orthogonal AARS/tRNA pair from archaea *M. jannaschii* for use in the *E. coli* cells. The wt of this pair, chosen by Schultz a co-workers,[106, 125] is originally incorporating Tyr (*Mj* TyrRS/*Mj* tRNA^{Tyr}) and was chosen based on Schimmel's observations about anticodon loop recognition. When used in *E. coli*, there is no interaction between the orthogonal *Mj* TyrRS and *Ec* tRNA^{Tyr} or *Ec* TyrRS and *Mj* tRNA^{Tyr} but the new *Mj* TyrRS/*Mj* tRNA^{Tyr} pair is efficient in translation.[139]

In order to incorporate the NAA at a specific site, a choice of a suppression codon needed to be considered. Since translation in *E. coli* is usually terminated by two stop codons, UGA and UAA in 93%, UAG amber stop codon was an ideal choice to use for an analogue suppression. In this order, *Mj* tRNA^{Tyr} anticodon site was mutated to CUA so that it specifically binds to the UAG in mRNA.

To improve specific aminoacylation of *Mj* tRNA^{Tyr} by only *Mj* TyrRS, 11 tRNA nucleotides were selected and randomized by mutagenesis. Result of this was the generation of a suppressor tRNA library. These mutants were passed through rounds of negative and positive selection. In the negative selection,

cognate synthetase is not present and an amber suppression codon is located in the area of barnase gene of toxicity. If tRNA is aminoacylated by an endogenous *Ec* TyrRS then ribonuclease barnase is the product of barnase gene suppression. This product is lethal for cells so the only surviving colonies should contain appropriate suppressor tRNA, or non-functional tRNA. Cognate synthetase is included in positive selection for the remaining tRNAs that follows this negative selection. This time, an amber codon is present in ß-lactamase gene. When cognate synthetase specifically aminoacylates tRNA, ß-lactamase gene is suppressed and cell gains its resistance towards ampicilin. Cells with tRNA nonspecific to the cognate synthetase or containing a non-functional tRNA will lack antibiotic resistance. After selection rounds, resulting mutant tRNA (mutRNA$^{Tyr}_{CUA}$) is able to restore cell antibiotic resistance by providing full ß-lactamase suppression but is also preferably aminoacylated by *Mj* TyrRS as by *Ec* TyrRS. *Mj* TyrRS together with mutRNA$^{Tyr}_{CUA}$ (*Mj* TyrRS/mutRNA$^{Tyr}_{CUA}$), with a high fidelity and low background of unspecific suppression, is an ideal orthogonal pair.[106]

Besides that, cognate *Mj* TyrRS needs to be evolved into an orthogonal synthetase that specifically recognizes a NAA.[106] Considering the crystal structure of *Bst* TyrRS,[85] five active site residues important for ligand binding[140] were chosen to be randomised by mutagenesis and a library of all possible *Mj* TyrRS mutants (1.6×10^9) was created (Figure 10). Yet again, positive and negative rounds of selection were applied to obtain the final orthogonal synthetase. In positive rounds, cells are co-transformed with two plasmids containing different *Mj* TyrRS mutants from a library and mutRNA$^{Tyr}_{CUA}$. The amber stop codon is introduced into the chloramphenicol acetyltransferase (CAT) gene.[141] Cells are grown in the presence of a NAA and are tested for survival in different chloramphenicol concentrations. Surviving ones either contain synthetase mutant aminoacylating the NAA or natural amino acid. Those are chosen for negative selection and tested for a growth in the absence of a NAA. The negative selection of a synthetase includes a suppression of amber nonsence mutations introduced into the barnase gene. Hence, only those incorporating natural amino acid survive. Going back to colonies from positive selection

plate, which did not survive in negative selection are the ones containing mutant *Mj* TyrRS most specific for nonnatural amino acid suppression.[106, 107]
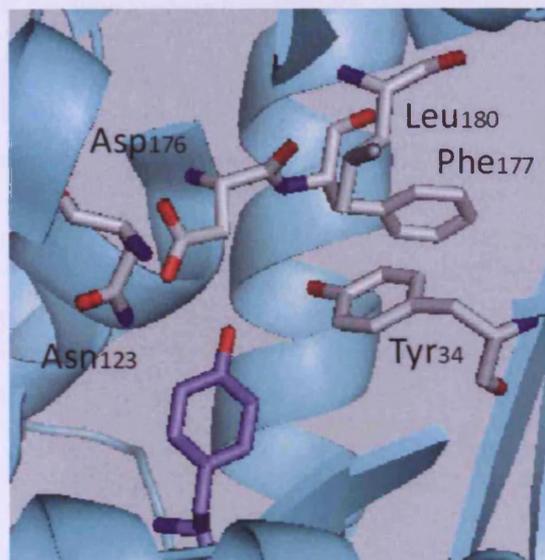
Figure 10: *Bst* **TyrRS active site.** Stereoview shows the important amino acid specificity determining residues and Tyr ligand (blue.) Based on this structure, *Mj* TyrRS active site residue analogues were randomized to select a synthetase mutant, specifically charging *O*Me-Tyr.[106]

Using the directed evolution scheme outlined above, the first NAA incorporated was *O*-methyltyrosine (*O*Me-Tyr). This method resulted in a highly specific AARS due to improved tRNA/synthetase pair orthogonality and because it involved directed evolution of a mutant tRNA and a synthetase library. Shortly after, the same *Mj* TyrRS was used to incorporate 3-(2-naphthyl)alanine in *E. coli.*[142] Another orthogonal and new AARS/tRNA pairs have been derived from *Mj* TyrRS/*Mj* tRNA[Tyr] pair. They displayed a high specificity towards different amino acid analogues, including Tyr-like or Phe-like[4]. Some of these are listed in Table 3.[143]

Table 3: List of NAAs incorporated by AARS/tRNA pairs, derived from *Mj*TyrRS/*Mj*tRNA^{Tyr} pair.

NAA	Common name	Organism(s) in which NAA is encoded	References
1	*p*-acetyl-L-phenylalanine	*E. coli*, yeast, mammalian cells	130, 144-146
2	*m*-acetyl-L-phenylalanine	*E. coli*	132
3	(di-ketone-containing analogue)	*E. coli*	147
4	*O*-allyltyrosine	*E. coli*	131
5	phenylselenocysteine	*E. coli*	148
6	*p*-propargyloxyphenylalanine	*E. coli*, yeast, mammalian cells	145, 146, 149, 150
7	*p*-azidophenylalanine	*E. coli*, yeast, mammalian cells	144-146, 150, 151
8	*p*-boronophenylalanine	*E. coli*	152
9	*O*-methyltyrosine	*E. coli*, yeast, mammalian cells	106, 144-146, 153, 154
10	*p*-aminophenylalanine	*E. coli*	155, 156
11	*p*-cyanophenylalanine	*E. coli*	157
12	*p*-iodophenylalanine	*E. coli*, yeast, mammalian cells	128, 144-146
13	*p*-bromophenylalanine	*E. coli*	128, 158, 159
14	(isotopically labelled analogues)	*E. coli*	10
15	*p*-nitrophenylalanine	*E. coli*	160
16	L-DOPA	*E. coli*	161
17	3-aminotyrosine	*E. coli*	162
18	3-iodotyrosine	*E. coli*, yeast, mammalian cells	163, 164
19	*p*-isopropylphenylalanine	*E. coli*	155
20	3-(2-naphtyl)alanine	*E. coli*	142

21	biphenylalanine	*E. coli*	[165]
22	*p*-hydroxyphenyllactic acid	*E. coli*	[166]
23	bipyridylalanine	*E. coli*	[165]
24	HQ-alanine	*E. coli*	[167]
25	*p*-benzoylphenylalanine	*E. coli*, yeast, mammalian cells	[133, 144-146, 154, 168]
26	*o*-nitrobenzyltyrosine	*E. coli*	[169]
27	2-nitrophenylalanine	*E. coli*	[170]
28	(phenylalanine- 4'- azobenzene)	*E. coli*	[171]
29	(diazirine photocrosslinker)	*E. coli*	[9]
30	(fluorescent analogue)	*E. coli*	[172]
31	*p*-carboxymethylphenylalanine	*E. coli*	[173]
32	3-nitrotyrosine	*E. coli*	[174]
33	sulfotyrosine	*E. coli*	[175]

4.1 Directed evolution & fidelity of TyrRS

Primarily, a synthetase active site is responsible for an amino acid recognition and activation.[176] The whole idea of the expanded genetic code is based on orthogonal AARS fidelity. In other words, the orthogonal AARS active site must be highly specific towards the amino acid analogue and must no longer activate the cognate amino acid. To achieve fidelity, several rounds of positive and negative selection afford to select a synthetase that successfully acylates a NAA.

However, before positive and negative selection, an AARS' active site is usually redesigned by site directed mutagenesis. Smith and co-workers (1978) first introduced site directed mutagenesis as a

powerful *in vivo* method, which enables one to substitute a codon with a different codon.[177] Thus, during translation in the place of this new codon, a new amino acid is introduced, that was not originally there. An amino acid can be substituted by any other 20 common amino acids. This method was also used[77] to systematically mutate the active site residues of *Bst* TyrRS and follow the effect of each mutation in ligand binding. In this approach, it was crucial to examine the new amino acid chain's contribution towards binding of a substrate. Next,[106] based on the active site of homologous *Bst* TyrRS crystal structure, five important *Mj* TyrRS active site residues were chosen for mutagenesis (described in section 4). These residues were mutated in order to create a library of all possible combinations of 20 amino acids, this is called saturation mutagenesis.[107] The library of all mutants was passed through rounds of positive and negative selection. In this fashion, the best *Mj* TyrRS mutant was selected that specifically recognizes *O*Me-Tyr with high efficiency and translational fidelity.[105]

For a better understanding of orthogonal synthetase adaptability towards a nonnatural substrate, Schultz and co-workers crystallized *Mj* TyrRS in complex with *p*-acetylphenylalanine (*p*AcPhe). This crystal structure demonstrated positions of mutated residues (Tyr^{32}Leu, Asp^{158}Gly, Ile^{159}Cys, and Leu^{162}Arg), that change synthetase specificity towards the NAA (Figure 11). In the wt *M. jannaschii* synthetase, similarly as in *Bst* TyrRS, the ligand Tyr has its -OH group hydrogen bonded to Tyr^{32} and Asp^{158} residues. To facilitate a complete change of the TyrRS fidelity towards *p*AcPhe, these two hydrogen bonds had to be deleted by mutagenesis of the crucial Tyr and Asp residues.[178] If Tyr^{32} and Asp^{158} would not be modified, synthetase specificity towards *p*AcPhe would suffer a loss because of the remained Tyr ligand affinity. In other words, in order to completely change *Mj* TyrRS specificity from Tyr to *p*AcPhe, hydrogen bonds forming between Tyr^{32} and Asp^{158} and Tyr -OH group must be targeted by mutagenesis.
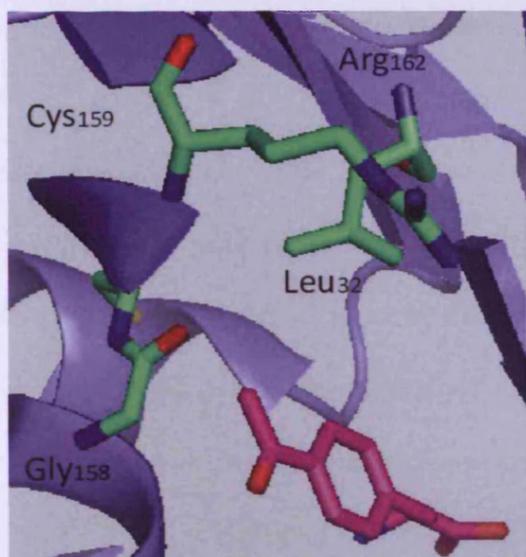
Figure 11: *Methanococcus jannaschii* **TyrRS active site.** Residues Tyr[32], Asp[158], Ile[159], and Leu[162] were mutated to Leu[32], Gly[158], Cys[159], and Arg[162], in order to change synthetase specificity towards *p*-acetylphenylalanine (pink).[178]

4.2   Incorporation of tryptophan analogues

L-Trp is a fundamental protein building block but at the same time is involved in the metabolism of some important molecules like hormones – serotonin or melatonin in animals, or indole alkaloids in plants. L-Trp and its derivates are targets of drug and antibiotic development.[179] L-Trp is a source for most of the absorption and fluorescence observed in proteins. In globular proteins, tryptophans are unique as they represent only 1% of all residues of the protein sequence.[4] Therefore, L-Trp has been used as a probe for fluorescence spectroscopy in order to study conformational changes[180] of a protein as: folding/unfolding, substrate binding etc.[181] A problem of data interpretation occurs if in a mixture of two proteins both comprised with a L-Trp residue, so it is difficult to distinguish between them. That is why L-Trp is not an ideal probe for biochemical interactions. In this case, one would look for a Trp analogue with a unique spectroscopic properties. However, it is very important that such an extrinsic probe have ideal properties so that it does not disturb protein structure and function.

Trp analogue incorporation into a native protein structure has been an important target of studying.[182] Ideally with the incorporation, protein structure or function should not be perturbed. One approach is an *in vivo* expression using a Trp auxotrophic strain. This strain is not able to synthesize L-

Trp by itself and contains a plasmid encoding a protein of interest under the control of inducible promoter.[127] Growth of the strain is attained in a media containing L-Trp. Cells are then harvested by centrifugation and resuspended in a media supplied by a Trp analogue. Next, protein expression is induced and the Trp analogue is misincorporated in place of related L-Trp.[183] Often this incorporation is toxic for bacterial growth if a structure of crucial proteins is changed by the presence of such species.[184, 185] Each NAA displays a different level of toxicity. In the instance of Trp analogues, despite toxicity, several have been able to undergo incorporation into bacterial proteins. Namely, this involves fluorinated tryptophans,[186, 187] azatryptophans,[185, 188] methyltryptophans, and hydroxytryptophans.[184] There is a great advantage of a Trp analogue that can be used in UV absorbance, fluorescence or ^{19}F NMR.[189-191] For example, if a Trp fluorophore integrates into a protein sequence during biosynthesis, a new spectrally enhanced protein (SEP) could be produced. This fluorophore has novel spectroscopic features and is distinct from other tryptophans. Trp analogues were confirmed to be useful for X-ray crystallography.[192] Selenium-containing Trp was used to feed *E. coli* Trp-auxotrophic strain, to be incorporated into human annexin V and barstar. Two such proteins were crystallized and their structures solved. Trp with the electron-rich selenium was essential for phase determination in X-ray crystallography. Some Trp analogues as [6,7]-selenatryptophan, can be used as pharmacologically active substrates in proteins.[98, 126] In order to obtain protein pH sensors, aminotryptophans were incorporated into a protein and showed a significant pH dependence UV shift. [164, 165, 177-180]

Over the years, much work has been done on investigation of synthetase specificity towards more than one substrate. For example, one natural amino acid can serve as a substrate for two different synthetases as well as a wt synthetase is able to charge its cognate tRNA with a NAA.[4] Incorporation of 7-azatryptophan (7-azaTrp) and 2-azatryptophan (2-azaTrp) was achieved in *E. coli* in 1956.[185] Shortly after that it was already clear that TrpRS from the pancreas is responsible for activation of some Trp analogues.[193] In 1968, Schlesinger reported that an auxotroph *E. coli* strain is able to grow on 7-azaTrp or 2-azaTrp and produce an active alkaline phosphatise.[194] In 1970,[184] Trp auxotrophic *B. subtilis* strain

was able to restore its growth on substrates like 5-methyl-tryptophan (5-M-Trp), 5-hydroxytryptophan (5-OH Trp), and 7-azaTrp which were radiolabeled. This experiment showed that after a Trp starvation period, when one of the analogues was added, the aminoacyl incorporation was renewed. Therefore, 5-M-Trp, 5-OH Trp, and 7-azaTrp are able to substitute Trp (5-OH Trp = 7-azaTrp < 5-M-Trp) although, by this substitution a certain level of toxicity is seen. In addition, a competition experiment between Trp and its analogues followed. When cells are grown in the presence of 5-OH Trp, they incorporate 5-OH Trp, but if an additional Trp is supplied they stop the analogue incorporation immediately. In addition, even large amounts of 5-OH Trp or 5-M-Trp did not inhibit the Trp incorporation, when this one is present in media in a small amount. In this study, 5-OH Trp was also indicated to act the same way as Trp in the pathway for the biosynthesis of aromatic amino acids. In 1975, Pratt and Ho[186] emphasized growth of an *E. coli* strain on different fluorotryptophan (F-Trp) analogues. These analogues were incorporated into *E. coli* enzymes lactose permease, β-galactosidase, and D-lactate dehydrogenase that resulted in different levels of activities. In the presence of 4-fluorotryptophan (4-F-Trp) about 75% of Trp was replaced in cellular protein by the 4-F-Trp. In case of 5- or 6-fluorotryptophan (5-F-Trp, 6-F-Trp), 50-60% of Trp was replaced by these analogues. The enzymes restored their activities best if they contained 4-F-Trp in their structures. A couple of years later, in order to replace three tryptophan residues by 5-OH Trp's in λ cI repressor, Ross (1992) used an *E. coli* auxotroph and achieved 95% efficiency.[127] The same year, 5-OH Trp was incorporated into oncomodulin by *in vivo* expression using an auxotroph *E. coli* strain.[71] In another report from Hogue and Szabo (1993) it was suggested[195] that if 7-azaTrp or 5-OH Trp are incorporated into proteins, they must have been forming an aminoacyladenylate in *Bs* TrpRS. Based on the spectroscopic properties of the analogues, these enzyme/aminoacyladenylate complexes could be analyzed. 7-azaTrp adenylate or 5-OH Trp adenylate were indeed shown to be bonded to *Bs* TrpRS and it was appreciated that there is a great benefit in studying protein structure and dynamics of these complexes. Laue[196] and Senear[197] used *in vivo* incorporated 5-OH Trp as an intrinsic probe for studying of protein-nucleic acid interaction between

DNA and bacteriophage lambda cI repressor. Nevertheless, James et al. incorporated 5-OH Trp into transferrin and its receptor, using the mammalian expression system. In this manner, a fluorescence changes could be monitored when iron was released.[198]

A second approach is an *in vivo* transcription-translation using stop codon suppression, which was described earlier in this thesis (section 4). In comparison with TyrRS, directed evolution of TrpRS has been neglected[199] and yet TrpRS does belong to the same Ic class and so it is very similar to TyrRS. Therefore, the main advantage for TrpRS, as for TyrRS in terms of directed evolution approach, is the lack of the editing domain and the active site residues recognition dependence. Moreover, TrpRS active site binds Trp, the largest natural amino acid, and for this reason has a potential for the incorporation of large NAA.[200] For this reason, a new orthogonal TrpRS/tRNA pair from *Saccharomyces cerevisiae* was evolved for use in *E. coli*.[199] Earlier, Schultz and co-workers[76] introduced a new *B. subtilis* TrpRS Val^{144}Pro/tRNA (*Bs* TrpRS Val^{144}Pro/*Bs* tRNA$^{Trp}_{UCA}$) pair that specifically recognizes 5-OH Trp in mammalian cells. However, we found this last example ("Selective incorporation of 5-hydroxytryptophan into protein in mammalian cells") in conflict with previous work done on directed evolution of TyrRS's and *in vivo* Trp analogue incorporations by native TrpRS's using auxotrophic strains. Thus, the next section is primarily focused on problems associated with this report, which is also the main focus of this project.

4.3 Selective incorporation of 5-hydroxytryptophan into protein in mammalian cells

5-OH Trp was chosen as a target for incorporation into a protein structure (Figure 12).[76] 5-OH Trp is a derivative of L-Trp. It is an absorption and fluorescence probe with an absorption band at lower energies than the absorption of tryptophan. Proteins containing such a probe are able to be distinguished even if they display number of L-Trp residues.[71] These spectroscopic properties together with redox properties give a useful tool to probe the protein structure *in vitro* and *in vivo*.[76]

Figure 12: 5-hydroxytryptophan structure.

The previous report claimed to have incorporated 5-OH Trp in response to the opal codon UGA, into bacteriophage T4 fibritin (*foldon*) domain, in mammalian cells (human kidney 293T) with more than 97% fidelity.[76] In order to achieve this, they[76] intended to generate an orthogonal *Bs* TrpRS Val[144]Pro/*Bs* tRNA$^{Trp}_{UCA}$ pair. The report referred to Yokoyama et al., who incorporated 3-iodo-L-Tyr in mammalian cells with 95% fidelity by generating a heterologous orthogonal pair consisting of a *B. stearothermophilus* amber suppressor tRNA^{Tyr} and mutant *Ec* TyrRS.[201] Furthermore, they referred to Xue et al., who described that *Bs* TrpRS/*Bs*tRNA is orthogonal in yeast and mammalian cells.[202] Therefore, they[76] firstly modified *Bs* tRNA^{Trp} sequence for eukaryotic translation, expressed and tested by Northern blot assay. In order to find out if wt *Bs* TrpRS is orthogonal and charges only the suppressor *Bs* tRNA$^{Trp}_{UCA}$, the *in vitro* aminoacylation assay was applied. Only this tRNA was found to be charged by a L-Trp whilst, as a control, the total tRNA isolated from mammalian tRNA remained uncharged. Next, they tested that the *Bs* TrpRS/*Bs* tRNA$^{Trp}_{UCA}$ pair is suppressing an opal mutation stop codon UGA68 within mammalian cells. A resulting full-length *foldon* protein was compared to the wt *foldon* protein and other controls on SDS/PAGE followed by Western blot. It was concluded from these experiments that *Bs* TrpRS/*Bs* tRNA$^{Trp}_{UCA}$ acted as an orthogonal pair for use in mammalian cells. The last approach was to create a *Bs* TrpRS that would specifically aminoacylate the suppressor *Bs* tRNA$^{Trp}_{UCA}$ with 5-OH Trp. Because the crystal structure of *Bs* TrpRS had not been solved, a crystal structure of similar TrpRS tryptophanyl-5'AMP complex from *B. stearothermophilus*[95] was taken by the

report[76] as a model to choose which active site residue to mutate. The main emphasis was directed towards the active site α-helix peptide comprising of residues Asp[140], Ile[141], Val[142], Pro[143], Val[144], and Gly[145]. If 5-OH Trp were to be positioned into the active site pocket in the same manner as is L-Trp as shown in the crystal structure, its 5-OH group would be pointing towards Val[144], which makes this interaction unfavourable. For this reason, they picked Val[144] as a target for site-directed mutagenesis to find an ideal residue that would be smaller and therefore better for 5-substitued Trp analogue accommodation. Altogether 19 different amino acid variants were generated by mutation of Val[144]. Next, they were tested if they are able to suppress the UGA in the mutant *foldon* gene in presence of 1 mM 5-OH Trp. Furthermore, the report stated that one mutant, *Bs* TrpRS Val[144]Pro, was capable of this suppression and there was no *foldon* protein product gained in absence of 5-OH Trp or tRNA$^{Trp}_{UCA}$ (verified by Western blot). High-resolution elestrospray ionization mass spectroscopy, fluorescence spectroscopy and cyclic voltammetry, applied on the mutant *foldon* gene, were other methods used by the group[76] to prove the presence of 5-OH in the mutant protein. To facilitate crosslinking of two monomers containing 5-OH Trp into a dimer, an electrochemical approach was applied and the resulting products separated with SDS/PAGE. On behalf of all of this, it was concluded that *Bs* TrpRS Val[144]Pro mutant efficiently aminoacylates tRNA$^{Trp}_{UCA}$ with 5-OH Trp and not any other endogenous amino acids and, thus, is capable of efficient opal stop codon suppression. According to the report it was startling that only one mutation was able to change the whole *Bs* TrpRS specificity from L-Trp to 5-OH Trp.[76]

It was concluded[76] that this mutation had a significant impact on *Bs* TrpRS fidelity. In order to explain that, the active site of the *Bs* TrpRS Val[144]Pro•5-OH Trp-5'AMP was modelled based on already known wt *Bst* TrpRS•L-Trp-5'AMP (PDB ID 3FIO)[95] (Figure 13). In this modelling 5-OH Trp is orientated differently to L-Trp so its indole ring flips over ~ 180° and Val[144]Pro mutation opens more space for this accommodation. Such a position allows the indole nitrogen, to be hydrogen bonded to Ser[7]. 5' OH group of the 5-OH Trp forms new hydrogen bonds with Asp[133] and His[44].[76]

Modelling of the complex between TrpRS and its substrates using INSIGHT II
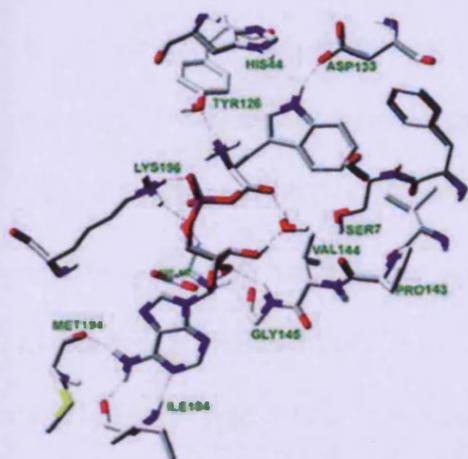
Wild type BsTrpRS•Trp-5'AMP complex
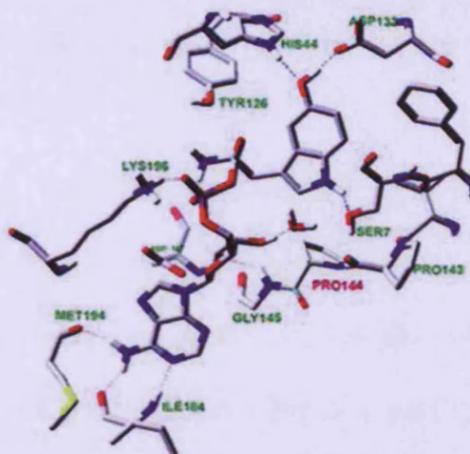
Val144ProBsTrpRS•5HTPP-5'AMP complex

Zhang Z. et.al. PNAS 2004;101:8882-8887

PNAS

Figure 13: Model of *Bs* TrpRS Val[144]Pro•5-OH-5'AMP built from wt *Bst* TrpRS•L-Trp-5'AMP, by previous report.[76]

4.4   Analysis of *Bs* TrpRS

The nonnatural incorporation method, that uses the orthogonal AARS/tRNA pair, faces two challenges. It is to achieve a high translation fidelity and efficiency, comparable with translation of the common amino acids. What does high translation fidelity and efficiency of the NAA incorporation really mean?

The translation fidelity is primarily linked to an AARS and its ability to specifically charge its substrate. While nearly all of the NAAs incorporated by *Mj* TyrRS mutants are poor substrates for wt AARSs, their incorporation *in vivo* depends mainly on a development of a new orthogonal tRNA synthetase pair. A new orthogonal synthetase needs to not only provide great fidelity for the tRNA aminoacylation reaction but also discriminate all the 20 common amino acids. Ideally, in nonnatural

amino acid suppression, the NAA will have a different chemical structure from any natural substrates. It is more challenging to achieve a high specificity of the orthogonal synthetase towards the analogue if this one is too similar to the original substrate. In this case, there is a higher chance that the analogue will be recognized by the endogenous synthetase.[106]

The efficiency of the NAA suppression is determined from the amount of the full-length protein produced after the suppression. If a NAA is incorporated in place of the UAG suppressor stop codon, translation is efficient and a full-length protein is produced. On the other hand, if NAA suppression fails and translation terminates in UAG, a truncated protein is produced. Therefore, the efficiency of the NAA incorporation is the ratio of full-length protein versus truncated protein in the presence of the NAA. A number of factors, such as the concentration of NAA in the media, are related to translation influence its efficiency. Firstly, it is the orthogonal AARS competency to adenylate NAA and charge the orthogonal tRNA. Other important factors are the release factor-1 (RF-1), concentration of the NAA, transport and metabolism of the NAA, elongation factor Tu (EF-Tu) and post-translation modifications.

It was reported that *Bs* TrpRS Val^{144}Pro mutant synthetase incorporated 5-OH Trp, in response to the opal stop codon, in mammalian cells. This was achieved with 20% - 40 % suppression efficiency and 97% fidelity[76] (Figure 14). For most nonnatural amino acid incorporations, the suppression efficiencies vary between 25% - 75% and translation fidelity is >99%. In addition, protein yields after expression usually range from several milligrams to tens of milligrams per litre of a cell culture.[105] Incorporation of amino acid analogues in an eukaryotic system is more difficult to achieve. The main issues usually are the orthogonal tRNA expression and its translational competency as well as the orthogonal synthetase evolution. Moreover, the mammalian cell system is more challenging in terms of the protein expression.[5]

Nevertheless, our concern is: does the suppression by mutant *Bs* TrpRS Val^{144}Pro really display high fidelity and efficiency in the mammalian system as was reported? The reason for this concern is

that *Bs* TrpRS Val^{144}Pro mutant **is not the outcome of a standard library selection** for 5-OH Trp

analogue.

Lane	1	2	3	4	5
Val144ProBsTrpRS	+	+	+	−	+
mutRNA$^{Trp}_{UCA}$	−	+	−	+	+
5-HTPP	−	−	+	+	+

Mutant TrpRS ➡

Foldon-His6 ➡
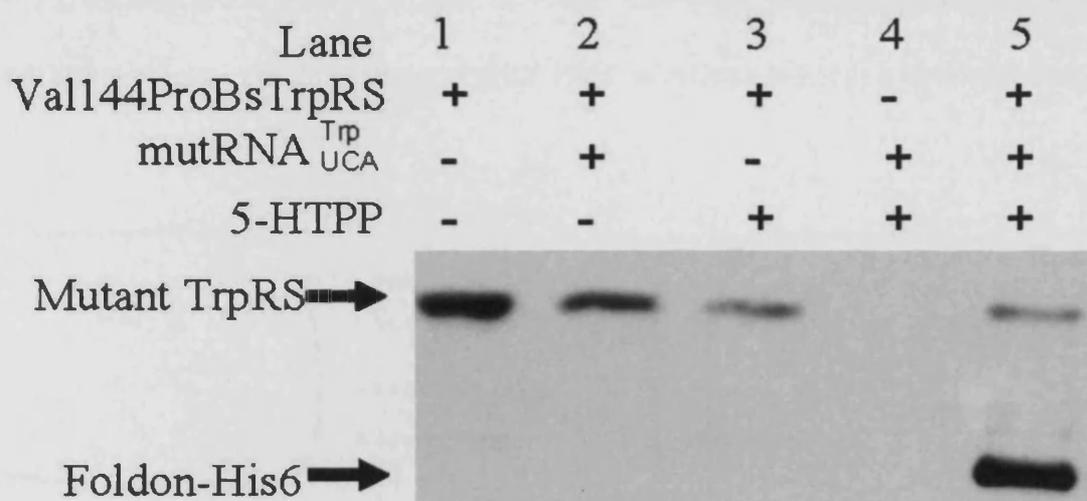
Figure 14: Result from the Schultz's and co. report (Western blot of the cell extracts with an anti-V5 antibody).[76] Incorporation of 5-OH Trp by *Bs* TrpRS Val^{144}Pro mutant into the *foldon* protein. The report used this experiment to show that in the absence of 5-OH Trp, tRNA$^{Trp}_{UCA}$, or *Bs* TrpRS Val^{144}Pro, there is no full length protein produced (lanes 1-4). On the other hand, if those three are present in the cell, the full-length *foldon* protein is produced.

Furthermore, the original report[76] stated that the wt *Bs* TrpRS does not use 5-OH Trp as a substrate

and so it is essential to evolve a new *Bs* TrpRS suppressor tRNA$^{Trp}_{UCA}$ pair if one wants to integrate this

Trp substituent in a protein structure. From previous work listed in section 4.2, it is obvious that 5-OH

Trp is a substrate in cells that they can utilise with the wt TrpRS. It was demonstrated that the wt,

bacterial[127] and human[198] TrpRSs (including *Bs* TrpRS[184]) **are able to recognize and charge 5-OH Trp**

if it is present in growth media (global 5-OH Trp incorporation) (Figure 15). In this incorporation, 5-OH

Trp is introduced at all L-Trp sites. Nevertheless, if only a small amount of L-Trp remains in the media,

incorporation of the natural substrate is preferred to 5-OH Trp.[184] Figure 15 shows, that if 5-OH Trp is to

be incorporated at one permissive site (UGA), a new orthogonal TrpRS/tRNA$^{Trp}_{UCA}$ pair needs to be

introduced into the cell.

This is a different situation as many examples of TyrRS directed evolution. *M. jannaschii* TyrRS wt does not normally display an affinity towards any of the nonnatural substrates that are now regularly encoded.[62] The affinity towards NAA, in case of TyrRS examples, is gained by directed evolution in the active site of the synthetase. In this manner, TyrRS loses its affinity towards a natural substrate.
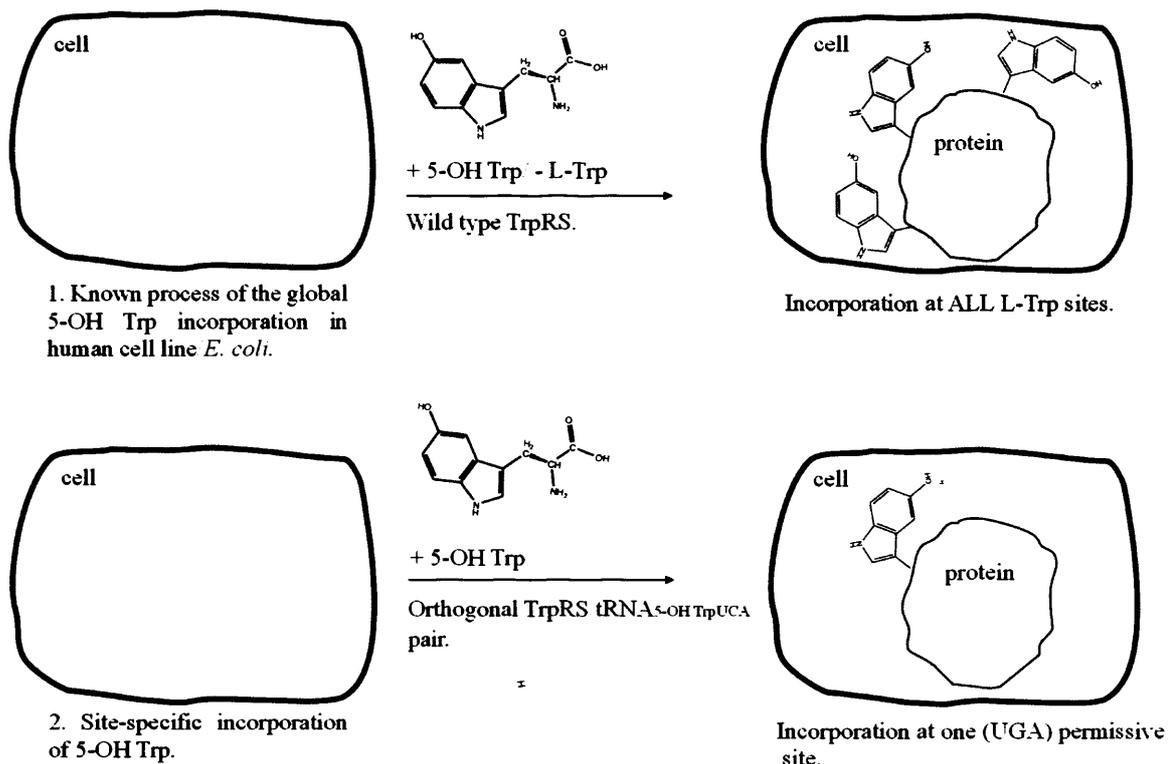
Figure 15: 5-OH Trp incorporation in a human cell line or *E. coli.* 1 - Global 5-OH Trp incorporation according to James[198] in human cell line and Ross[127] in *E. coli*, with no L-Trp access (- L-Trp). 5-OH Trp is incorporated at ALL possible L-Trp sites. 2 - 5-OH Trp could be specifically incorporated at one permissive site (UGA), if the orthogonal TrpRS/tRNA$^{Trp}_{UCA}$ pair is in charge.

If one investigates the active site of a normal wt TrpRS, one can see quite readily, why the active wt TrpRS would still be expected to incorporate a substrate like 5-OH Trp. This is an important detail of our work because we essentially argue that there was a fatal flaw in their research design, and this was the fact that the human TrpRS would be expected to competitively incorporate both 5-OH Trp and L-Trp

into all cellular proteins. This may be the reason for Schultz and coworkers choice of an obscure test system (i.e. their "foldon" protein) which had just one Trp residue. Had the protein had two or more Trp residues, one could have its corresponding codon mutated to the opal codon and targeted for 5-OH Trp, whilst both 5-OH Trp and Trp- would have been competitively targeted to the other L-Trp residues precisely because the endogenous TrpRS has a reasonably high affinity for 5-OH Trp. Again, this latter point has in fact been proven for several bacterial TrpRSs and even human TrpRSs, where it has been shown that the endogenous wt TrpRS, in the presence of lower L-Trp concentrations, will incorporate 5-OH Trp into all proteins.

The fact that the 5-OH Trp might be incorporated into all proteins does not by itself negate the utility of the Schultz study, because one can imagine a scenario where it would be advantageous to site-specifically incorporate 5-OH into a protein. However, our work and analysis goes much further and strongly suggests that even the targeted incorporation of 5-OH Trp at a reprogrammed opal stop codon would proceed preferentially at L-Trp and not the NAA. We will discuss at length how we came to this conclusion, but the real explanation is no more complicated than understanding that a key molecular determinant from Schultz's *Bs* TrpRS mutant is still located in the same position that the residue is found in the wt enzyme.

The hydrogen bond's importance in the molecular determination was already described in chapter 3. Active site synthetase residues are responsible for the substrate specificity of synthetases. They bind the substrate with hydrogen bonds and discriminate against non-substrates. As was studied the *Bst* TrpRS structure[95], the Asp[132] residue is hydrogen bonded to the tryptophan indole nitrogen and is important for L-Trp recognition.[98] This makes the Asp[132] the main molecular determinant in the *Bst* TrpRS active site. Structurally 5-OH Trp is very similar to L-trp and the only difference is the presence of the -OH group in position 5 of 5-OH Trp. Despite this, the structure of the indole ring with nitrogen remains the same for either L-trp or 5-OH Trp. For this reason, we suggest that the recognition of these two ligands by *Bs* TrpRS active site Asp[133] (equivalent of Asp[132] in *Bst* TrpRS) residue happens in the

same manner. This was already confirmed when wt *Deinococcus radiodurans* TrpRS (*Dr* TrpRS) was co-crystallized with 5-OH Trp. The position of the analogue indole ring did not change and remained similar as L-trp in the crystal structure of the same synthetase (PDB ID 2A4M)[203] (Figure 16). Therefore, the indole nitrogen of 5-OH Trp is hydrogen bonded to the aspartate carboxylate group Asp[157] (equivalent of Asp[132] in *Bs* TrpRS) in *Dr* TrpRS.
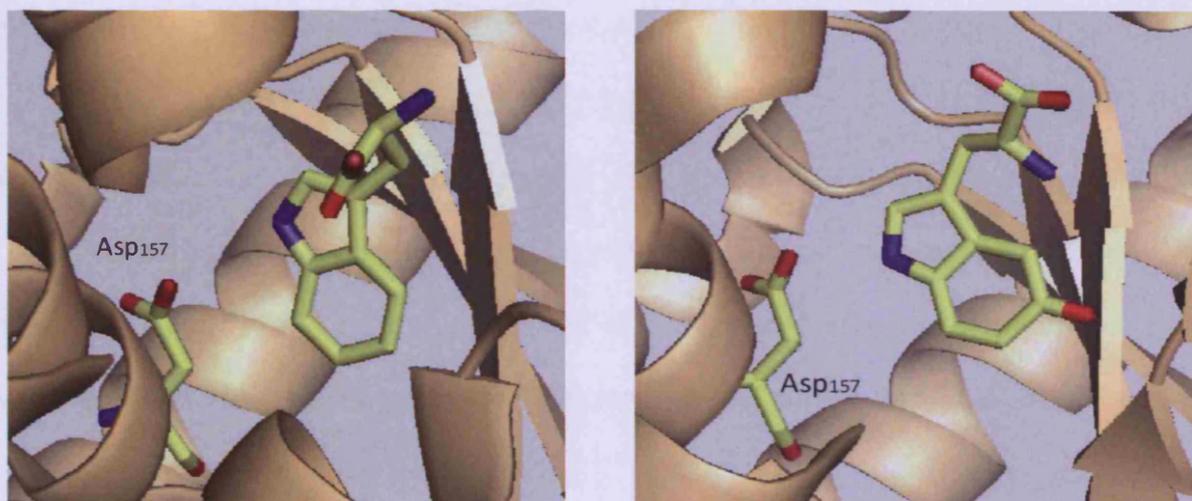
Figure 16: Wt *Deinococcus radiodurans* TrpRS active site crystal structure model with L-trp (left) and 5-OH Trp (right). Nitrogens are indicated in blue, and oxygens red. In both, indole nitrogens points towards the Asp[157] carboxylate group.

The original report[76] explained 5-OH Trp accommodation by modelling the *Bs* TrpRS Val[144]Pro•5-OH Trp-5'AMP active site (Figure 13). 5-OH Trp flips over by 180° and original hydrogen bond between the indole nitrogen and Asp[133] disappears. This suggests that this model disagrees with what was shown in the *Dr* TrpRS crystal structure, where position of 5-OH Trp does not change comparing to the natural substrate. Moreover, the active site of *Bs* TrpRS Val[144]Pro mutant **comprises an unmodified Asp[133] residue** that is the major L-trp or 5-OH Trp determinant. Therefore, the burden is twofold as the synthetase mutant must suppress the enzyme's affinity for L-Trp and 5-OH Trp and switch the enzyme's fidelity to *only* 5-OH Trp.[204]

It was also stated[76] (unlike in *Mj* TyrRS approach) that for selective incorporation of 5-OH Trp into proteins, the Val[144]Pro mutation is enough for enzyme to switch its specificity so that it does not use L-Trp. This is because Val[144] points towards the 5 position of L-Trp in wt *Bs* TrpRS synthetase and clashes with the 5-OH group of 5-OH Trp.[76] However, in the TyrRS directed evolution,[205] usually 4-8 active site residues needed to be randomized to create a TyrRS library. Next, a new TyrRS mutant was selected from the library and displayed a high fidelity towards its nonnatural substrate. However, the original report[76] did not consider the directed evolution approach. Instead, Val[144] was site directly mutated to each of the other 19 amino acids. These 19 *Bs* TrpRS mutants were assayed for their specificity to aminoacylate tRNA$^{Trp}_{UCA}$ with 5-OH Trp. After this test, it was concluded that mutant *Bs* TrpRS Val[144]Pro, specific for 5-OH Trp incorporation displays high efficiency and translation fidelity.

Previous reports showed that one mutation of an active site residue was not enough for a synthetase to change its fidelity towards a nonnatural substrate, **unless an auxotrophic strain was used**. For instance, a Thr[415]Gly mutation in yeast PheRS resulted in site specifically incorporation of 6-chlorotryptophan, 6-bromotryptophan, 5-bromotryptophan, and benzothienylalanine.[206] However, for this attempt to be successful, Phe-auxotrophic strains were used so that the level of the natural amino acid was very low and Phe was not competing the activation of analogues. Yokoyama and co-workers[201] engineered *Ec* TyrRS to specifically recognize 3-iodo-L-tyrosine and not L-Tyr in aminoacylation reaction. As in the *B. stearothermophilus* (Tyr[34]), the identity determinant Tyr[37] is hydrogen bonded to Tyr -OH group in the wt *Ec* TyrRS. This residue was chosen for mutagenesis together with Gln[179] and Gln[195], which are highly conserved among the prokaryotic TyrRS species. The effect of each mutation was examined separately, by randomising those three residues. By introducing one mutation in the active site, the most of mutants remained higher or the same activity towards tyrosine in comparing to 3-iodo-L-tyrosine. Two mutations enhanced the synthetase specificity towards the nonnatural substrate, especially in case of TyrRS Val[37]Cys[195]. Yet SDS-PAGE analysis shows that full-length protein is produced, even if 3-iodo-L-tyrosine is not present. That indicates that the fidelity of TyrRS Val[37]Cys[195]

toward the nonnatural analogue is only moderate. That is why if only one mutation in a synthetase active site is to be applied for incorporation of a NAA, using either auxotrophic strain or depleted media conditions is necessary to achieve a high fidelity. This was not the case when the single mutant *Bs* TrpRS Val^{144}Pro was used for 5-OH Trp incorporation.[76]

What is the impact of Val^{144}Pro single mutation in *Bs* TrpRS? We know that the Val^{144} is not the major determinant of L-Trp specificity. In Söll's work[69], activities of different *Bst* TrpRS single mutants were studied, including a Val^{141}Gln mutant. Val^{141} is a residue located in a similar region as Val^{144} in *B. subtilis*. This Val^{141}Gln mutant has its catalytic efficiency reduced ~300 fold for L-Trp and ~1000 fold for ATP. Therefore, it is more likely that the **activity for ATP is affected** in *Bs* TrpRS Val^{144}Pro most. There is no reason to think that this mutant would have changed its activity only towards 5-OH Trp. Recently, Schimmel and co-workers[75] were able to disturb L-Trp recognition by human TrpRS completely by targeting an important recognition residue (Asp^{133}-like in *Bs* TrpRS). Therefore, such a disruption is essential for *Bs* TrpRS to discriminate L-Trp or 5-OH Trp initially. The single Val^{144}Pro mutation in the *Bs* TrpRS Val^{144}Pro mutant does not convince us about this synthetase high discrimination ability. Hence we ask how a 97% fidelity, and good protein yields were obtained with *Bs* TrpRS Val^{144}Pro mutant synthetase for 5-OH Trp incorporation in mammalian cells?[76]

5. Our hypothesis

Fundamentally, in our approach we examine the impact of the $Val^{144}Pro$ mutation in *Bs* TrpRS. We contend that at least one main identity determinant (Asp^{132}) was not randomized and has the same importance as Asp^{158} (*Mj* TyrRS) or Asp^{176} (*Bst* TyrRS) for the natural substrate recognition especially as no approach towards mutagenesis of this residue was considered in the original work.[76]

Reflecting on the reports we have previously discussed, there is a chance that *Bs* TrpRS $Val^{144}Pro$ mutant would comprise an affinity towards the natural substrate. This would suggest a low fidelity of *Bs* TrpRS $Val^{144}Pro$ mutant, which is contradicting the original report. In addition, the $Val^{144}Pro$ mutation might have the potential to reduce a catalytic efficiency of *Bs* TrpRS.

5.1   Experimental approach

AARS fidelity can be tested and measured by many standard techniques (section 2.4). Since a critical analysis of the original work[76] raises questions, the aim of our project is to directly evaluate the TrpRS enzymes using standard *in vitro* and *in vivo* techniques.

First, we wanted to crystallize the *Bs* TrpRS $Val^{144}Pro$ mutant in the presence of L-trp and 5-OH Trp because, based on the solved crystal structures, we would like to inspect the mutant synthetase active site and the effect of $Val^{144}Pro$ mutation. We are also attempting to see which of two competing ligands specifically binds in the active site and their orientation. In this manner, we hope to confirm the known mechanism of amino acid recognition by these synthetases.

Next, we wanted to apply a complementation assay by using *E. coli* tryptophan auxotrophic strain. Through this method, we hoped to find if an exogenous synthetase (either wt *Bs* TrpRS or *Bs* TrpRS $Val^{144}Pro$) is able to charge *E. coli* $tRNA^{Trp}$ with either L-trp or 5-OH Trp and therefore complement depleted phenotype under L-Trp-starved conditions.

73

Finally, we wanted to measure wt *Bs* TrpRS and *Bs* TrpRS Val^{144}Pro affinities towards L-Trp and 5-OH Trp by ITC. In this way, we want to examine their binding abilities towards both substrates and to indirectly measure the fidelity.

Through these three approaches; crystallography, complementation assay and ITC we wanted to reinvestigate the conclusions found in the original report.[76] This is for the sake of understanding the role of fidelity and what is required for these enzymes to achieve it, as it is the key element for highly efficient nonnatural amino acid incorporation. We also wanted to confirm our hypothesis, that the *Bs* TrpRS Val^{144}Pro mutant is likely a low fidelity synthetase.

75

Material and Methods

6. Material

6.1   Cloning and expression vectors

In this work, the following cloning and expression vectors were used (Table 4):

Table 4: Cloning and expression vectors

Vector	Antibiotic/Resistance	Tag	Copy Number	Application
1. GS43463 pBSK 3871 bp (Epoch Biolabs)	Ampicillin (Amp^R)	-	High	Custom synthesis of *Bs* TrpRS Val^{144}Pro
2. pET151/D-TOPO 5760 bp (Invitrogen)	Ampicillin (Amp^R)	V5, N-term 6×His	Low	Cloning of the synthetase genes wt *Bs* TrpRS and *Bs* TrpRS Val^{144}Pro Protein expression of wt *Bs*TrpRS, *Bs* TrpRS Val^{144}Pro and *Bs* TrpRS Val^{144}Pro + tag
3. pGP1-2 7140 bp (LGC Standards)	Kanamycin (Kan^R)	-	Low	T7 polymerase expression; T7 polymerase facilitates a gene expression in pET151

1. GS43463 pBSK (Epoch Biolabs) - pBlueScript II SK is a high copy number vector with a pUC origin of replication (Figure 17). It is commonly used for cloning, DNA sequencing, RNA transcripts *in vitro* production, site-directed mutagenesis and gene mapping. This vector contains an extensive polylinker with 21 distinguished restriction sites. T7/T3 RNA promoters that margin the polylinker, can be used for in vitro RNA transcription. Part of the vector is also *lacZ* gene fragment that enables the insert detection. The ampicillin resistance is achieved by presence of *bla* ORF. pBlueScript II SK can be also replicated as a ssDNA because it comprises the f1 phage origin of replication.

GS43463 pBSK containing the *Bs* TrpRS Val^{144}Pro insert was used for side-directed mutagenesis of Pro^{144}Val.
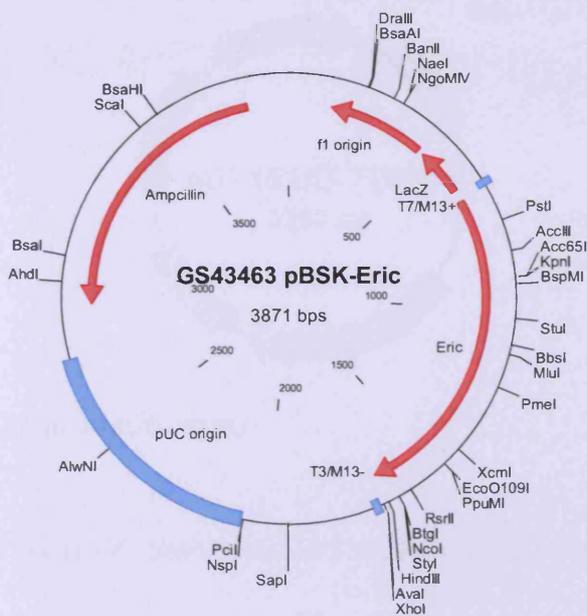
Figure 17: GS43463 pBSK

2. pET151/D-TOPO (Invitrogen) – it is a low copy number vector with a pBR322 origin (Figure 18). In order to facilitate a low copy replication, rop ORF interacts with pBR322 origin. Furthermore, the vector contains T7 promoter, *lac* operator (*lac*O), ribosome binding site (RBS), ATG initiating codon, polyhistidine (6×His) region, V5 epitope, TEV recognition site, T7 reverse priming site, *bla* promoter, ampicilin (*bla*) resistance gene, *lacI* ORF. The presence of T7*lac* promoter enables a high level, IPTG-inducible expression. Both, 6×His and V5 epitope fusion tags together with TEV recognition site are N-terminal oriented. *LacI* ORF inhibits the basal gene transcription. pET151/D-TOPO is designed for cloning of blunt PCR products.

pET151/D-TOPO is used as a cloning and expression vector. In the complementation assay, pET151/D-TOPO is combined with a compatible kanamycin resistant pGP1-2.
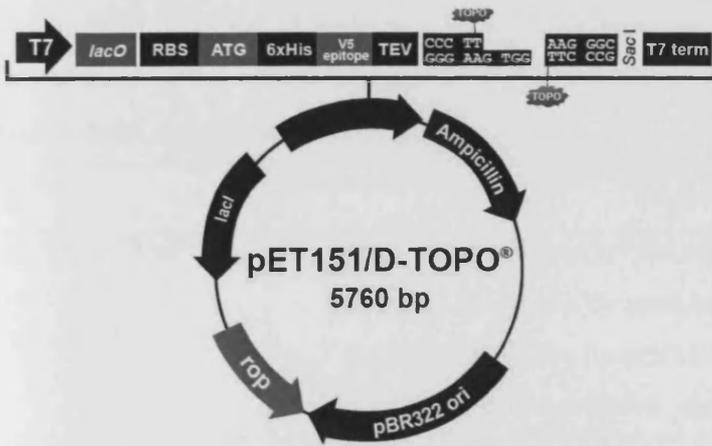
Figure 18: pET151/D-TOPO

3. pGP1-2 (LGC Standards) is a medium copy number plasmid with p15A origin of replication (Figure 19). This plasmid contains T7 polymerase gene which transcription is under the PL promoter. The temperature sensitive cI857 repressor controls T7 polymerase expression which allows non - BL21(DE3) strains to be used.

pGP1-2, producing T7 polymerase, used with pET151/D-TOPO facilitated expression of a gene synthetase in pET151/D-TOPO.
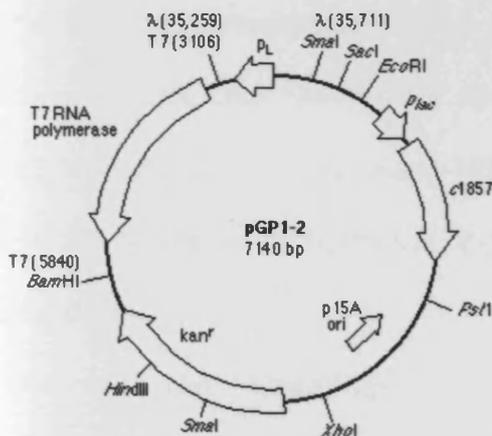
Figure 19: pGP1-2

6.2 DNA template

Table 5: DNA template used in this study

DNA	Description
Bs TrpRS Val^{144}Pro 1092 nt (Sequence I)	• *Bs* TrpRS Val^{144}Pro (in GS43463 pBSK) was purchased from Epoch Biolabs and used for the site directed mutagenesis • *Bs* TrpRS Val^{144}Pro (in pET151/D-TOPO) was obtained by cloning and used for protein expression, complementation assay and ITC
wt Bs TrpRS 1092 nt (Sequence II)	• wt *Bs* TrpRS was obtained by site-directed mutagenesis of *Bs* TrpRS Val^{144}Pro (in GS43463 pBSK) and used for cloning • wt *Bs* TrpRS (in pET151/D-TOPO), obtained by cloning, was used for protein expression, complementation assay and ITC
Bs TrpRS Val^{144}Pro + tag 1167 nt (Sequence III)	• *Bs* TrpRS Val^{144}Pro + tag in (pET151/D-TOPO) was obtained from ET* and used for protein expression and crystallization

* Dr. Eric M. Tippmann

Sequence I, II and III are given in Appendices

6.3 Bacterial strains

For all DNA and protein strategies done in this work, following strains were used:

***Escherichia coli* One Shot® TOP 10** (Invitrogen): F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ-. This strain was used as a host for cloning of *Bs* TrpRS Val^{144}Pro, *Bs* TrpRS Val^{144}Pro + tag or wt *Bs* TrpRS genes and plasmid propagation.

***Escherichia coli* GeneHogs®** (Invitrogen): F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) 80*lac*ZΔM15Δ *lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ *rps*L(Str^R) *nup*G. This strain was used as a host for cloning of *Bs* TrpRS Val^{144}Pro, *Bs* TrpRS Val^{144}Pro + tag or wt *Bs* TrpRS genes and plasmid propagation.

Escherichia coli **One Shot® BL21 Star**^{TM} **(DE3)** (Invitrogen): F⁻ *ompT hsdS*$_B$ (r$_B$⁻m$_B$⁻) *gal dcm rne131* (DE3). This host was used for *Bs* TrpRS Val^{144}Pro, *Bs* TrpRS Val^{144}Pro + tag or wt *Bs* TrpRS genes expression.

Escherichia coli **KY4040** (Coli Genetic Stock Center): F-, *tyrT58*(AS), *trpS44, mel-1*. This strain was used for the complementation assay, when strain was double-transformed with pET151 *Bs* TrpRS Val^{144}Pro /pGP1-2 or pET151 wt *Bs* TrpRS /pGP1-2.

6.4   Enzymes

Table 6: Enzymes used in this study and their suppliers

Enzyme	Supplier
Platinum® *Taq* DNA Polymerase	Invitrogen
*Dpn*I restriction enzyme	*NEW ENGLAND BioLabs*
ß-Agarase I	*NEW ENGLAND BioLabs*
Pwo DNA polymerase	Roche
EconoTaq PLUS DNA polymerase	Lucigen® Corporation
Topoisomerase I	Invitrogen

6.5 Kits

Table 7: Kits used in this study and their suppliers

Kit	Supplier
QIAquick® PCR Purification Kit (50)	QIAGEN
e.Z.N.A.™ Gel Extraction Kit	OMEGA Bio-Tek
Zyppy™ Plasmid Miniprep Kit	ZYMO RESEARCH
QIAprep® Spin Miniprep Kit (50)	QIAGEN
Champion™ pET Directional TOPO® Expression Kit	Invitrogen

6.6 Chromatography columns

Table 8: Chromatography columns used in this study

Resin	Supplier
Nickel-nitrilotriacetic acid (Ni-NTA) agarose	Qiagen
Superdex 75 100/300 GL	Amersham Biosciences
Superdex 200 100/300 GL	Amersham Biosciences

6.7 Media and supplements

All media components were usually ordered from *MELFORD* or *SIGMA-ALDICH*.

Luria-Bertani (LB) medium

1 % Bacto Tryptone	10 g
0.5 % Yeast extract	5 g
1 % NaCl	10 g
ddH$_2$O	Up to 1 litre (pH 7.5)

Luria-Bertani (LB) agar

1 % Bacto Tryptone	10 g
0.5 % Yeast extract	5 g
1 % NaCl	10 g
1.5 % Agar	15 g
ddH$_2$O	Up to 1 litre (pH 7.5)

2×YT medium

1 % Bacto Tryptone	16 g
0.5 % Yeast extract	10 g
1 % NaCl	5 g
ddH$_2$O	Up to 1 litre (pH 7.5)

Super Optimal (SOC) medium

1 % Bacto Tryptone	20 g
0.5 % Yeast extract	5 g
5 M NaCl	2 ml
1 M KCl	2.5 ml
1 M MgCl$_2$	10 ml
1 M MgSO$_4$	10 ml
1 M glucose	20 ml
$_{dd}$H$_2$O	Up to 1 litre (pH 7.5)

M9 minimal medium

Stock solutions	Final concentration in media
25 % NaCl	0.5 %
100 × Metal salts	1 ×
50 % Glycerol	1 %
1M MgCl$_2$	0.05 mM
20 mg/ml CaCl$_2$	0.02 mg/ml
20 mg/ml Biotine	0.08 mg/ml
20 mg/ml Thiamine	0.08 mg/ml
10 × M9's	1 ×
0.1 M Leucine	0.3 mM

In order to achieve high sterility, media were either autoclave for 20 minutes/121°C or filter (Millipore) sterilized (0.4 µm pore size).

6.8   Antibiotics

All antibiotics were purchased from *MELFORD*.

Table 9: Antibiotics used in this study

Antibiotic	Stock solution	Final concentration in media
Ampicillin	100 mg/ml in H_2O	100 µg/ml
Carbenicilin	100 mg/ml in H_2O	100 µg/ml
Kanamycin	50 mg/ml in H_2O	50 µg/ml

6.9   Oligonucleotides

Oligonucleotides were designed in EditSeq and SeqMan. The annealing temperature (Tm) was determined with the online Finnzyme Tm calculator. Oligonucleotides were synthesized and lyophilizated by IDT (INTEGRATED DNA TECHNOLOGIES). Obtained oligonucleotides were usually dissolved in $_{dd}H_2O$ to the final concentration 100 µM, aliquoted and stored in - 20°C.

Table 10: List of used oligonucleotides

ID	Sequence	Tm [°C]
	Oligonucleotides used for overlap PCR	
Pro144Val_F	GATCTGGTCCCG<u>GTC</u>GGTGAAGACCAGA	69
Pro144Val_R	GATCCGTCCCGTACAGCAGATATC	60
	Oligonucleotides used for cloning into pET 151	
TrpRS_F	CACCATGAAACAGACCATTTTCAGCG	65
TrpRS_R+stop	TTATTAGCGGCGTTTGCGACCTAAAC	60
	Oligonucleotides used for sequencing	
T7F (invitogen)	TAATACGACTCACTATAGGG	48
T7R (invitogen)	TAGTTATTGCTCAGCGGTGG	52
M13F (invitogen)	GTAAAACGACGGCCAG	46
M13R (invitogen)	CAGGAAACAGCTATGAC	45

6.10 Software

Table 11: List of used programs, databases and bioinformatic tools

Name	Link
BLAST	http://www.ncbi.nlm.nih.gov/blast/
CCP4-6.1.3	http://www.ccp4.ac.uk/about.php
COOT-0.6.1	http://www.biop.ox.ac.uk/coot/
EditSeq	http://www.dnastar.com/t-sub-products-lasergene-editseq.aspx
ExPASy Proteomics Server	http://expasy.org
Finnzyme Tm calculator	https://www.finnzymes.fi/tm_determination.html
IC ITC	http://www.cf.ac.uk/chemy/staffinfo/poc/software/ic_itc/
MolProbity	http://molprobity.biochem.duke.edu/
Molrep	http://www.ccp4.ac.uk/html/molrep.html
PROCHECK	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
ProtParam	http://expasy.org/tools/protparam.html
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/
PyMOL 0.99rc6	http://www.pymol.org/
RSCB Protein Data Bank	http://www.rcsb.org/pdb/home/home.do
Refmac5	http://www.ccp4.ac.uk/dist/html/refmac5.html
SeqMan	http://www.dnastar.com/t-sub-products-lasergene-seqmanpro.aspx
Translate tool	http://expasy.org/tools/dna.html
UNICORN 5.2	http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/ aktadesign_platform~system_software

7. Methods

7.1   DNA purification, concentration and sequencing

For plasmid and DNA purification, Zyppy™ Plasmid Miniprep Kit or QIAprep® Spin Miniprep Kit (50) were used. DNA concentration was determined by using nanodrop (Labtech International Ltd.). All DNA constructs were sequenced by Cardiff University DNA Sequencing Core, and analyzed in DNASTAR SeqMen program.

7.2   Mutagenesis

The 1012 bp DNA fragment (6.2) encoding *Bs* TrpRS with desired mutation Val^{144}Pro was synthesised and provided in the GS43463 pBSK vector (6.1). Platinum® *Taq* DNA Polymerase PCR reaction (Table 12), using two overlap PCR primers Pro144Val_F and Pro144Val_R (6.9), was designed to provide the CCC mutation of Pro^{144} into GTC Val^{144}. After PCR, vector was digested by *Dpn*I restriction enzyme (6.4) and low melting agarose gel purified (7.2.3). This vector was transformed in Gene-Hog chemicompetent cells, which were incubated at 37°C overnight on 2YT agar plates with carbenicillin. Next, single colonies were picked from the plate and grown as 5 ml LB saturated cultures at 37°C overnight. DNA from every overnight culture was purified separately and its concentration measured. Furthermore, these DNA samples were sent for sequencing with M13F and M13R primers.

Table 12: Overlap PCR reaction's components and conditions

Overlap PCR[I] of *Bs* TrpRS Val[144]Pro			
PCR reaction components		**PCR reaction conditions**	
10×PCR Mg-free buffer	5 µl	denaturing 94°C, 3 min	
10 mM dNTP mixture	1 µl		
50 mM MgCl₂	1.5 µl	denaturing 94°C, 30 sec	35 ×
primer mix - Pro144Val_F and	1 µl	annealing 55°C, 30 sec	
Pro144Val_R (10 µM each)		extension 72°C, 5 min/10 sec	
template DNA[II] (100 ng/µl)	0.5 µl	final extension 72°C, 5 min	
Platinum® *Taq* DNA Polymerase	0.2 µl	hold 4°C	
ddH₂O	Up to 50 µl		

I)   PCR was run on TC-3000 Personal 25-well Thermal Cycler

II)   Sequence of DNA template is given in Appendices

7.2.1 Agarose gel electrophoresis

In order to visualize PCR or plasmid DNA samples, the agarose gel electrophoresis was used as a DNA presence detection method. 1 % agarose gel was prepared by warming 0.5 g agarose (Acros Organics) mixed with 50 ml 1×Tris-acetate-EDTA (TAE) buffer (AppliChem) in a microwave, until it completely dissolves. This agarose solution was cooled down to about 60 °C and Ethidium Bromide (*BIO RAD*) was added to the final concentration of 5 µg/ml so that DNA could be visualized on UV after electrophoresis. Next, the solution was poured into the casting tray and allowed to solidify. Before loading on the gel, the DNA samples (3-7 µl) were mixed with 6× loading dye (10 mM Tris-HCl pH 7.6, 60 % glycerol, 60 mM EDTA and 0.03% bromophenol blue) to the final concentration 1×. Samples were loaded on the gel together with 1 kb DNA Ladder (BioLabs) in order to detect molecular weight of sample bands. Electrophoresis was usually run at 80-100 V in an electrophoretic chamber, for 30-90 min. Gel, containing DNA bands was visualized with UV (VersaDoc™ Imaging System, *BIO RAD*).

7.2.2 DpnI digestion

In order to cleave the original DNA template in the PCR reaction, DpnI digestion was carried on. DpnI cleaves DNA recognition site that is methylated. 1µl of DpnI (20 U/ µl) was applied to the PCR tube, mixed, incubated for 1 hour/37°C and purified.

7.2.3 Low melting agarose gel purification

The entire PCR reaction was run on a 1 % low melting agarose (Promega) gel in order to separate and purify PCR product. A gel slice containing DNA band of a right size was simply cut from the gel and allowed to melt for 10 min in 70°C preheated water bath. Subsequently, this water bath was brought to a temperature of 42°C and 10× ß-Agarase I Reaction Buffer (10 mM Bis-Tris-HCl, 1 mM EDTA, pH 6.5) was added to the final concentration 1×. Futher, ß-Agarase I was added, 1U (1 µl) for every 200 µl of melted agarose gel and left for 1 hour/42°C. The PCR product was purified with QIAquick® PCR Purification Kit (50) or e.Z.N.A.™ Gel Extraction Kit.

7.2.4 Heat-shock transformation

Plasmid transformation was done by a heat-shock. One stock of chemically competent cells – 100 µl (stored in - 80°C) was first melted on ice. Next, 0.5 - 3 µl of a plasmid DNA was gently mixed with the cells and incubated for another 10 - 30 min. The mixture was then heat-shocked for 30 - 45 s in 42°C and cool on ice for 1 min. Cells were rescued by adding 200 – 500 µl of SOC media (6.7) and shaken for 1 hour/37°C. Finally, the cell mixture was spread on the LB agar plate and incubated at 37°C overnight.

7.3   Cloning into pET151/D-TOPO

The DNA fragment encoding wt *Bs* TrpRS (6.2) was amplified from GS43463 pBSK (6.1) by Pwo DNA polymerase PCR (Table 13) with TrpRS_F and TrpRS_R+stop primers (6.9). After PCR, fragment with CACC at the 5′ end was low melting agarose gel purified (7.2.3). In order to express *Bs* TrpRS with 6×His tag at the N terminus in *E. coli,* the PCR fragment encoding the synthetase was ligated (7.3.1) into a pET151/D-TOPO vector (6.1), by using Champion™ pET Directional TOPO® Expression Kit (6.5). This construct was further transformed in *E. coli* One Shot® TOP 10 chemical competent cells (6.3). The cells were incubated at 37°C overnight on LB agar plates with 100 μg/ml carbenicillin. In order to check the presence of the insert DNA in the TOPO vector, 6 colonies were picked and a colony PCR (Table 13) was carried on. For the colony PCR, EconoTaq PLUS DNA polymerase was used in a mixture of ready-to-use PCR EconoTaq PLUS 2× Master Mix and T7F, TrpRS_R primers (6.9). The same colonies were inoculated in 5 ml volume of LB media with carbenicilin and grown at 37°C overnight. DNA from every culture was purified separately; its concentration measured and sent for sequencing with T7F, T7R and TrpRS R (6.9).

In the same manner, *Bs* TrpRS Val^{144}Pro (6.2) was amplified (from *Bs* TrpRS Val^{144}Pro + tag in pET151/D-TOPO) and cloned into pET151/D-TOPO.

Table 13: Pwo DNA polymerase and EconoTaq PCR reaction's components and conditions.

Amplification[1] of *Bs* TrpRS Val[144]Pro and wt *Bs* TrpRS for cloning into pET151/D-TOPO				
PCR reaction components		**PCR reaction conditions**		
1 mM dNTP mixture	30 µl	denaturing 94°C, 2 min		
primer mix - TrpRS_F and TrpRS_R (10 µM each)	6 µl	denaturing 94°C, 15 sec		30 ×
template DNA[II] (200 ng/µl),	1 µl	annealing 60°C, 30 sec		
10× PCR buffer	15 µl	extension 72°C, 1 min/20 sec		
Pwo DNA polymerase	1 µl	final extension 72°C, 4 min		
ddH₂O	Up to 150 µl	hold 4°C		
Colony PCR[1]				
PCR reaction components		**PCR reaction conditions**		
EconoTaq PLUS 2× Master Mix	25 µl	denaturing 94°C, 2 min		
primer T7F (100 µM)	1 µl	denaturing 94°C, 30 sec		35×
primer TrpRS_R (100 µM)	1 µl	annealing 60°C, 30 sec		
colony DNA		extension 72°C, 1 min/20 sec		
ddH₂O	Up to 50 µl	final extension 72°C, 2 min		
		hold 4°C		

I)   PCR was run on TC-3000 Personal 25-well Thermal Cycler

II)   Sequence of DNA template is given in Appendices

7.3.1   Ligation reaction

Directional TOPO® Cloning is based on a reverse reaction of Topoisomerase I phosphodiester cleavage. For ligation reaction of wt *Bs* TrpRS, *Bs* TrpRS Val[144]Pro constructs with pET151/D-TOPO vector (6.1), we used molar ratio 1 : 0.5 of PCR product:TOPO vector. The reaction, with total volume of 6 µl contained 1 µl Fresh PCR product, 1 µl Salt Solution, 3.5 µl water and 0.5 µl TOPO vector. These components were mixed together, incubated at 22°C for 30 min, put on ice and used for transformation.

7.4 Expression of mutants *Bs* TrpRS Val^{144}Pro, *Bs* TrpRS Val^{144}Pro + tag and wt *Bs* TrpRS

Starter cultures of *E. coli* One Shot® BL21 Star™ (DE3) (6.3), containing the pET151 with *Bs* TrpRS mutants or wt (6.2) were grown with vigorous shaking (210 rpm) at 37°C to an OD 0.6 - 0.9. Proteins were overexpressed after induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG, MELFORD) for 4-5 hours.

7.5 Purification of *Bs* TrpRS-6×His

Cells were pelleted by centrifugation at 4000 ×g for 20 min (4 °C). A quantity of 1.5 g of wet cells from a 250 ml culture was resuspended in 10 ml lysis buffer (50 mM NaH$_2$PO, pH 8.0; 300 mM NaCl; 10 mM imidazole) and frozen overnight at -80°C. The cell pellet was thawed on ice. 1 mM protease inhibitor, 1 mg/ml lysozyme and benzonase nuclease (2 µl / 250 ml starting culture) were added to the cells and mixed on rotator for 30 min, followed by sonication on ice. The lysate was centrifuged at 10,000 ×g for 40 min at 4°C to pellet the cellular debris. In order to purify the *Bs* TrpRS-6×His, nickel-nitrilotriacetic acid slurry (Ni-NTA, 250 µl) was added to the recovered supernatant and incubated with gentle mixing for 2 hour at 4°C. The mixture was applied to a minicolumn for gravity flow chromatography. The resin was washed with 50 ml each of 15, 20, mM imidazole and 20 ml of 30, 40, 50, 75 mM imidazole wash buffer (50 mM NaH$_2$PO, pH 8.0; 300 mM NaCl). The synthetase was eluted with elution buffer (50 mM NaH$_2$PO, pH 8.0; 300 mM NaCl; 250 mM imidazole). The presence of synthetase in the elution fractions was determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE; 7.5.1) and fractions containing the synthetase were pooled and concentrated by centrifugation in AMICON tube (Millipore). The protein was further purified by FPLC size-exclusion column (7.5.2).

7.5.1 SDS-PAGE

SDS-PAGE was used as a protein detection method. Normally, 15 µl protein samples were mixed with 5× Sample buffer (see below) and denatured in 80°C/10 min. Next, samples were run on polyacrylamide gel (Acrylamide solution, *SIGMA-ALDRICH*) composed of 13 % Separating and 4 % Stocking gel (see below), at 150 – 200 V for 60-90 minutes. Protein ladder (10-250 kDa, *NEW ENGLAND* BioLabs) was run together with samples, in order to determine molecular weight of the sample bands. After the run, gel was stained in Coomassie Brilliant Blue staining buffer (see below) and distained in water. Gels were visualized (VersaDoc™ Imaging System, *BIO RAD*).

5× Sample buffer

Tris-HCl pH 6.8	0.2 M
2-mercaptoethanol	10 mM
SDS	10 %
Glycerol	20 %
Bromophenol blue	0.05 %

10× Catode buffer

Tris-HCl pH 8.25	1 M
Tricine	1 M
SDS	1 %

10× Anode buffer

Tris-HCl pH 8.9	2.1 M

Separating gel (4 short gels)

40 % Acrylamide solution	10 ml
Gel buffer*	10 ml
70 % Glycerol	4 ml
H_2O	6.2 ml
10 % APS	133 µl
TEMED	13,2 µl

Stocking gel (4 short gels)

40 % Acrylamide solution	2.12 ml
Gel buffer*	5 ml
H_2O	13.44 ml
10 % APS	160 µl
TEMED	16 µl

*Gel buffer

Tris-HCl pH 8.45	3 M
SDS	0.3 %

Coomassie Brilliant Blue staining buffer

Ethanol	250 ml
Acetic acid	80 ml
Coomassie Brilliant Blue R-250	2.5 g
H_2O	Up to 1 litre

7.5.2   Analytical gel filtration

In order to obtain maximum protein purity, we used Superdex 200 100/300 GL or Superdex 75 10/300 GL size-exclusion columns. The purification was usually run with the flow rate of 0.4 ml/min, on ÄKTApurifier UPC (GE Healthcare) in buffer A: 50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5 mM β-mercaptoethanol and 0.5 mM EDTA. Before actual experiments, purity of the protein was determined by SDS-PAGE. To analyze molecular weight of the protein samples for ITC, buffer I (50 mM Tris (pH 8), 300 mM NaCl, 2% glycerol, 0.5 mM EDTA) was used.

7.6   Protein concentrations

Protein concentrations for *Bs* TrpRSs (molecular mass: *Bs* TrpRS Val^{144}Pro = 40,978 Da, wt *Bs* TrpRS = 40,980 Da and *Bs* TrpRS Val^{144}Pro + tag = 43,833.4 Da) were determined from their extinction coefficient at 280 nm (ε_{280} = 27,975 M^{-1}.cm^{-1}, ε_{280} = 27,975 M^{-1}.cm^{-1} and ε_{280} = 33,475 M^{-1}.cm^{-1}) and confirmed by nanodrop (Labtech International Ltd.) and Bradford protein assay (*BIO RAD*) with bovine serum albumin (BSA) as a standard. The extinction coefficient was calculated from protein sequence (ProtParam tool).

7.7   Complementation of *E. coli* KY4040

Competent *E. coli* KY4040 (6.3) were co-transformed with pGP1-2 (Kan^R) (6.1), which encodes T7 RNA polymerase, and with pET plasmids encoding either wt or mutant *Bs* TrpRS (Amp^R) (6.2). As a control, we also transformed *E. coli* KY4040 with an empty pET151$_e$ vector. Cells were plated, cultivated overnight at 37°C on M9 minimal media (6.7) plates containing 50 μg/ml ampicilin, 25 μg/ml kanamycin and 0.5 mM L-trp. Prior to the complementation experiment, four single colonies were selected from each of the plates and resuspended in a separate 20 μl volume of deionized water. In order

95

to obtain a similar density of the *E. coli* KY4040 cells containing three different constructs, we measured the cell absorbance (600 nm) on the nanodrop (Labtech International Ltd.). M9 minimal media plates containing either 0.05 μM L-trp or 5 μM 5-OH Trp were inoculated by the same amount of the *E. coli* KY4040 containing pGP1-2 and either of three different pET151 (pET151 wt *Bs* TrpRS, pET151 *Bs* TrpRS Val^{144}Pro or pET151$_e$). This inoculation was done in three separate sections on the plate.

7.8 Isothermal titration calorimetry

Calorimetric measurements were carried out using a VP-isothermal titration calorimeter (MicroCal) at 25°C. Prior to titration, concentrated *Bs* TrpRS's protein samples were dialyzed overnight against buffer I: 50 mM Tris (pH 8), 300 mM NaCl, 2% glycerol, 0.5 mM EDTA. The final dialysate was used to prepare ligand solutions.

97

Results and Discussion

Because of the complexicity of the mammalian system, we have not repeated techniques that the original report used. Instead, we employed standard methods for the AARS evaluation. These are crystallization, complementation assay and ITC. In this manner, we have examined the $Val^{144}Pro$ *Bs* TrpRS for use in the orthogonal system.

Experimental part of this project was done at Cardiff University Physical Organic Chemistry (POC) Centre. Crystallization data were collected at the Diamond Light Source in the UK, by Sam Hart from York Structural Biology Lab. Crystallization and structure refinement were carried out with help of Prof. Matthias Bochtler, Dr Honorata Czapinska and Dr Ania Piasecka. ITC experiments were performed with help of Dr Niklaas J. Buurma and Dr Lavinia Onell.

8. Mutagenesis, expression and purification

In order to confirm a single Pro[144]Val mutation in *Bs* TrpRS (6.2), I isolated a number of plasmid samples (GS43463 pBSK/*Bs* TrpRS) from bacterial cultures (6.9). After sequencing, one of the samples showed the GTC (Val[144]) mutation in the desired place (Figure 20).


Figure 20: Sequencing chromatogram (SeqMan). Sequences show the place of Pro[144]Val (GTC) mutation in both M13F and M13R samples. These are aligned with the original *Bs* TrpRS Val[144]Pro (CCC) sequence.

I verified the cloning reaction of wt *Bs* TrpRS or *Bs* TrpRS Val[144]Pro into pET151/D-TOPO by colony PCR, agarose gel electrophoresis and sequencing (7.1, 7.2.1, 7.3). The positive clones that contained the insert could be distinguished on the electrophoresis gel as they were bigger than the negative clones (empty pET151*e*). Sequencing also showed the right position of *Bs* TrpRS's in pET151/D-TOPO vector.

In order to obtain high protein yields, I overexpressed wt *Bs* TrpRS, *Bs* TrpRS Val[144]Pro, *Bs* TrpRS Val[144]Pro + tag in pET151/D-TOPO with optimized conditions and temperature (7.4). When Ni-NTA purification was applied, part of the protein started to elute in the wash buffer containing 20 mM imidazol but this fraction displayed high impurity. The rest of the protein eluted in 30, 40, 50, 75, 250 mM imidazol fractions with similar purities (Figure 21). Therefore, when repeating Ni-NTA purification, after 30 mM imidazol wash, I eluted protein directly in 250 mM imidazol buffer. This was in order to obtain highly concentrated samples for ITC and crystallization. For the size exclusion on Superdex 75

100/300 GL (Amersham Biosciences), I combined fractions 30 mM (40 mM, 50 mM, 75 mM if used) and 250 mM imidazol. In the size exclusion purification, the greatest part of the target protein was typically found in fractions 10, 11, 12 and 13 (Figure 22). After protein chromatography, SDS-PAGE confirmed sufficient purity of the concentrated protein samples for crystallization and ITC (Figure 23). For protein crystallography, I used the purest fractions 11 and 12 for further dialysis and concentration whereas for ITC, I typically combined and dialyzed fractions 10, 11, 12 and 13.
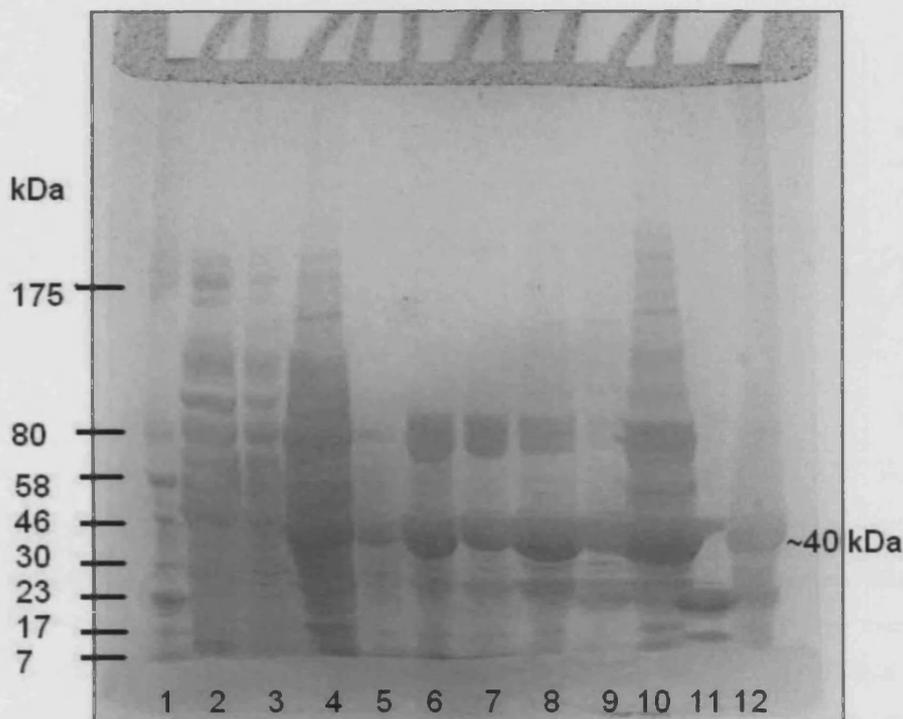
Figure 21: SDS protein gel analysis of *Bs* TrpRS expression at different steps of Ni-NTA purification. Electrophoresis was run on a 13% SDS gel and stained by Coomassie blue. 1 – prestained protein marker (*BioLabs*) with molecular weights of 175, 80, 58, 46, 30, 23, 17 and 7 kDa, respectively, 2 - flow through fraction, 3 - first wash fraction with 15 mM imidazol, 4 - wash fraction with 20 mM imidazol, 5 - wash fraction with 30 mM imidazol, 6 - wash fraction with 40 mM imidazol, 7 - wash fraction with 50 mM imidazol, 8 - wash fraction with 75 mM imidazol, 9 - eluted fraction with 250 mM imidazol, 10 - control TrpRS, 11 - control Hcp1 protein ~ 20 kDa, 12 - control TrpRS.
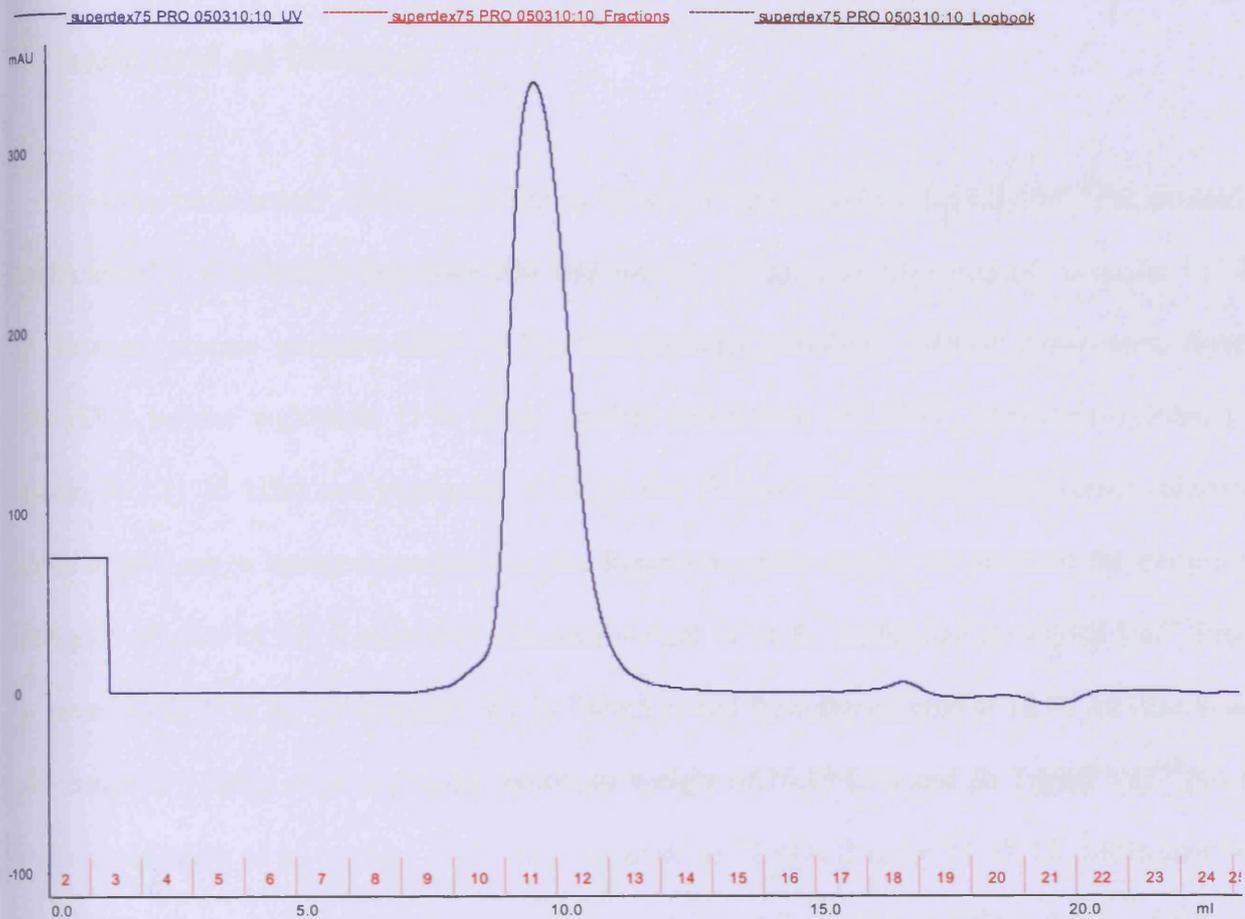
Figure 22: Size exclusion purification of *Bs* TrpRS on Superdex 75 10/300 GL (Amersham Biosciences). The purification was performed on ÄKTApurifier UPC (GE Healthcare), with 0.4 ml/min flow rate and 1.8 MPa maximum column pressure, in a room temperature (in buffer A: 50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5mM β-mercaptoethanol and 0.5 mM EDTA). The elution profile was monitored at 280 nm. Protein was typically found in fractions 10, 11, 12 and 13.
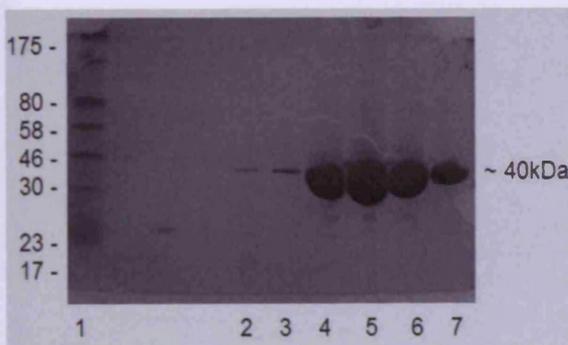
Figure 23: SDS protein gel analysis of *Bs* TrpRS after size-exclusion protein chromatography. Electrophoresis was run on a 13% SDS gel and stained by Coomassie blue. 1 – prestained protein marker (*BioLabs*) with molecular weights of 175, 80, 58, 46, 30, 23, 17 and 7 kDa, respectively, 2 – 8 fraction, 3 – 9 fraction, 4 – 10 fraction, 5 – 11 fraction, 6 – 12 fraction, 7 –13 fraction.

9. Analytical gel filtration

In order to determine Molecular weights of wt *Bs* TrpRS and *Bs* TrpRS Val^{144}Pro protein samples experimentally, I calibrated Superdex 200 100/300 GL (Amersham Biosciences) in buffer I (7.8) with a gel filtration protein standard (*BIO RAD*). This standard contained: protein aggregates, thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), equine myoglobin (17 kDa), vitamin B12 (1.35 kDa) and tryptophan (204.23 Da) (Figure 24. A). From the protein standard elution profile at 280 nm, a linear dependence of the logarithm of the molecular mass on the elution time was calculated (Figure 24. D). I analyzed molecular weight of wt *Bs* TrpRS and *Bs* TrpRS Val^{144}Pro by using the same buffer I (as for calibration). Wt *Bs* TrpRS eluted from the column at 12.76 ml (this was taken as an average of 3 runs), responds to the molecular weight of 76,19 kDa and *Bs* TrpRS Val^{144}Pro at 12,715 ml (this was taken as an average of 3 runs), responds to 78 kDa (Figure 24. B, C). Molecular weights of analyzed synthetases suggested that both wt *Bs* TrpRS and *Bs* TrpRS Val^{144}Pro form a dimer in buffer I (for the ITC experiment).

Figure 24: Analytical gel filtration on Superdex 200 10/300 GL (Amersham Biosciences). This experiment indicated that wt *Bs* TrpRS and *Bs* TrpRS Val[144]Pro form a dimer. The gel filtration was performed on ÄKTApurifier UPC (GE Healthcare), with 0.4 ml/min flow rate and 1.5 MPa maximum column pressure, in a room temperature (in buffer I: 50 mM Tris (pH 8), 300 mM NaCl, 2% glycerol, 0.5 mM EDTA). The elution profile was recorded at 280 nm. **A** – gel filtration standard (*BIO RAD*) elution profile that indicates that; protein aggregates eluted from the column at 7.62 ml, thyroglobulin at 8.25 ml, bovine γ-globulin at 11.34 ml, chicken ovalbumin at 14.24 ml, equine myoglobin at 16.21 ml and vitamin B12 at 19.81. **B** - wt *Bs* TrpRS that eluted at 12.77 ml. **C** – *Bs* TrpRS Val[144]Pro that eluted at 12.71 ml. **D** – calibration curve with gel filtation standard (*BIO RAD*). **E** - Molecular weight - MW [kDa], logarithm of the molecular weight – Log (MW) and the elution time – V [ml] values for the *BIO RAD* standard, wt *Bs* TrpRS and *Bs* TrpRS Val[144]Pro. Wt *Bs* TrpRS and *Bs* TrpRS Val[144]Pro MW values were calculated from the linear function of the logarithm of the standard molecular weight on the elution time. Wt *Bs* TrpRS responds to MW = 76.19 kDa and *Bs* TrpRS Val[144]Pro to MW = 78 kDa. These values are also indicated in bold.

A.

B.

C.

D.

Calibration curve

Legend: ◆ 8.25 ◆ 11.34 ◆ 12.715 ◆ 12.76 ◆ 14.24 ◆ 16.21 ◆ 19.81
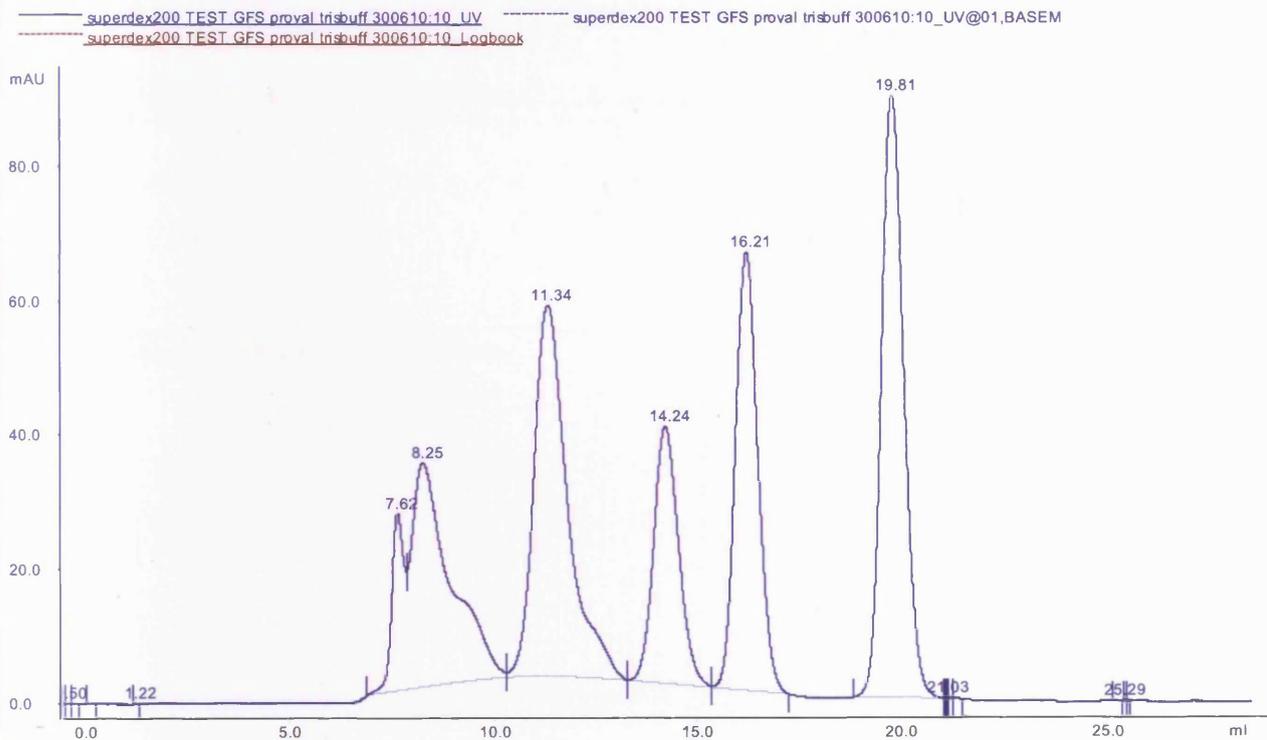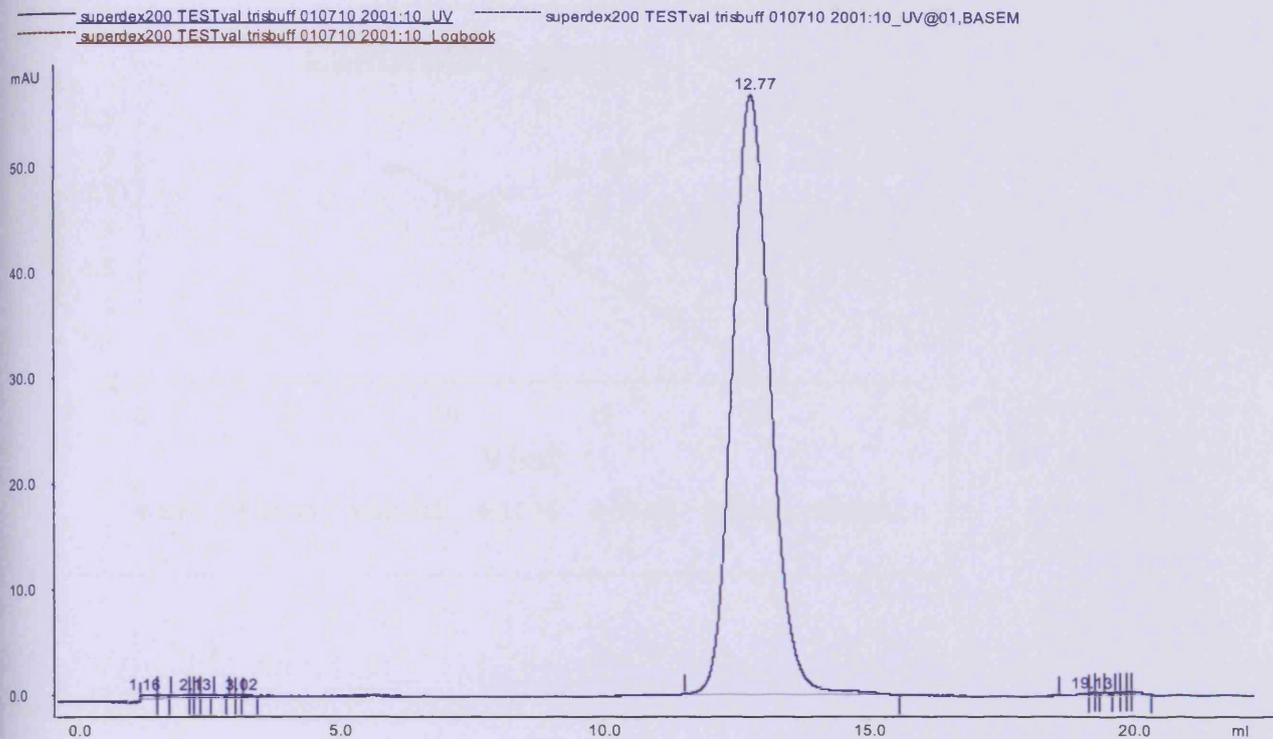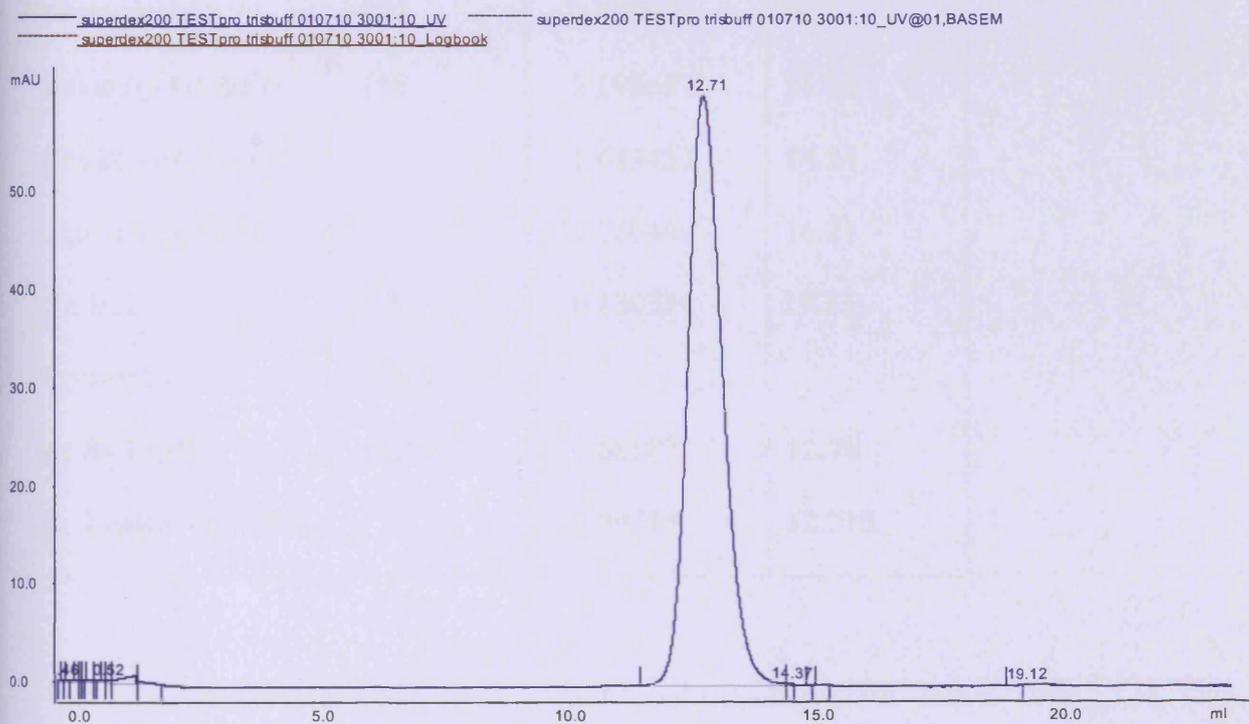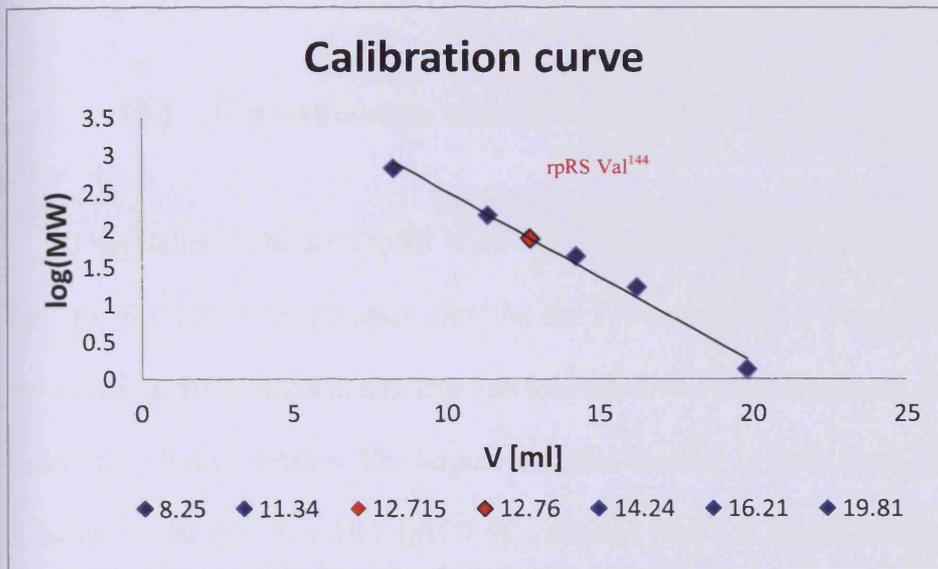
E.

Protein	MW [kDa]	Log (MW)	V [ml]
Protein Aggregates			7.62
Thyroglobulin	670	2.826075	8.25
Bovine γ- Globulin	158	2.198657	11.34
Chicken ovalbumin		1.643453	14.24
Equine myoglobin		1.230449	16.21
Vit B12	1.35	0.130334	19.81
Tryptophan	0.204		
wt *Bs* TrpRS	**76,19**	**1.881873**	**12.76**
***Bs* TrpRS Val^{144}Pro**		**1.89214**	**12.715**

10. Crystal structure of *Bs* TrpRS Val[144]Pro

10.1 Crystallization and data collection

I crystallized the *Bs* TrpRS Val[144]Pro + tag mixed with equal concentrations of L-Trp and 5-OH Trp. (EC 6.1.1.2) at temperature 16°C by the vapour diffusion technique. A Cartesian robot was used to pre-screen crystallization conditions and identified two promising lead conditions. Both of them could be scaled up to larger volume. The largest and best-looking crystals were grown by mixing 9 mg/ml protein in buffer C: 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM 2-Mercaptoethanol, 0.5 mM EDTA, 0.5 mM L-Trp, 0.5 mM 5-OH Trp, with an equal volume of buffer 0.1 M MES (pH 6.5), 12% PEG 20 000 and equilibrated against this buffer in the reservoir solution. Smaller and less regular crystals could also be obtained by using the same buffer C, mixing with a half volume of buffer 25% PEG 1500, 66.4 mM SPG* (pH 10), 33.6 mM SPG* (pH 4) in a crystallization drop and equilibrated against this buffer in the reservoir (SPG* = 1 M succinic acid, 1 M $NaH_2PO_4 \cdot H_2O$, 1 M glycine). The larger and nicer crystals turned out to be very difficult to protect for flash cryo-cooling. Eventually, with help of Prof. Matthias Bochtler we found a buffer that allowed us to record and index a diffraction pattern on a synchrotron beamline, but the pattern was nevertheless clearly marred by the problems in the cryo-protection step. The smaller and less attractive crystals could be cryo-protected straightforwardly (0.05 M $MgCl_2 \cdot 6H_2O$, 0.05 M HEPES sodium pH 7, 25% PEG 550, 14% glycerol), but diffracted only to lower resolution. Although the two groups of crystals were not superficially similar in morphology, a comparison of the indexing results suggested that all crystals belonged to the same crystal form. Therefore, the cryo-buffer for the crystals of less attractive morphology was tested for the larger crystals and turned out to work. Data was collected by Sam Hart, integrated and scaled by Prof. Matthias Bochtler with XDS and SCALA (CCP4-6.1.3).

10.2 Crystal structure determination by molecular replacement

Bs TrpRS Val^{144}Pro + tag crystals were orthorhombic and with the synchrotron radiation diffracted to 2.8 Å. Indexing, done by Prof Matthias Bochtler, yielded a unit cell with dimension 49.2 × 119.8 × 127.3 Å. Based on solvent content; crystals were expected to contain two protomers in the asymmetric unit. Due to non-measured data on the reciprocal coordinate axis, which was later assigned as 0k0, the pattern of systematically absent reflections on this axis was unknown and thus the space group remained ambiguous (P2(1)2(1)2(1) or P2(1)22(1) equivalent to P2(1)2(1)2 in the standard nomenclature). The number of reflections measured for the axis later assigned as 001 was also quite limited. Due to the uncertainty about the two 2-fold axes, translation searches were separately carried out for all orthorhombic space groups. For molecular replacement searches (Prof Matthias Bochtler), we used the MOLREP program. As a search model, we used a monomer of the TrpRS from *B. stearothermophilus* (PDB ID 3FIO),[207] which is 76% identical in sequence to the *B. subtilis* homologue, but did not obtain clear-cut results. Assuming the conservation of the dimer interface, we next tried to position the full *B. stearothermophilus* dimer in the unit cell. The rotation search now yielded two rotational solutions (corresponding to the two ways to superimpose homologous dimers) that stood out from the noise. For these rotational solutions, the translation function scores and R-factors stood out from the noise for space group P2(1)2(1)2(1), but not for other screw axes assignments. Inspection of the molecular replacement solution showed satisfactory molecular packing in all three dimensions of the crystal.

10.3 Refinement

In spite of the very significant sequence similarity between the template and the target structure, the initial R-factor for the molecular replacement solution was rather high, indicating that movements and other local changes needed to be corrected manually (done by me under supervision of Dr Honorata

Czapinska). This was done with the modeling program COOT. Cycles of model improvement were alternated with refinement cycles using the REFMAC maximum likelihood refinement program. The Ramachandran plot statistics showed quality of the structure refinement. 91.9 % of the peptide dihedral angles are in the most favoured regions, 7.7 % in the additionally allowed regions, 0.4 % in the generously allowed region and 0 % in the disallowed region. Due to the limited resolution of the diffraction data, no attempt was made to build water molecules in the structure. We also did not attempt to use global anisotropic temperature factors (TLS) because refinement appeared satisfactory also in the absence of these assumptions. Crystal model evaluation was done with PROCHECK and MolProbity.

The crystal structure, atomic coordinates and structure factors have been deposited as; tryptophan-tRNA synthetase Val144Pro mutant from *B. subtillis*, in Protein Data Bank (PDB ID 3PRH).[204] The statistics for the best dataset, collected at the Diamond Light Source in the UK, are presented in Table 14.

Table 14: Data collection and refinement statistics

Data collection statistics	
Space group	P2(1)2(1)2(1)
a (Å)	49.258
b (Å)	119.816
c (Å)	127.353
Resolution range (Å)	30.0 - 2.8
Total reflections	137536
Unique reflections	19307
Completeness (%) (last shell)	99.9 (100.0)
I/σ (last shell)	13.5 (2.6)
R(sym) (%) (last shell)	5.0 (30.0)
B(iso) from Wilson (Å^2)	71.8
Refinement statistics	
Protein atoms excluding H	4958
Solvent molecules	0
R-factor (%)	23.52
R-free (%)	27.50
Rmsd bond lengths (Å)	0.012
Rmsd angles (°)	1.3
Ramachandran core region (%)	91.9
Ramachandran allowed region (%)	7.7
Ramachandran additionally allowed region (%)	0.4
Ramachandran disallowed region (%)	0.0

10.4 Overall structure

Bs TrpRS (Figure 25) is a typical homodimeric α2 enzyme. As in the case of the Carter's *Bst* TrpRS (PDB ID 3FIO),[207] each monomer of *Bs* TrpRS contains two unequal domains. The bigger domain (residues 1-200) has a Rossmann fold (RF) built of 4 – stranded parallel ß-strands linked with α-helical

108

crossover connections. In this region, we noticed the characteristic TIGH and KMSKS consensus sequences (in the N- and C-terminal parts of RF, respectively). The small domain (SD; residues 207-280) consists of 4 high motion helices with one of them stretching through the whole length of a monomer (residues 265-326) and terminating at the dimer axis.
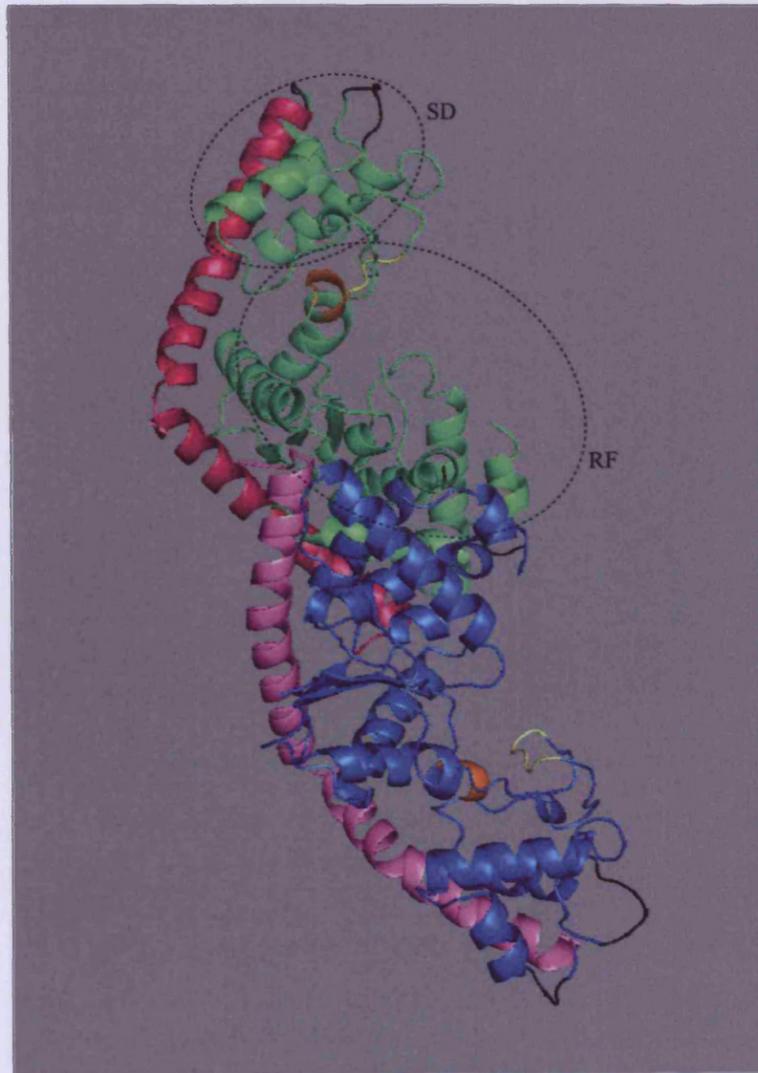
Figure 25: *Bs* **TrpRS Val[144]Pro homodimer crystal structure model.** Two homologue monomers are distinguished as blue and green. Each monomer has its bigger domain, the Rossmann dinucleotide-binding region comprising a Rossmann fold (RF) and a small domain (SD). Two important consensus sequences are TIGH (orange) and KMSKS (yellow); and regions with a great mobility are in black. A long helix (pink or light pink) expands from the SD onto the dimer interface.

10.5 Active site

Because the crystal of the *Bs* TrpRS Val^{144}Pro mutant diffracted to only 2.8 Å, we could not find electron density for either L-Trp or 5-OH-L-Trp in the synthetase active site (Figure 26). However, we superimposed the structures of *Bs* TrpRS Val^{144}Pro and *Bst* TrpRS (PDB ID 3FIO)[207] and found the expected ligand binding position (Figure 27).
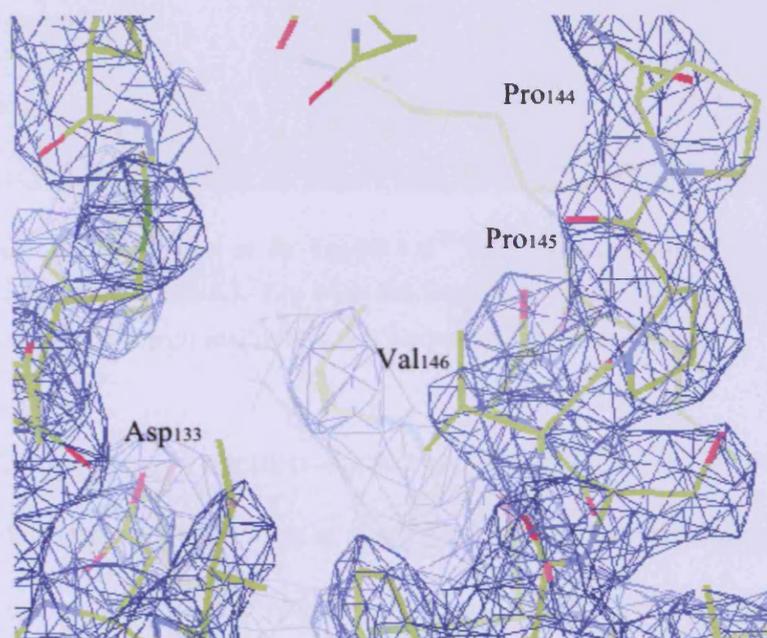
Figure 26: *Bs* TrpRS Val^{144}Pro active site electron density map.
Positions of Asp^{133}, Pro^{144}, Pro^{145} and Val^{146} are indicated. This is the site where a ligand would be expected to bind.
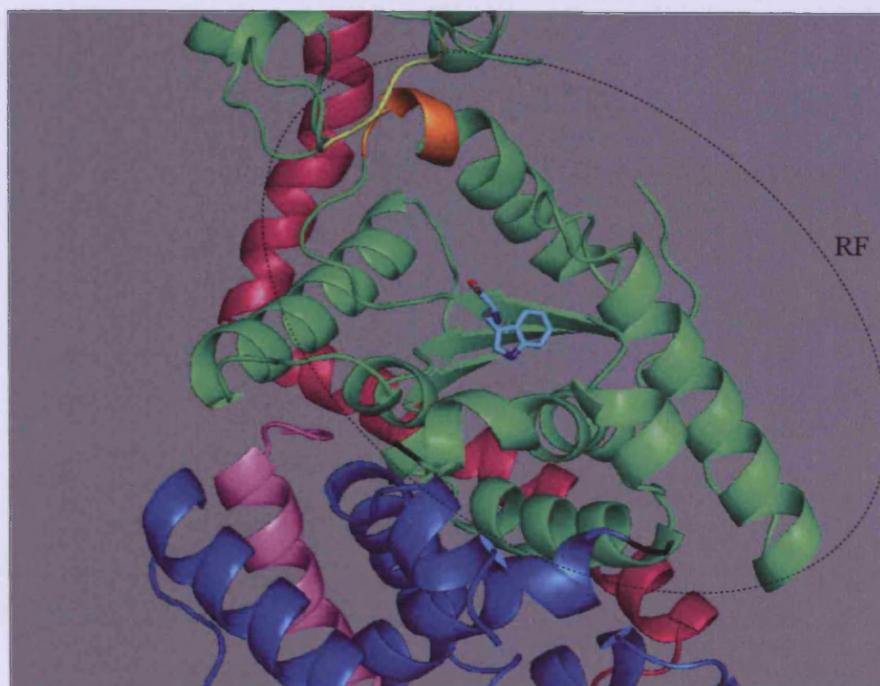
Figure 27: Stereoview of *Bs* TrpRS Val¹⁴⁴Pro Rossmann fold binding site (green) with a modeled L-Trp from *Bst* TrpRS structure. TIGH and KMSKS are coloured in orange and yellow; and the long extended helix in pink.

After the superposition of the two synthetase structures (*Bs* TrpRS Val¹⁴⁴Pro and *Bst* TrpRS - PDB ID 3FIO[207]), we were able to have a closer look at any differences in the position of active site elements. In particular, we were curious whether the position of Asp¹³³ (equivalent of Asp¹³² in *Bst* TrpRS), the main L-Trp determinant, was altered due to the Val¹⁴⁴Pro mutation. However, we did not detect any significant differences in the conformation of this residue (Figure 28). Despite the fact that residue 144 is only about 6 Å from 5-position of L-Trp, the Val¹⁴⁴Pro mutation does not have a considerable effect on the synthetase secondary structure. However, we were able to detect a loop disposition in place of Val¹⁴⁴Pro mutation.
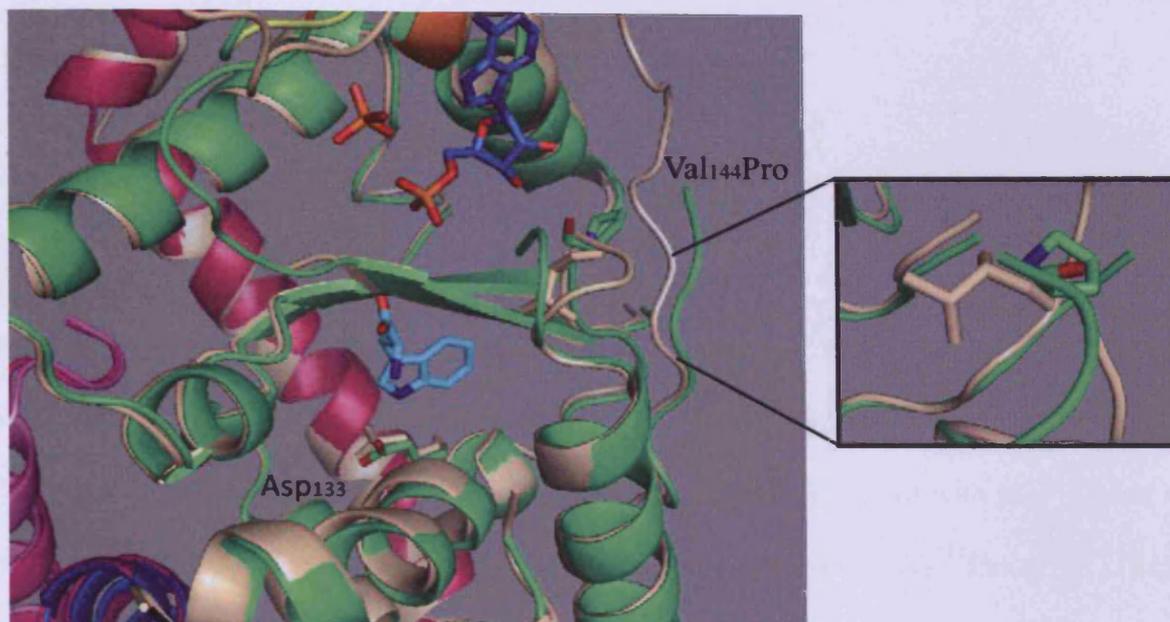
Figure 28: *Bs* TrpRS Val^{144}Pro and *Bst* TrpRS (brown) structures alignment. L-trp (cyan blue), AMP (dark blue) and inorganic phosphate (orange) comes from *Bst* TrpRS structure and no ligand was present in *Bs* TrpRS Val^{144}Pro structure. RF comprises Val^{144}Pro mutation (green) from *Bs* TrpRS Val^{144}Pro and Val^{143} (brown) from *Bst* TrpRS that are magnified in an extra window. In this area, we also see the loop disposition caused by the mutation. Asp^{133} residue has not changed its position and is identical to Asp^{132} in *Bst* TrpRS.

From our analysis of the unmodified Asp^{133} in *Bs* TrpRS Val^{144}Pro active site, it appears that this gatekeeper residue in an aminoacylation reaction would recognize L-Trp and it is unlikely that the synthetase mutant would switch its specificity to *only* 5-OH Trp. This does not correspond to what Schultz has described.[76] In terms of the Val^{144}Pro mutation, previous work[207] suggested that this mutation gives rise to some conformational changes in the place where Asp^{146} binds the ribose of ATP. This could have an effect on the catalytic efficiency of the synthetase. In order to investigate how Val^{144}Pro mutation does influence the aminoacylation, I carried on with the complementation assay.

11.   Complementation assay results

I used *E. coli* KY4040 strain (6.3)[208,209] to test the activity of exogenous *Bs* TrpRS's. This auxotrophic strain displays a mutation in the *trpS* structural gene and therefore TrpRS is unstable. Growth of the strain could be stabilized by access of tryptophan in the presence of ATP and the growth rate depends on the concentration of tryptophan.[208] In order to complement the auxotrophic phenotype under Trp-starved conditions, I co-transformed *E. coli* KY4040 with pGP1-2 and one of three pET151 plasmids (i.e. pET151 wt *Bs* TrpRS, pET151 *Bs* TrpRS Val^{144}Pro or pET151$_e$) (6.1, 6.2). In order to facilitate T7 polymerase expression in pGP1-2 (6.1), under the temperature sensitive cI857 repressor, I grew cells at 37°C. Only in presence of wt *Bs* TrpRS, KY4040 cultures were complemented and grew after 48 h, either in presence of 0.05 μM L-Trp or 5 μM 5-OH Trp (Figure 29). This result confirmed that 5-OH Trp is substrate for wt *Bs* TrpRS as oppose to the original report.[76] There was no growth detected in presence of *Bs* TrpRS Val^{144}Pro and L-Trp or 5-OH Trp, which was identical with the control experiment (where no synthetase was expressed) (Figure 29). This result showed that *Bs* TrpRS Val^{144}Pro has no ability to charge a tRNA^{Trp} with either L-Trp or 5-OH Trp.

All experiments were repeated 3 times.
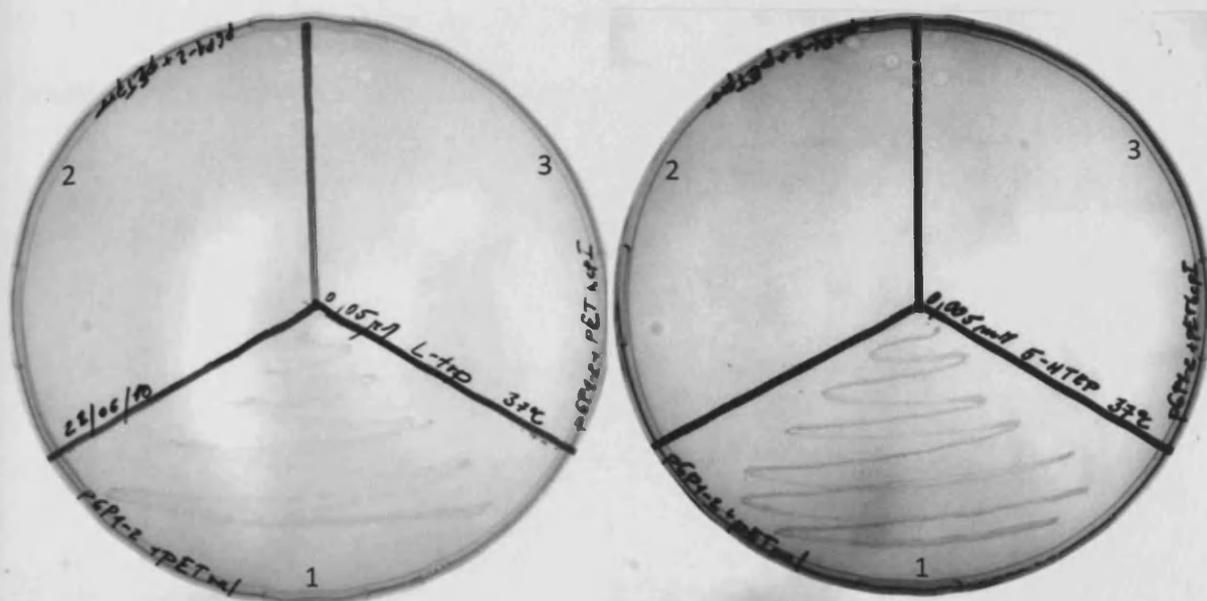
Figure 29: *E. coli* **KY4040 complementation assay in presence of 0.05 μM L-Trp (left) and 5 μM 5-OH Trp (right).** Cell cultures were grown in three sections on M9 minimal media (6.7) plates containing 50 μg/ml ampicilin, 25 μg/ml kanamycin, at 37°C/48 h. 1 - pET151 wt *Bs* TrpRS /pGP1-2, 2 - pET151 TrpRS Val^{144}Pro/ pGP1-2, 3 - pET151$_e$ (control)/pGP1-2.

A complementation assay was used by Soll et al.[69] in order to evaluate *in vivo* activities of wt and mutant *Bst* TrpRS enzymes. Growth of different transformants was compared after 1, 2 and 4 days. Similarly to our *Bs* TrpRS Val^{144}Pro, *Bst* TrpRS Val^{141}Gln mutant was also not able to complement KY4040 phenotype after 4 days while wt *Bst* TrpRS grew after 1 day. As was already mentioned in chapter 4.4, the catalytic efficiency of the *Bst* TrpRS Val^{141}Gln mutant was reduced ~300 fold for L-Trp and ~1000 fold for ATP. Previously studied conformational changes of Asp^{146}, binding the ribose of ATP,[207] in the Val^{144}Pro mutant suggested the explanation for this catalytic efficiency reduction. Our results only confirmed what was already proposed. The Val^{144}Pro mutation in *Bs* TrpRS perhaps affects the ATP ribose binding and therefore lowers the catalytic efficiency of the synthetase. Charging of tRNA with either L-Trp or 5-OH Trp is disturbed and therefore *Bs* TrpRS Val^{144}Pro is not able to complement the auxotrophic phenotype. Thus our results are in disagreement with the original report[76]. They reported 20% - 40 % translation efficiency and 97% translation

fidelity of 5-OH Trp incorporation by *Bs* TrpRS Val^{144}Pro and, thus, should have successfully

complemented the phenotype of this strain in the presence of 5-OH Trp.

12. Isothermal titration calorimetry (ITC) for wt *Bs* TrpRS and *Bs* TrpRS Val[144]Pro analysis

In order to examine the affinity of *Bs* TrpRS wt and *Bs* TrpRS Val[144]Pro towards L-Trp and 5-OH Trp, I applied ITC.

ITC is a powerful and adaptable technology that measures heat (q) of a binding interaction in solution.[210] ITC directly determines association constant (K_a), enthalpy changes (ΔH) and binding stoichiometry (n). From those, one can calculate dissociation constant (K_d), Gibbs free energy change (ΔG), and entropy change (ΔS) of a reaction. Moreover, ITC can also measure catalytic reactions, conformational rearrangements and changes in protonation. Defining of a full binding curve depends on ligand solubility and magnitude of binding affinity[74]. ITC is applicable in biomolecular science, drug development and materials engineering.[210] Focus of this work is on a protein - synthetase interaction with a ligand - amino acid.

ITC works on a simple principle.[210] Figure 30 shows a design of the typical ITC instrument. There are two identical cells, held in temperature equilibrium. A sample cell is a place where the macromolecule is applied, whereas a reference cell always contains water or buffer. The sample cell is temperature controlled. A thermopile/thermocouple circuit connects two cells. Throughout the experiment, an instrumental syringe injects a ligand solution into the sample cell, at a fixed temperature in the range 5-60 °C.[74] The q of this reaction is either released or taken up (exo- or endo-thermic process). This thermal difference is detected by semi-conductor thermopiles. Subsequently, the instrument releases a thermal power ($\mu cal/s$), in order to compensate for the heat change. This brings the cells back into the temperature equilibrium.
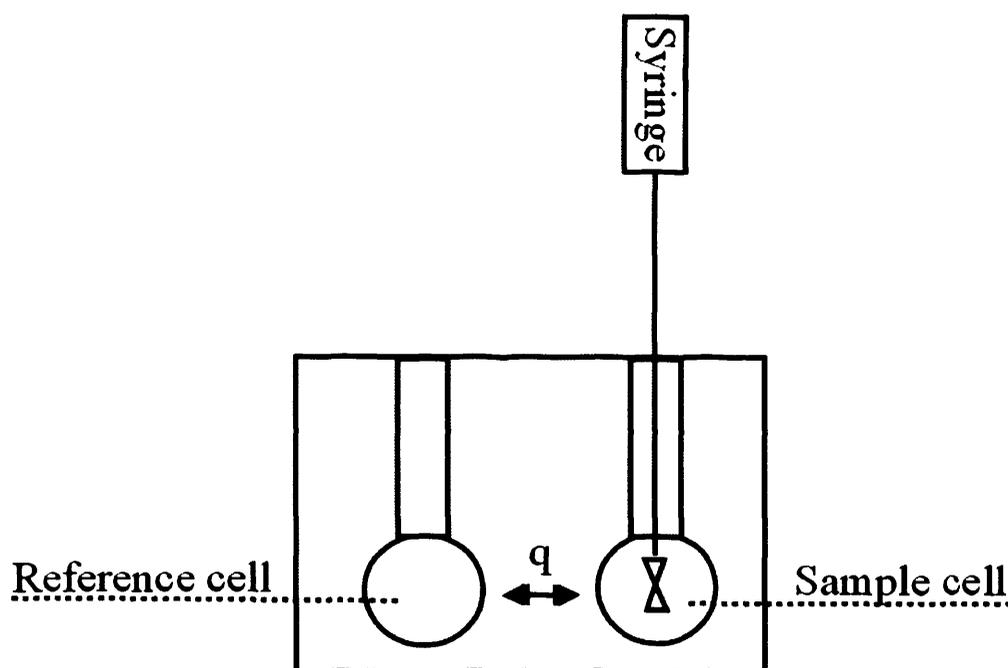
Figure 30: Typical ITC instrument set up. The sample cell and the reference cell are kept in the temperature equilibrium. The sample cell contains a macromolecule solution into which a ligand solution is being injected from the syringe. Whilst the experiment runs and single injections are applied, syringe stirs in the sample cell.

ITC measures amount of the reaction heat for every injection with sensitivity to several hundred nanojoules. ΔH is directly calculated by integrating the thermal power *versus* time.[210] A typical experiment is projected as a number of peaks, one peak for every injection. Furthermore, the integration of each peak (and correction to a per mole basis) defines specific points of a binding curve. Identifying of a binding curve enables to determine degree of complex formation. In this manner, K_a can be calculated and K_d derived from the equation; K_a (K_d^{-1}). Moreover, from these values, one can calculate (ΔG) and (ΔS) as follows:[210]

$$\Delta G = -RT \ln K_a; \qquad \Delta G = \Delta H - T\Delta S$$

Besides, one can run the experiment in different temperatures and determine the heat capacity of binding (ΔC_p):

$$\Delta C_p = d(\Delta H) / dT$$

ITC technology has developed largely so that a wide profile of an enzyme binding reaction can be obtained.[211-213] Despite of this, ITC for enzyme binding kinetics analyzes has not been opened up much yet. However, in this project, the main interest was to determine K_d for wt *Bs* TrpRS and *Bs* TrpRS Val[144]Pro mutant in reactions with either L-Trp or 5-OH Trp.

ITC have been used in several cases to determine activities of number of synthetases. The importance of two tryptophan residues was determined by ITC on human T2-TrpRS and its multifunctionality was proven by this method.[75] T2-TrpRS binds L-Trp with K_d of 11 μM. Activities of different T2-TrpRS mutants towards L-Trp were shown to be reduced or not measurable. In this way, crucial residues for the synthetase activity were found. By using the same human T2-TrpRS and ITC method, activity towards L-Trp and ATP was measured and compared to the full-length TrpRS.[214] It was found that T2-TrpRS and TrpRS have comparable affinities for L-Trp (11.3 μM and 6.8 μM) but not similar for ATP. TrpRS binds ATP with K_d of 20 μM but T2-TrpRS > 0.5 mM. Furthermore, ITC was applied to examine the effect of Mg^{2+} presence on ATP titration with TrpRS from *Bacillus stearothermophilus*.[215]

When designing the ITC experiment, several requirements needed to be considered.[210] Firstly, it is the sample volume. Except for low volume designed calorimeters, ITC requires a large volume of sample. In our experiments I used VP-isothermal titration calorimeter (MicroCal), with ~ 1,4 ml sample cell volume. For an easy manipulation, one experiment requires ~ 3 ml of a protein sample. Secondly, because ITC measures overall reaction heat, additional heat that does not belong to the reaction itself needs to be subtracted. For example, this additional heat could be coming from protonation or salvation conformational transition. Therefore, additional dilution experiments are required to be run before the actual experiment and subtracted. In our case, such dilution experiments were; buffer into buffer, ligand solution into buffer and buffer into the synthetase solution, but they did not result in significant heat release. For this reason, I did not deduct this dilution heat of the additional experiments but the heat from the last injection of the actual experiment was subtracted

118

from every injection instead. In the third place, we need to consider that not every reaction results in a large signal to noise. This can be resolved by changing the reaction temperature, as ΔH is generally temperature-dependence. In our experiments with synthetases from *Bacillus subtilis*, the temperature test is limited to only physiological temperatures, between 25-35 °C. However, I tried two different temperatures, 25°C and 30°C, but the difference was trivial for the objective of our study. Because of some precipitation formations in our reaction mixtures during the ITC runs, I sticked to only a "friendlier" 25°C temperature. On the other hand, increased concentration of the macromolecule does enhance the reaction heat and therefore signal to noise, in our experiments. For this reason, the synthetase concentration I used ranges between 81 μM- 181 μM. Finally, ITC can measure K_d value in range between $\sim$ 1nM $< K_d < \sim$ several hundred μM. If K_d of a macromolecule is too low (tight binding) and the concentration is too high, the most of the binding sites saturate in the first injection. In this case, only one-step reaction is achieved and the binding curve cannot be analyzed. This can be resolved by the macromolecule sample dilution but only until the point that ITC can still detect the signal-noise difference. In contrast, it is also challenging to measure very high K_d of a macromolecule (weak interaction)[216] binding to a ligand, because of both concentrations (macromolecule and ligand) need to reach binding saturation. Yet again, this can be resolved by concentrating of the macromolecule sample but only until the sample is still soluble.

In our study, *Bs* TrpRS is relatively a poor binder when compared to *Ec* TyrRS (results not shown). Limitations of studying poor binding complexes by ITC have been well explored.[217] In order to proceed with ITC experimental data, a nonlinear least squares curve fitting needs to be applied that faithfully describe the binding event. Since the 1980s,[218] when highly sensitive microcalorimeters came onto the market, applying the so-called Wiseman isotherm[219] for the fitting process has become common. This was after Wiseman et al. pointed out that the shape of such binding isotherm changes depending on K_a and the concentration of macromolecule ([M]t) that defines a so-called c value. The c value can be also taken as the ratio of [M]t and K_d.

$$c = K_a[M]t = [M]t/K_d$$

Working with systems where the c value is between $10 - 500$ is highly preferable.[219, 220] Binding

isotherm is sigmoidal if $c > 10$ (Figure 31).[217] Therefore, the entire binding curve can be fitted and the

final points are almost in a vertical order due to a full saturation. It is more difficult to fit a binding

curve, if $c < 10$, due to the curve incompleteness and a challenge to achieve a full macromolecule

saturation (Figure 31).[217]
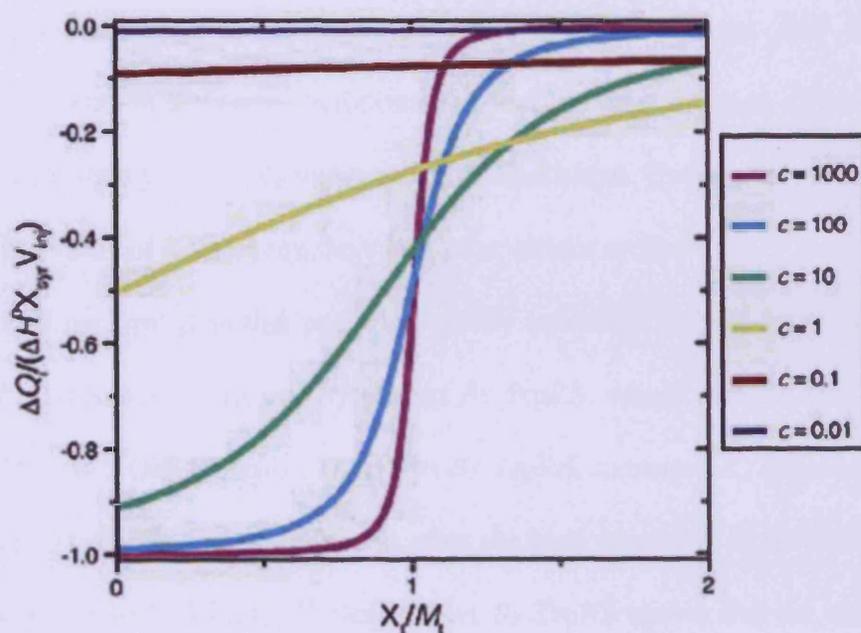
Figure 31: Wiseman isotherms projected by Turnbull et al.[217] that change their shape as the c value changes. C value is defined by Ka and [Mt]. When $c < 10$, the binding curve loses its sigmoidal shape.

For the low affinity systems, c value is in a range between $0.01 - 10$.[217] When dealing with these

systems ($K_a < 10\ 000\ M^{-1}$; $K_d > 100\ \mu M$), solubility usually becomes a significant issue. This is

because higher concentrations of macromolecule and ligand ($[M]t \gg K_d$) is required in order to

satisfy the c value rule. Even if $[M]t \ll K_d$ and we cannot improve the $[M]t$, the final concentration of

ligand has to be several times higher than K_d. Nevertheless, it was shown that it is possible to

determine K_a and therefore ΔG^0 for the low affinity systems, with four following conditions.[217] (1)

Significant amount of the binding isotherm is required to use for fitting. (2) Stoichiometry of a

reaction is required to be known. (3) Concentrations of both macromolecule and ligand are well determined. (4) ITC experiment needs to result a sufficient signal to noise difference.

In our experiments, I used the nanodrop and Bradford assay to determine the macromolecule concentration. Yet, for the low affinity systems when repeating the same experiment using the same concentrations of the macromolecule and ligand, the value of ΔH^o fluctuates. This can be due to the error in the concentration measurements. An earlier suggestion[217] was to fix the n value in fitting and then the differences in ΔH^o may go unnoticed. However, despite the ΔH^o fluctuations, when I repeated our experiments in the same conditions and similar concentrations of the macromolecule and ligand, I found a constancy of the K_d value in each experiment. Eventually, this was what we aimed for, to compare the value of K_d between the wt and the variant synthetases.

ITC titrations performed in this project typically indicated one binding event. In the first ITC experiment, I titrated 5 mM L-Trp into 81 µM wt *Bs* TrpRS, which resulted in K_d = 149 µM (Figure 32). Titration of 5 mM 5-OH Trp into 118 µM wt *Bs* TrpRS, increases K_d approximately 7 folds; K_d = 1079 µM (Figure 33). This is in accordance to what we have expected, as the wt *Bs* TrpRS prefers its natural substrate L-Trp to 5-OH Trp.[184] Besides, wt *Bs* TrpRS shows that the activity towards 5-OH Trp, was as described previously.[184, 195] On the other hand, *Bs* TrpRS Val[144]Pro has reduced activity towards L-Trp; $K_d \geq 2000$ µM, when 7 mM L-Trp is titrated with 151 µM of the protein (Figure 34). The last experiment was the titration of 7 mM 5-OH Trp into 181 µM *Bs* TrpRS Val[144]Pro (Figure 35). Due to a weak interaction and therefore fitting limitations, I estimated the lowest possible K_d value to 2700 µM. Because of the experimental limitations mentioned earlier, I could not use higher concentrations of synthetases and ligands mainly because of a bad solubility and a significant heat of dilution experiments. For all ITC enthalpograms, the upper panel represents the raw ITC data whereas the lower panel shows integrated heat data for every injection. Heat released (kcal/mol) as a function of tryptophan concentration was fitted to a one-site model. ITC data analysis was done by IC-ITC program (Dr Niklaas Buurma).[74, 221]
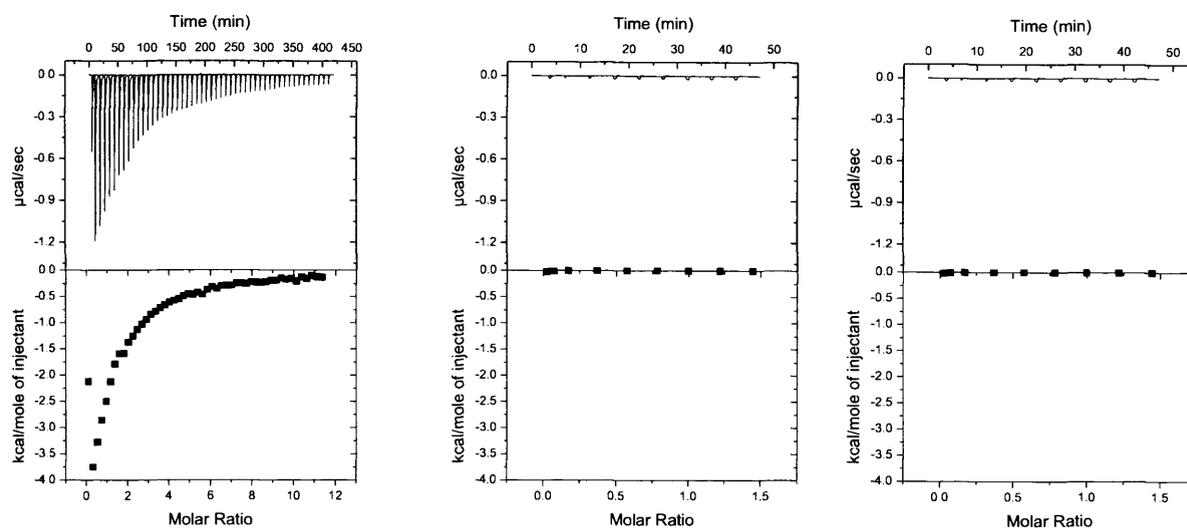
A. $K_d = 149 \ \mu M$

$\Delta S^\circ = -89 \ JK^{-1} mol^{-1}$

B.

C.

Figure 32: A; Titration of L-Trp (5 mM) with wt *Bs* TrpRS (81 μM) at pH 8 (50 mM Tris, 300 mM NaCl, 2% glycerol, 0.5 mM EDTA) at 25°C. Each peak corresponds to 5 μL of ligand injected into the sample cell. B; The heat of dilution when ligand is injected into buffer and C; buffer is injected into wt *Bs* TrpRS in the same conditions as A.
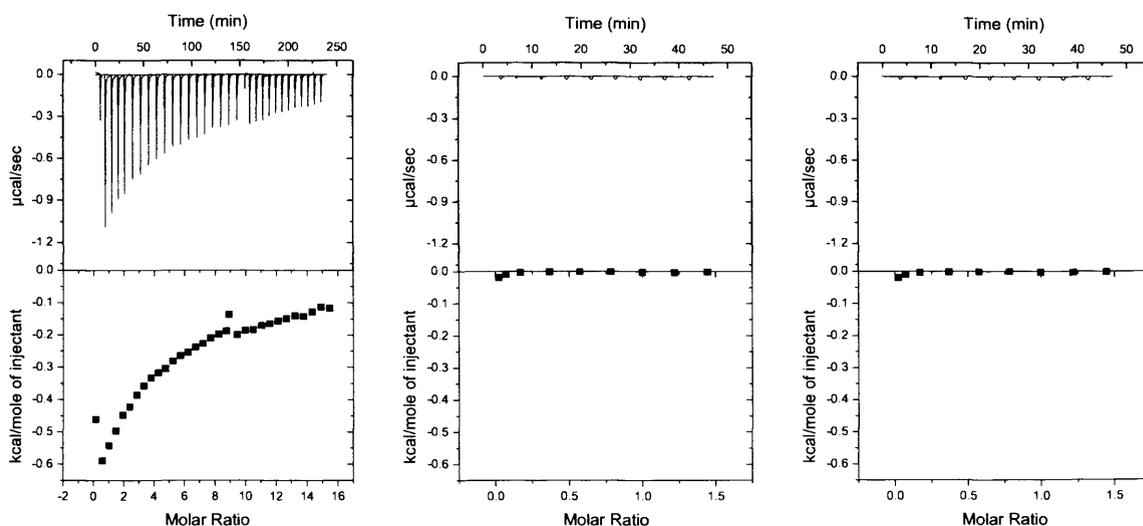
A. $K_d = 1079 \ \mu M$

$\Delta S^\circ = -2.3 \ JK^{-1} \ mol^{-1}$

B.

C.

Figure 33: A; Titration of 5-OH Trp (5 mM) with wt *Bs* TrpRS (118 μM) at pH 8 (50 mM Tris, 300 mM NaCl, 2% glycerol, 0.5 mM EDTA), at 25°C. Each peak corresponds to 10 μL of ligand injected into the sample cell. B; The heat of dilution when ligand is injected into buffer and C; buffer is injected into wt *Bs* TrpRS in the same conditions as A.

A. $K_d \geq 2000\ \mu M$ B.   C.

$\Delta S^\circ = -12.6\ JK^{-1}\ mol^{-1}$
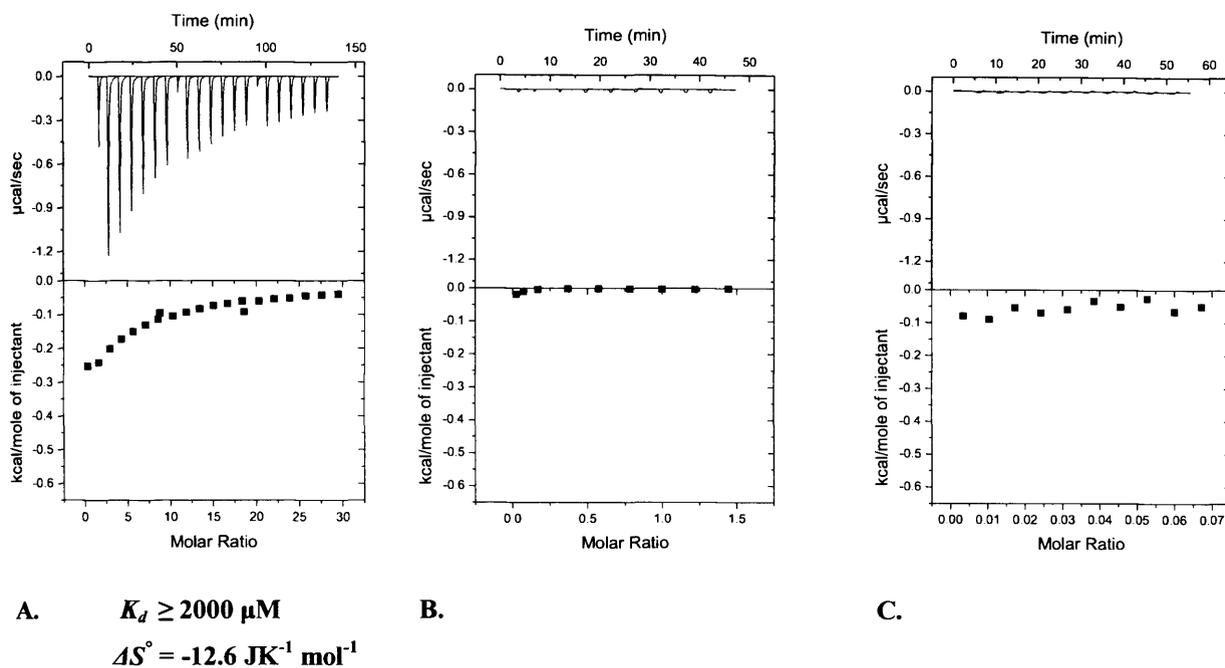
Figure 34: A; Titration of L-Trp (7 mM) with *Bs* TrpRS Val^{144}Pro (151 µM) at pH 8 (50 mM Tris, 300 mM NaCl, 2% glycerol, 0.5 mM EDTA), at 25°C. Each peak corresponds to 40 µL of ligand injected into the sample cell. B; The heat of dilution when ligand is injected into buffer C; buffer is injected into wt *Bs* TrpRS in the same conditions as A.

A. $K_d \geq 2700\ \mu M$ B.   C.

$\Delta S^\circ = -14\ JK^{-1}\ mol^{-1}$
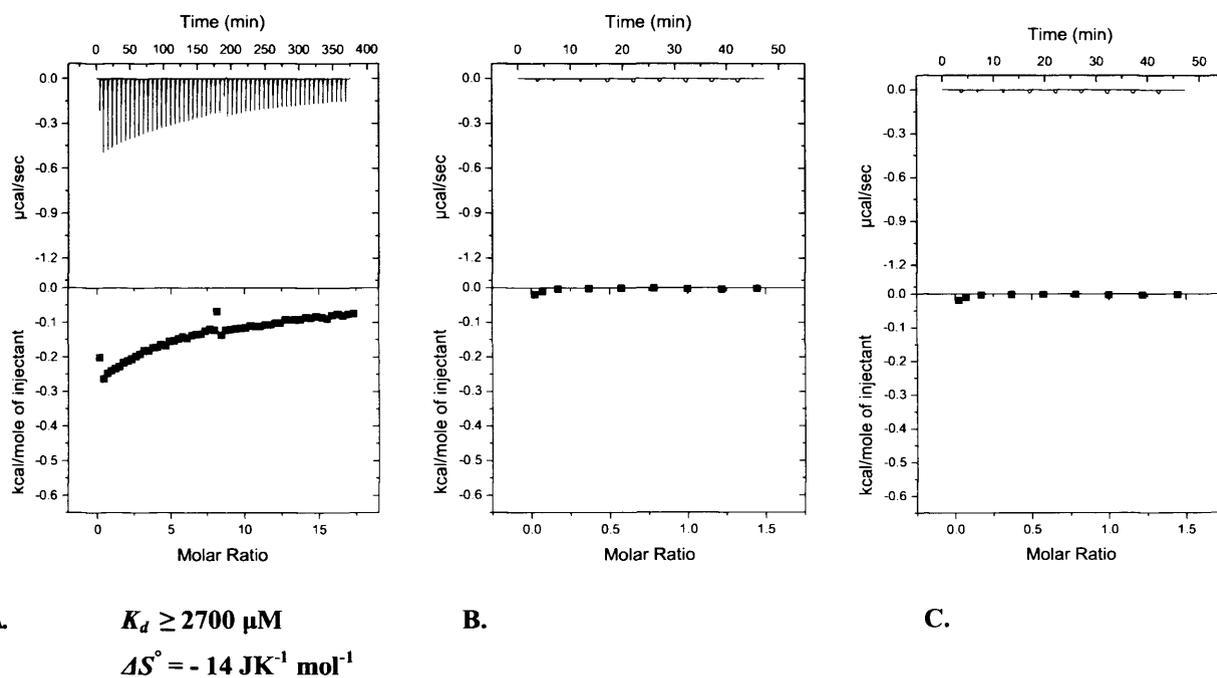
Figure 35: A; Titration of 5-OH Trp (7 mM) with *Bs* TrpRS Val^{144}Pro (181 µM) at pH 8 (50 mM Tris, 300 mM NaCl, 2% glycerol, 0.5 mM EDTA), at 25°C. Each peak corresponds to 10 µL of ligand injected into the sample cell. B; The heat of dilution when ligand is injected into buffer and C; buffer is injected into *Bs* TrpRS Val^{144}Pro in the same conditions as A.

123

Different studies done on TrpRS activity enable us to compare with our results. Wong et al. reported[222] that TrpRS from *Bacillus subtilis* activates L-Trp as well as the three fluorinated analogues, DL-4-fluoro-, DL-5-fluoro-, or DL-6-fluorotryptophan (4F-, 5F-, and 6F-Trp), in the ATP-pyrophosphate exchange reaction. Their relative activities follow the same order: L-Trp > 4F-Trp > 6F-Trp > 5F-Trp (Table 15). On the other hand, this represents the reverse order of relative hydrophobicity and its importance in TrpRS specific recognition of these compounds. K_M values estimated for L-Trp and Trp analogs show that the position of fluoro-substituent is critical for the active site binding. Position 5 seems to be the most unfavorable when considering the interaction of 5-F-Trp with *Bs* TrpRS. From our observations, 5-OH Trp interacting with *Bs* TrpRS is similarly unfavorable.

Table 15: Enzymatic selectivity of *Bs* TrpRS for tryptophan and the fluoroanalogues in ATP-PPi exchange.[222]

Substrate	K_M
L-Trp	164 ± 12
4F-Trp	419 ± 34
5F-Trp	2606 ±188
6F-Trp	1621 ± 169

From the ITC data I have collected, the previous report is to be contradicted.[76] *Bs* TrpRS Val[144]Pro, with $K_d \geq 2000$ µM for L-Trp and $K_d \geq 2700$ µM for 5-OH Trp, displays low binding affinity for both substrates. There is no preference of the mutant synthetase towards 5-OH Trp. Moreover, *Bs* TrpRS Val[144]Pro binds L-Trp and this interaction is even greater than binding with 5-OH Trp. These experiments strongly indicate that 5-OH Trp would not have been incorporated with fidelity reported by Schultz under the conditions they describe.

Conclusions

Without doubt, the revalidation of a scientific work is essential. In this project, we reinvestigated the key elements of the *Bs* TrpRS's achieved fidelity in the expanded genetic code. This investigation was important as its analysis can be used to further introduce new functional groups into proteins by providing a guide for designing future synthetases for other NAAs. In consideration of the original report,[76] we examined the effect of the $Val^{144}Pro$ mutation in *Bs* TrpRS using standard methods i.e. crystallography, complementation assay and ITC and found:

I. The crystal structure of *Bs* TrpRS $Val^{144}Pro$ shed more light on the ligand recognition site. The position of Asp^{133} does not change compared to the homologues residue in the *Bst* TrpRS active site. Therefore, there is no reason to expect a changed fidelity from L-Trp to 5-OH Trp. The active site loop is slightly displaced near the $Val^{144}Pro$ mutation but this had no appreciable effect on the position of Asp^{133}.

II. The mutation of $Val^{144}Pro$ produces catalytically inactive synthetase in the complementation assay therefore the synthetase translation efficiency and/or fidelity are severly diminished towards 5-OH Trp.

III. *Bs* TrpRS $Val^{144}Pro$ exhibits limited affinity towards ligands, L-Trp and 5-OH Trp in ITC experiment, with an actual preference indicated for L-Trp, which is in direct conflict with published data.

The results of our experiments contradict the original report.[76] Despite the fact that we did not follow the same methods as they did, the aim was to collect evidence using standard biochemical means to allow their complex system to be more easily interrogated.

126

Our results are clearly relevant for all attempts to expand the genetic code by nonnatural amino acid mutagenesis. It is instructive to consider the incredible benefits that society has incurred from using standard mutagenesis (i.e. mutagenesis involving 20 common amino acids), and how it has added to the understanding of how proteins function, helped our understanding of the active site mechanisms of enzymes, and has assisted in the design of novel proteins. Thus, given that the inherent benefit from nonnatural amino acid mutagenesis may very well rival standard mutagenesis, our results are important from the standpoint that they should considerably hasten a highly desirable NAA incorporation technology for a mammalian expression system. It would appear that this was not achieved with *Bs* TrpRS Val^{144}Pro mutant that we examined. Therefore, certain conditions need to be addressed if the same approach is to repeated successfully. This includes that one apply what has been already well studied and documented for a given AARS, and to follow the standard methodology for demonstrating high fidelity NAA incorporation.

128

References

1. Cozzone, A. J. Proteins: Fundamental Chemical Properties *ENCYCLOPEDIA OF LIFE SCIENCES* [Online], 2002, p. 1-10.
2. Bock, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F., *Molecular Microbiology* **1991**, *5* (3), 515-520.
3. Srinivasan, G.; James, C. M.; Krzycki, J. A., *Science* **2002**, *296* (5572), 1459-1462.
4. Budisa, N., *Angewandte Chemie-International Edition* **2004**, *43* (47), 6426-6463.
5. Wang, Q.; Parrish, A. R.; Wang, L., *Chemistry & Biology* **2009**, *16* (3), 323-336.
6. Chapeville, F.; Lipmann, F.; Von Ehrenstein, G.; Weisblum, B.; Ray Jr, W. J.; Benzer, S., *PNAS* **1962**, *48*, 1086-1092.
7. Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G., *Science* **1989**, *244* (4901), 182-188.
8. Hendrickson, W. A., *Science* **1991**, *254* (5028), 51-58.
9. Tippmann, E. M.; Liu, W.; Summerer, D.; Mack, A. V.; Schultz, P. G., *Chembiochem* **2007**, *8* (18), 2210-2214.
10. Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X. S.; Zhang, Q.; Hu, H. Y.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H., *Journal of the American Chemical Society* **2008**, *130* (29), 9268-9281.
11. Ibba, M.; Soll, D., *Annual Review of Biochemistry* **2000**, *69*, 617-650.
12. Price, N. C.; Stevens, L., *Fundamentals of enzymology: The cell and molecular biology of catalytic proteins*. Oxford University Press; Oxford University Press: 1999; p. 478.
13. Schimmel, P. R.; Soll, D., *Annual Review of Biochemistry* **1979**, *48*, 601-648.
14. Kraeveskii, A. A.; Kiselev, L. L.; Gottikh, B. P., *Molecular Biology* **1974**, *7* (5), 634-639.
15. Delarue, M.; Moras, D., *Bioessays* **1993**, *15* (10), 675-687.
16. Ibba, M.; Soll, D., *Genes & Development* **2004**, *18* (9), 1106-1106.
17. Stadtman, T. C., *Annual Review of Biochemistry* **1996**, *65*, 83-100.
18. Woese, C. R.; Olsen, G. J.; Ibba, M.; Soll, D., *Microbiology and Molecular Biology Reviews* **2000**, *64* (1), 202-236.
19. Bult, C. J.; White, O.; Olsen, G. J.; Zhou, L. X.; Fleischmann, R. D.; Sutton, G. G.; Blake, J. A.; FitzGerald, L. M.; Clayton, R. A.; Gocayne, J. D.; Kerlavage, A. R.; Dougherty, B. A.; Tomb, J. F.; Adams, M. D.; Reich, C. I.; Overbeek, R.; Kirkness, E. F.; Weinstock, K. G.; Merrick, J. M.; Glodek, A.; Scott, J. L.; Geoghagen, N. S. M.; Weidman, J. F.; Fuhrmann, J. L.; Nguyen, D.; Utterback, T. R.; Kelley, J. M.; Peterson, J. D.; Sadow, P. W.; Hanna, M. C.; Cotton, M. D.; Roberts, K. M.; Hurst, M. A.; Kaine, B. P.; Borodovsky, M.; Klenk, H. P.; Fraser, C. M.; Smith, H. O.; Woese, C. R.; Venter, J. C., *Science* **1996**, *273* (5278), 1058-1073.
20. Smith, D. R.; DoucetteStamm, L. A.; Deloughery, C.; Lee, H. M.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D. Y.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.; Church, G. M.; Daniels, C. J.; Mao, J. I.; Rice, P.; Nolling, J.; Reeve, J. N., *Journal of Bacteriology* **1997**, *179* (22), 7135-7155.
21. Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Soll, D.; Nureki, O., *Nature* **2009**, *457* (7233), 1163-1167.
22. Cusack, S.; Yaremchuk, A.; Tukalo, M., *Embo Journal* **2000**, *19* (10), 2351-2361.
23. Goldgur, Y.; Mosyak, L.; Reshetnikova, L.; Ankilova, V.; Lavrik, O.; Khodyreva, S.; Safro, M., *Structure* **1997**, *5* (1), 59-68.
24. Retailleau, P.; Huang, X.; Yin, Y. H.; Hu, M.; Weinreb, V.; Vachette, P.; Vonrhein, C.; Bricogne, G.; Roversi, P.; Ilyin, V.; Carter, C. W., *Journal of Molecular Biology* **2003**, *325* (1), 39-63.
25. Burbaum, J. J.; Schimmel, P., *Journal of Biological Chemistry* **1991**, *266* (26), 16965-16968.
26. Arnez, J. G.; Moras, D., *Trends in Biochemical Sciences* **1997**, *22* (6), 211-216.

27. Schimmel, P., *Annual Review of Biochemistry* **1987**, *56*, 125-158.
28. Lloyd, A. J.; Thomann, H. U.; Ibba, M.; Soll, D., *Nucleic Acids Research* **1995**, *23* (15), 2886-2892.
29. Berg JM, T. J., Stryer L., *Biochemistry. 5th edition*. W H Freeman: New York, 2002.
30. Vasil'eva, I. A.; Moor, N. A., *Biochemistry-Moscow* **2007**, *72* (3), 247-263.
31. Gribskov, M.; McLachlan, A. D.; Eisenberg, D., *PNAS* **1987**, *84* (13), 4355-4358.
32. Cusack, S.; Berthetcolominas, C.; Hartlein, M.; Nassar, N.; Leberman, R., *Nature* **1990**, *347* (6290), 249-255.
33. Ruff, M.; Krishnaswamy, S.; Boeglin, M.; Poterszman, A.; Mitschler, A.; Podjarny, A.; Rees, B.; Thierry, J. C.; Moras, D., *Science* **1991**, *252* (5013), 1682-1689.
34. Rould, M. A.; Perona, J. J.; Soll, D.; Steitz, T. A., *Science* **1989**, *246* (4934), 1135-1142.
35. Brunie, S.; Mellot, P.; Zelwer, C.; Risler, J. L.; Blanquet, S.; Fayat, G., *Journal of Molecular Graphics* **1987**, *5* (1), 18-21.
36. Brick, P.; Blow, D. M., *Journal of Molecular Biology* **1987**, *194* (2), 287-297.
37. Carter, C. W., *Annual Review of Biochemistry* **1993**, *62*, 715-748.
38. Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D., *Nature* **1990**, *347* (6289), 203-206.
39. Ribas de Pouplana, L.; Schimmel, P., *Cell* **2001**, *104* (2), 191-193.
40. Ibba, M.; Morgan, S.; Curnow, A. W.; Pridmore, D. R.; Vothknecht, U. C.; Gardner, W.; Lin, W.; Woese, C. R.; Soll, D., *Science* **1997**, *278* (5340), 1119-1122.
41. Cassio, D.; Waller, J. P., *European Journal of Biochemistry* **1971**, *20* (2), 283-300.
42. Webster, T.; Tsai, H.; Kula, M.; Mackie, G. A.; Schimmel, P., *Science* **1984**, *226* (4680), 1315-1317.
43. Hountondji, C.; Dessen, P.; Blanquet, S., *Biochimie* **1986**, *68* (9), 1071-1078.
44. Ravel, J. M.; Wang, S. F.; Heinemey.C; Shive, W., *Journal of Biological Chemistry* **1965**, *240* (1), 432-438.
45. Schmitt, E.; Panvert, M.; Blanquet, S.; Mechulam, Y., *Nucleic Acids Research* **1995**, *23* (23), 4793-4798.
46. Fersht, A. R., *Biochemistry* **1987**, *26* (25), 8031-8037.
47. Fersht, A. R.; Knilljones, J. W.; Bedouelle, H.; Winter, G., *Biochemistry* **1988**, *27* (5), 1581-1587.
48. Fersht, A. R.; Shindler, J. S.; Tsui, W. C., *Biochemistry* **1980**, *19* (24), 5520-5524.
49. Hartman, M. C. T.; Josephson, K.; Szostak, J. W., *PNAS* **2006**, *103* (12), 4356-4361.
50. O'Donoghue, P.; Luthey-Schulten, Z., *Microbiology and Molecular Biology Reviews* **2003**, *67* (4), 550-575.
51. Ibba, M.; Soll, D., *Science* **1999**, *286* (5446), 1893-1897.
52. Dale, T.; Uhlenbeck, O. C., *Trends in Biochemical Sciences* **2005**, *30* (12), 659-665.
53. Ataide, S. F.; Ibba, M., *Acs Chemical Biology* **2006**, *1* (5), 285-297.
54. Starzyk, R. M.; Webster, T. A.; Schimmel, P., *Science* **1987**, *237* (4822), 1614-1618.
55. Nureki, O.; Vassylyev, D. G.; Tateno, M.; Shimada, A.; Nakama, T.; Fukai, S.; Konno, M.; Hendrickson, T. L.; Schimmel, P.; Yokoyama, S., *Science* **1998**, *280* (5363), 578-582.
56. Fukai, S.; Nureki, O.; Sekine, S.; Shimada, A.; Tao, J. S.; Vassylyev, D. G.; Yokoyama, S., *Cell* **2000**, *103* (5), 793-803.
57. Fukunaga, R.; Yokoyama, S., *Journal of Biological Chemistry* **2005**, *280* (33), 29937-29945.
58. Tukalo, M.; Yaremchuk, A.; Fukunaga, R.; Yokoyama, S.; Cusack, S., *Nature Structural & Molecular Biology* **2005**, *12* (10), 923-930.
59. Derewenda, Z. S.; Lee, L.; Derewenda, U., *Journal of Molecular Biology* **1995**, *252* (2), 248-262.
60. Desiraju, G. R., *Accounts of Chemical Research* **1996**, *29* (9), 441-449.
61. Senes, A.; Ubarretxena-Belandia, I.; Engelman, D. M., *PNAS* **2001**, *98* (16), 9056-9061.
62. Wells, T. N. C.; Fersht, A. R., *Nature* **1985**, *316* (6029), 656-657.

63. Francklyn, C. S.; First, E. A.; Perona, J. J.; Hou, Y. M., *Methods* **2008**, *44* (2), 100-118.
64. Cleland, W. W., *Biochimica Et Biophysica Acta* **1963**, *67* (2), 188-196.
65. Fersht, A., *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding.* W.H. Freeman: New York, 1998.
66. Dusha, I.; Denes, G., *Analytical Biochemistry* **1977**, *81* (1), 247-250.
67. Fersht, A. R.; Ashford, J. S.; Bruton, C. J.; Jakes, R.; Koch, G. L. E.; Hartley, B. S., *Biochemistry* **1975**, *14* (1), 1-4.
68. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S., *Nature Structural Biology* **2003**, *10* (6), 425-432.
69. Praetorius-Ibba, M.; Stange-Thomann, N.; Kitabatake, M.; Ali, K.; Soll, I.; Carter, C. W.; Ibba, M.; Soll, D., *Biochemistry* **2000**, *39* (43), 13136-13143.
70. Xu, F.; Jia, J.; Jin, Y. X.; Wang, D. T. P., *Protein Expression and Purification* **2001**, *23* (2), 296-300.
71. Hogue, C. W. V.; Rasquinha, I.; Szabo, A. G.; Macmanus, J. P., *Febs Letters* **1992**, *310* (3), 269-272.
72. Loftfield, R. B., *Biochemical Journal* **1963**, *89* (1), 82-92.
73. Fersht, A. R., *Biochemistry* **1977**, *16* (5), 1025-1030.
74. Buurma, N. J.; Haq, I., *Methods* **2007**, *42* (2), 162-172.
75. Zhou, Q. S.; Kapoor, M.; Guo, M.; Belani, R.; Xu, X. L.; Kiosses, W. B.; Hanan, M.; Park, C.; Armour, E.; Do, M. H.; Nangle, L. A.; Schimmel, P.; Yang, X. L., *Nature Structural & Molecular Biology* **2010**, *17* (1), 57-61.
76. Zhang, Z. W.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G., *PNAS* **2004**, *101* (24), 8882-8887.
77. Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G., *Nature* **1985**, *314* (6008), 235-238.
78. Fersht, A. R., *Trends in Biochemical Sciences* **1987**, *12* (8), 301-304.
79. Alber, T.; Matthews, B. W., *THE USE OF X-RAY CRYSTALLOGRAPHY TO DETERMINE THE RELATIONSHIP BETWEEN THE STRUCTURE AND STABILITY OF MUTANTS OF PHAGE T4 LYSOZYME.* 1987; p 289-298.
80. Brown, K. A.; Brick, P.; Blow, D. M., *Nature* **1987**, *326* (6111), 416-418.
81. Villafranca, J. E.; Howell, E. E.; Oatley, S. J.; Xuong, N. H.; Kraut, J., *Biochemistry* **1987**, *26* (8), 2182-2189.
82. Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R., *Cell* **1984**, *38* (3), 835-840.
83. Wilkinson, A. J.; Fersht, A. R.; Blow, D. M.; Winter, G., *Biochemistry* **1983**, *22* (15), 3581-3586.
84. Winter, G.; Fersht, A. R.; Wilkinson, A. J.; Zoller, M.; Smith, M., *Nature* **1982**, *299* (5885), 756-758.
85. Brick, P.; Bhat, T. N.; Blow, D. M., *Journal of Molecular Biology* **1989**, *208* (1), 83-98.
86. Fersht, A. R.; Leatherbarrow, R. J.; Wells, T. N. C., *Biochemistry* **1987**, *26* (19), 6030-6038.
87. Fersht, A. R.; Leatherbarrow, R. J.; Wells, T. N. C., *Trends in Biochemical Sciences* **1986**, *11* (8), 321-325.
88. Fersht, A. R.; Leatherbarrow, R. J.; Wells, T. N. C., *Nature* **1986**, *322* (6076), 284-286.
89. Ho, C. K.; Fersht, A. R., *Biochemistry* **1986**, *25* (8), 1891-1897.
90. Leatherbarrow, R. J.; Fersht, A. R.; Winter, G., *PNAS* **1985**, *82* (23), 7840-7844.
91. Leatherbarrow, R. J.; Fersht, A. R., *Biochemistry* **1987**, *26* (26), 8524-8528.
92. Bedouelle, H.; Winter, G., *Nature* **1986**, *320* (6060), 371-373.
93. Landes, C.; Perona, J. J.; Brunie, S.; Rould, M. A.; Zelwer, C.; Steitz, T. A.; Risler, J. L., *Biochimie* **1995**, *77* (3), 194-203.
94. Yang, X. L.; Otero, F. J.; Skene, R. J.; McRee, D. E.; Schimmel, P.; de Pouplana, L. R., *PNAS* **2003**, *100* (26), 15376-15380.

95. Doublie, S.; Bricogne, G.; Gilmore, C.; Carter, C. W., *Structure* **1995**, *3* (1), 17-31.
96. Brunie, S.; Zelwer, C.; Risler, J. L., *Journal of Molecular Biology* **1990**, *216* (2), 411-424.
97. Perona, J. J.; Rould, M. A.; Steitz, T. A.; Risler, J. L.; Zelwer, C.; Brunie, S., *PNAS* **1991**, *88* (7), 2903-2907.
98. Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L., *Protein Science* **2001**, *10* (7), 1281-1292.
99. Jia, J.; Xu, F.; Chen, X. L.; Chen, L.; Jin, Y. X.; Wang, D. T. P., *Biochemical Journal* **2002**, *365*, 749-756.
100. Ilyin, V. A.; Temple, B.; Hu, M.; Li, G. P.; Yin, Y. H.; Vachette, P.; Carter, C. W., *Protein Science* **2000**, *9* (2), 218-231.
101. Burke, S. A.; Lo, S. L.; Krzycki, J. A., *Journal of Bacteriology* **1998**, *180* (13), 3432-3440.
102. Polycarpo, C.; Ambrogelly, A.; Berube, A.; Winbush, S. A. M.; McCloskey, J. A.; Crain, P. F.; Wood, J. L.; Soll, D., *PNAS* **2004**, *101* (34), 12450-12454.
103. Herring, S.; Ambrogelly, A.; Gundllapalli, S.; O'Donoghue, P.; Polycarpo, C. R.; Soll, D., *Febs Letters* **2007**, *581* (17), 3197-3203.
104. Tuite, M. F.; Santos, M. A. S., *Biochimie* **1996**, *78* (11-12), 993-999.
105. Wang, L.; Xie, J.; Schultz, P. G., *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35*, 225-249.
106. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., *Science* **2001**, *292* (5516), 498-500.
107. Wang, L.; Schultz, P. G., *Angewandte Chemie-International Edition* **2005**, *44* (1), 34-66.
108. Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P., *Annual Review of Biochemistry* **2004**, *73*, 147-176.
109. Li, C. H.; Yamashiro, D.; Gospodarowicz, D., *PNAS* **1983**, *80* (8 I), 2216-2220.
110. Johnson, A. E.; Woodward, W. R.; Herbert, E.; Menninger, J. R., *Biochemistry* **1976**, *15* (3), 569-575.
111. Saks, M. E.; Sampson, J. R.; Nowak, M. W.; Kearney, P. C.; Du, F. Y.; Abelson, J. N.; Lester, H. A.; Dougherty, D. A., *Journal of Biological Chemistry* **1996**, *271* (38), 23169-23175.
112. Tze Fei Wong, J., *PNAS* **1983**, *80* (20 I), 6303-6306.
113. Coin, I.; Beyermann, M.; Bienert, M., *Nature Protocols* **2007**, *2* (12), 3247-3256.
114. Amblard, M.; Fehrentz, J. A.; Martinez, J.; Subra, G., *Molecular Biotechnology* **2006**, *33* (3), 239-254.
115. Kent, S. B. H., *Annual Review of Biochemistry* **1988**, *57*, 957-989.
116. Dawson, P. E.; Kent, S. B. H., *Annual Review of Biochemistry* **2000**, *69*, 923-960.
117. Muir, T. W.; Sondhi, D.; Cole, P. A., *PNAS* **1998**, *95* (12), 6705-6710.
118. Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S., *Journal of Biological Chemistry* **1978**, *253* (13), 4517-4520.
119. Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S., *Journal of the American Chemical Society* **1989**, *111* (20), 8013-8014.
120. Beene, D. L.; Dougherty, D. A.; Lester, H. A., *Current Opinion in Neurobiology* **2003**, *13* (3), 264-270.
121. Hendrickson, W. A.; Horton, J. R.; Lemaster, D. M., *Embo Journal* **1990**, *9* (5), 1665-1672.
122. Hortin, G.; Boime, I., *Methods in Enzymology* **1983**, *96*, 777-784.
123. Ibba, M.; Hennecke, H., *Febs Letters* **1995**, *364* (3), 272-275.
124. Tang, Y.; Tirrell, D. A., *Biochemistry* **2002**, *41* (34), 10635-10645.
125. Young, T. S.; Schultz, P. G., *Journal of Biological Chemistry* **2010**, *285* (15), 11039-11044.
126. Boles, J. O.; Henderson, J.; Hatch, D.; Silks, L. A., *Biochemical and Biophysical Research Communications* **2002**, *298* (2), 257-261.
127. Ross, J. B. A.; Senear, D. F.; Waxman, E.; Kombo, B. B.; Rusinova, E.; Huang, Y. T.; Laws, W. R.; Hasselbacher, C. A., *PNAS* **1992**, *89* (24), 12023-12027.

132

128. Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G., *Nature Biotechnology* **2004,** *22* (10), 1297-1301.

129. Carlson, J. E.; Steward, L. E.; Rusinova, R.; Chamberlin, A. R.; Ross, J. B. A., *Biophysical Journal* **1996,** *70* (2 PART 2), A210.

130. Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G., *PNAS* **2003,** *100* (1), 56-61.

131. Zhang, Z. W.; Wang, L.; Brock, A.; Schultz, P. G., *Angewandte Chemie-International Edition* **2002,** *41* (15), 2840-2842.

132. Zhang, Z. W.; Smith, B. A. C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G., *Biochemistry* **2003,** *42* (22), 6735-6746.

133. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G., *PNAS* **2002,** *99* (17), 11020-11024.

134. Wang, L.; Schultz, P. G., *Chemical Communications* **2002,** (1), 1-11.

135. Liu, D. R.; Magliery, T. J.; Pasternak, M.; Schultz, P. G., *PNAS* **1997,** *94* (19), 10092-10097.

136. Furter, R., *Protein Science* **1998,** *7* (2), 419-426.

137. Steer, B. A.; Schimmel, P., *Journal of Biological Chemistry* **1999,** *274* (50), 35601-35606.

138. Himeno, H.; Hasegawa, T.; Ueda, T.; Watanabe, K.; Shimizu, M., *Nucleic Acids Research* **1990,** *18* (23), 6815-6819.

139. Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G., *Journal of the American Chemical Society* **2000,** *122* (20), 5010-5011.

140. Pastrnak, M.; Magliery, T. J.; Schultz, P. G., *Helvetica Chimica Acta* **2000,** *83* (9), 2277-2286.

141. Stemmer, W. P. C., *Nature* **1994,** *370* (6488), 389-391.

142. Wang, L.; Brock, A.; Schultz, P. G., *Journal of the American Chemical Society* **2002,** *124* (9), 1836-1837.

143. Liu, C. C.; Schultz, P. G., Adding New Chemistries to the Genetic Code. In *Annual Review of Biochemistry, Vol 79*, Annual Reviews: Palo Alto, 2010; Vol. 79, pp 413-444.

144. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G., *Science* **2003,** *301* (5635), 964-967.

145. Young, T. S.; Ahmad, I.; Brock, A.; Schultz, P. G., *Biochemistry* **2009,** *48* (12), 2643-2653.

146. Liu, W. S.; Brock, A.; Chen, S.; Chen, S. B.; Schultz, P. G., *Nature Methods* **2007,** *4* (3), 239-244.

147. Zeng, H. Q.; Me, J. M.; Schultz, P. G., *Bioorganic & Medicinal Chemistry Letters* **2006,** *16* (20), 5356-5359.

148. Wang, J.; Schiller, S. M.; Schultz, P. G., *Angewandte Chemie-International Edition* **2007,** *46* (36), 6849-6851.

149. Deiters, A.; Schultz, P. G., *Bioorganic & Medicinal Chemistry Letters* **2005,** *15* (5), 1521-1524.

150. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G., *Journal of the American Chemical Society* **2003,** *125* (39), 11782-11783.

151. Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G., *Journal of the American Chemical Society* **2002,** *124* (31), 9026-9027.

152. Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G., *Angewandte Chemie-International Edition* **2008,** *47* (43), 8220-8223.

153. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., *Journal of the American Chemical Society* **2004,** *126* (44), 14306-14307.

154. Wang, W. Y.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L., *Nature Neuroscience* **2007,** *10* (8), 1063-1072.

155. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G., *Nature Biotechnology* **2002,** *20* (10), 1044-1048.

156. Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G., *Journal of the American Chemical Society* **2003,** *125* (4), 935-939.

157. Schultz, K. C.; Supekova, L.; Ryu, Y. H.; Xie, J. M.; Perera, R.; Schultz, P. G., *Journal of the American Chemical Society* **2006**, *128* (43), 13984-13985.

158. Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G., *Journal of Organic Chemistry* **2003**, *68* (1), 174-176.

159. Turner, J. M.; Graziano, J.; Spraggon, G.; Schultz, P. G., *PNAS* **2006**, *103* (17), 6483-6488.

160. Tsao, M. L.; Summerer, D.; Ryu, Y. H.; Schultz, P. G., *Journal of the American Chemical Society* **2006**, *128* (14), 4572-4573.

161. Alfonta, L.; Zhang, Z. W.; Uryu, S.; Loo, J. A.; Schultz, P. G., *Journal of the American Chemical Society* **2003**, *125* (48), 14662-14663.

162. Seyedsayamdost, M. R.; Xie, J.; Chan, C. T. Y.; Schultz, P. G.; Stubbe, J., *Journal of the American Chemical Society* **2007**, *129* (48), 15060-15071.

163. Sakamoto, K.; Murayama, K.; Oki, K.; Lraha, F.; Kato-Murayama, M.; Takahashi, M.; Ohtake, K.; Kobayashi, T.; Kuramitsu, S.; Shirouzu, M.; Yokoyama, S., *Structure* **2009**, *17* (3), 335-344.

164. Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S., *Nucleic Acids Research* **2002**, *30* (21), 4692-4699.

165. Xie, J. M.; Liu, W. S.; Schultz, P. G., *Angewandte Chemie-International Edition* **2007**, *46* (48), 9239-9242.

166. Guo, J. T.; Wang, J. Y.; Anderson, J. C.; Schultz, P. G., *Angewandte Chemie-International Edition* **2008**, *47* (4), 722-725.

167. Lee, H. S.; Spraggon, G.; Schultz, P. G.; Wang, F., *Journal of the American Chemical Society* **2009**, *131* (7), 2481-2483.

168. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S., *Nature Methods* **2005**, *2* (3), 201-206.

169. Deiters, A.; Groff, D.; Ryu, Y. H.; Xie, J. M.; Schultz, P. G., *Angewandte Chemie-International Edition* **2006**, *45* (17), 2728-2731.

170. Peters, F. B.; Brock, A.; Wang, J. Y.; Schultz, P. G., *Chemistry & Biology* **2009**, *16* (2), 148-152.

171. Bose, M.; Groff, D.; Xie, J. M.; Brustad, E.; Schultz, P. G., *Journal of the American Chemical Society* **2006**, *128* (2), 388-389.

172. Wang, J. Y.; Xie, J. M.; Schultz, P. G., *Journal of the American Chemical Society* **2006**, *128* (27), 8738-8739.

173. Xie, J. M.; Supekova, L.; Schultz, P. G., *Acs Chemical Biology* **2007**, *2* (7), 474-478.

174. Neumann, H.; Hazen, J. L.; Weinstein, J.; Mehl, R. A.; Chin, J. W., *Journal of the American Chemical Society* **2008**, *130* (12), 4028-4033.

175. Liu, C. C.; Schultz, P. G., *Nature Biotechnology* **2006**, *24* (11), 1436-1440.

176. Splan, K. E.; Musier-Forsyth, K.; Boniecki, M. T.; Martinis, S. A., *Methods* **2008**, *44* (2), 119-128.

177. Hutchison, C. A.; Phillips, S.; Edgell, M. H.; Gillam, S.; Jahnke, P.; Smith, M., *Journal of Biological Chemistry* **1978**, *253* (18), 6551-6560.

178. Turner, J. M.; Graziano, J.; Spraggon, G.; Schultz, P. G., *Journal of the American Chemical Society* **2005**, *127* (43), 14976-14977.

179. Phillips, R. S.; Cohen, L. A.; Annby, U.; Wensbo, D.; Gronowitz, S., *Bioorganic & Medicinal Chemistry Letters* **1995**, *5* (11), 1133-1134.

180. Hudson, B. S.; Harris, D. L.; Ludescher, R. D.; Ruggiero, A.; Cooney-Freed, A.; Cavalier, S. A., *FLUORESCENCE PROBE STUDIES OF PROTEINS AND MEMBRANES.* **1986**; p 159-202.

181. Vivian, J. T.; Callis, P. R., *Biophysical Journal* **2001**, *80* (5), 2093-2109.

182. Ross, J. B. A.; Szabo, A. G.; Hogue, C. W. V., Enhancement of protein spectra with tryptophan analogs: Fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. In *Fluorescence Spectroscopy*, Academic Press Inc: San Diego, 1997; Vol. 278, pp 151-190.

183. Broos, J.; Gabellieri, E.; Biemans-Oldehinkel, E.; Strambini, G. B., *Protein Science* **2003**, *12* (9), 1991-2000.

184. Barlati, S.; Ciferri, O., *Journal of Bacteriology* **1970**, *101* (1), 166-172.

185. Pardee, A. B.; Shore, V. G.; Prestidge, L. S., *Biochimica Et Biophysica Acta* **1956**, *21* (2), 406-407.

186. Pratt, E. A.; Ho, C., *Biochemistry* **1975**, *14* (13), 3035-3040.

187. Bronskill, P. M.; Wong, J. T. F., *Biochemical Journal* **1988**, *249* (1), 305-308.

188. Brawerman, G.; Ycas, M., *Archives of Biochemistry and Biophysics* **1957**, *68* (1), 112-117.

189. Sykes, B. D.; Weingart.Hi; Schlesin.Mj, *PNAS* **1974**, *71* (2), 469-473.

190. Pratt, E. A.; Ho, C., *Federation Proceedings* **1974**, *33* (5), 1463-1463.

191. Minks, C.; Huber, R.; Moroder, L.; Budisa, N., *Biochemistry* **1999**, *38* (33), 10649-10659.

192. Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N., *Journal of Molecular Biology* **2001**, *309* (4), 925-936.

193. Sharon, N.; Lipmann, F., *Arch Biochem and Biophys* **1957**, *69*, 219-227.

194. Schlesin.S, *Journal of Biological Chemistry* **1968**, *243* (14), 3877-3883.

195. Hogue, C. W. V.; Szabo, A. G., *Biophysical Chemistry* **1993**, *48* (2), 159-169.

196. Laue, T. M.; Senear, D. F.; Eaton, S.; Ross, J. B. A., *Biochemistry* **1993**, *32* (10), 2469-2472.

197. Senear, D. F.; Laue, T. M.; Ross, J. B. A.; Waxman, E.; Eaton, S.; Rusinova, E., *Biochemistry* **1993**, *32* (24), 6179-6189.

198. James, N. G.; Byrne, S. L.; Mason, A. B., *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2009**, *1794* (3), 532-540.

199. Hughes, R. A.; Ellington, A. D., *Nucleic Acids Research* **2010**, *38* (19), 6813-6830.

200. Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L., **2002**, *41* (21), 4066-4069.

201. Kiga, D.; Sakamoto, K.; Kodama, K.; Kigawa, T.; Matsuda, T.; Yabuki, T.; Shirouzu, M.; Harada, Y.; Nakayama, H.; Takio, K.; Hasegawa, Y.; Endo, Y.; Hirao, I.; Yokoyama, S., *PNAS* **2002**, *99* (15), 9715-9720.

202. Guo, Q.; Gong, Q. G.; Tong, K. L.; Vestergaard, B.; Costa, A.; Desgres, J.; Wong, M. S.; Grosjean, H.; Zhu, G.; Wong, J. T. F.; Xue, H., *Journal of Biological Chemistry* **2002**, *277* (16), 14343-14349.

203. Buddha, M. R.; Crane, B. R., *Nature Structural & Molecular Biology* **2005**, *12* (3), 274-275.

204. Antonczak, A. K.; Simova, Z.; Yonemoto, I. T.; Bochtler, M.; Piasecka, A.; Czapinska, H.; Brancale, A.; Tippmann, E. M., *PNAS* **2011**, *108* (4), 1320-1325.

205. Xie, J. M.; Schultz, P. G., *Nature Reviews Molecular Cell Biology* **2006**, *7* (10), 775-782.

206. Kwon, I.; Tirrell, D. A., *Journal of the American Chemical Society* **2007**, *129* (34), 10431-10437.

207. Laowanapiban, P.; Kapustina, M.; Vonrhein, C.; Delarue, M.; Koehl, P.; Carter, C. W., *PNAS* **2009**, *106* (6), 1790-1795.

208. Ito, K.; Hiraga, S.; Yura, T., *Genetics* **1969**, *61* (3), 521-538.

209. Ito, K., *Molecular and General Genetics* **1972**, *115* (4), p 349.

210. Falconer, R. J.; Penkova, A.; Jelesarov, I.; Collins, B. M., *Journal of Molecular Recognition* **2010**, *23* (5), 395-413.

211. Todd, M. J.; Gomez, J., *Analytical Biochemistry* **2001**, *296* (2), 179-187.

212. Bianconi, M. L., *Biophysical Chemistry* **2007**, *126* (1-3), 59-64.

213. Egawa, T.; Tsuneshige, A.; Suematsu, M.; Yonetani, T., *Analytical Chemistry* **2007**, *79* (7), 2972-2978.

214. Yang, X. L.; Guo, M.; Kapoor, M.; Ewalt, K. L.; Otero, F. J.; Skene, R. J.; McRee, D. E.; Schimmel, P., *Structure* **2007**, *15* (7), 793-805.

215. Kapustina, M.; Weinreb, V.; Li, L.; Kuhlman, B.; Carter Jr, C. W., *Structure* **2007**, *15* (10), 1272-1284.

216. Tellinghuisen, J., *Analytical Biochemistry* **2008**, *373* (2), 395-397.

217. Turnbull, W. B.; Daranas, A. H., *Journal of the American Chemical Society* **2003**, *125* (48), 14859-14866.

218. Brandts, J. F.; Lin, L. N.; Wiseman, T.; Williston, S.; Yang, C. P., *American Laboratory* **1990**, *22* (8), p 30.

219. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N., *Analytical Biochemistry* **1989**, *179* (1), 131-137.

220. Zhang, Y. L.; Zhang, Z. Y., *Analytical Biochemistry* **1998**, *261* (2), 139-148.

221. Buurma, N. J.; Haq, I., *Journal of Molecular Biology* **2008**, *381* (3), 607-621.

222. Xu, Z. J.; Love, M. L.; Ma, L. Y. Y.; Blum, M.; Bronskill, P. M.; Bernstein, J.; Grey, A. A.; Hofmann, T.; Camerman, N.; Wong, J. T. F., *Journal of Biological Chemistry* **1989**, *264* (8), 4304-4311.

137

Appendices

Following are sequences of all three used constructs. CCG or GTC nucleotides encoding Pro[144] or Val[144] are indicated bold and underlined. Underlined is also a sequence corresponding to an extra tag of TrpRS Pro[144] + tag in Sequence III.

Sequence I: *Bs* TrpRS Val[144]Pro

ATGAAACAGACCATTTTCAGCGGCATCCAGCCATCCGGCAGCGTCACACTGGGTAATTATA

TTGGCGCTATGAAACAATTCGTTGAGCTGCAGCATGATTATAACTCCTATTTTTGCATTGTG

GATCAGCACGCGATTACGGTGCCACAGGATCGTCTGGAACTGCGTAAAAACATCCGGAATC

TGGCAGCACTGTACCTGGCAGTGGGTCTGGACCCAGAAAAAGCGACCCTGTTTATCCAGTC

AGAGGTACCGGCCCATGCCCAAGCAGGTTGGATGATGCAATGTGTGGCATACATTGGTGAG

CTGGAACGCATGACTCAGTTCAAAGATAAAAGCAAAGGCAACGAAGCAGTCGTATCAGGC

CTGTTAACATATCCGCCGCTGATGGCTGCGGATATCCTGCTGTACGGGACGGATCTGGTCCC

G<u>**CCG**</u>GGTGAAGACCAGAAACAACACCTGGAATTAACGCGTAATCTGGCTGAACGCTTTAAC

AAGAAGTACAATGATATTTTTACCATCCCGGAAGTTAAAATTCCGAAGGTTGGGGCGCGCA

TCATGAGTTTAAACGATCCGTTAAAGAAAATGTCCAAATCTGACCCGAATCAGAAAGCGTA

CATTACCCTGCTGGACGAACCTAAACAACTGGAAAGAAGATCAAATCTGCGGTGACTGAC

TCTGAAGGTATCGTAAAATTTGACAAAGAAAATAAGCCTGGCGTTAGTAATCTGCTGACGA

TTTACAGTATCTTAGGCAACACTACCATCGAAGAACTGGAGGCGAAGTATGAGGGGAAAGG

CTATGGCGAGTTCAAAGGGGACCTGGCGGAAGTGGTTGTTAACGCTCTGAAACCTATTCAG

GATCGTTATTATGAGCTGATTGAGAGCGAAGAACTGGATCGTATTTTAGATGAGGGTGCCG

AACGTGCCAATCGGACCGCCAACAAAATGTTAAAAAAAATGGAAAACGCCATGGGTTTAG

GTCGCAAACGCCGCTAA

Sequence II: wt *Bs* TrpRS

ATGAAACAGACCATTTTCAGCGGCATCCAGCCATCCGGCAGCGTCACACTGGGTAATTATA

TTGGCGCTATGAAACAATTCGTTGAGCTGCAGCATGATTATAACTCCTATTTTTGCATTGTG

GATCAGCACGCGATTACGGTGCCACAGGATCGTCTGGAACTGCGTAAAAACATCCGGAATC

TGGCAGCACTGTACCTGGCAGTGGGTCTGGACCCAGAAAAAGCGACCCTGTTTATCCAGTC

AGAGGTACCGGCCCATGCCCAAGCAGGTTGGATGATGCAATGTGTGGCATACATTGGTGAG

CTGGAACGCATGACTCAGTTCAAAGATAAAAGCAAAGGCAACGAAGCAGTCGTATCAGGC

CTGTTAACATATCCGCCGCTGATGGCTGCGGATATCCTGCTGTACGGGACGGATCTGGTCCC

G**GTC**GGTGAAGACCAGAAACAACACCTGGAATTAACGCGTAATCTGGCTGAACGCTTTAAC

AAGAAGTACAATGATATTTTTACCATCCCGGAAGTTAAAATTCCGAAGGTTGGGGCGCGCA

TCATGAGTTTAAACGATCCGTTAAAGAAAATGTCCAAATCTGACCCGAATCAGAAAGCGTA

CATTACCCTGCTGGACGAACCTAAACAACTGGAAAAGAAGATCAAATCTGCGGTGACTGAC

TCTGAAGGTATCGTAAAATTTGACAAAGAAAATAAGCCTGGCGTTAGTAATCTGCTGACGA

TTTACAGTATCTTAGGCAACACTACCATCGAAGAACTGGAGGCGAAGTATGAGGGGAAAGG

CTATGGCGAGTTCAAAGGGGACCTGGCGGAAGTGGTTGTTAACGCTCTGAAACCTATTCAG

GATCGTTATTATGAGCTGATTGAGAGCGAAGAACTGGATCGTATTTTAGATGAGGGTGCCG

AACGTGCCAATCGGACCGCCAACAAAATGTTAAAAAAAATGGAAAACGCCATGGGTTTAG

GTCGCAAACGCCGCTAA

Sequence III: *Bs* TrpRS Val¹⁴⁴Pro + tag

ATGAAACAGACCATTTTCAGCGGCATCCAGCCATCCGGCAGCGTCACACTGGGTAATTATA

TTGGCGCTATGAAACAATTCGTTGAGCTGCAGCATGATTATAACTCCTATTTTTGCATTGTG

GATCAGCACGCGATTACGGTGCCACAGGATCGTCTGGAACTGCGTAAAAACATCCGGAATC

TGGCAGCACTGTACCTGGCAGTGGGTCTGGACCCAGAAAAAGCGACCCTGTTTATCCAGTC

AGAGGTACCGGCCCATGCCCAAGCAGGTTGGATGATGCAATGTGTGGCATACATTGGTGAG

CTGGAACGCATGACTCAGTTCAAAGATAAAAGCAAAGGCAACGAAGCAGTCGTATCAGGC

CTGTTAACATATCCGCCGCTGATGGCTGCGGATATCCTGCTGTACGGGACGGATCTGGTCCC

G<u>CCG</u>GGTGAAGACCAGAAACAACACCTGGAATTAACGCGTAATCTGGCTGAACGCTTTAAC

AAGAAGTACAATGATATTTTTACCATCCCGGAAGTTAAAATTCCGAAGGTTGGGGCGCGCA

TCATGAGTTTAAACGATCCGTTAAAGAAAATGTCCAAATCTGACCCGAATCAGAAAGCGTA

CATTACCCTGCTGGACGAACCTAAACAACTGGAAAGAAGATCAAATCTGCGGTGACTGAC

TCTGAAGGTATCGTAAAATTTGACAAAGAAAATAAGCCTGGCGTTAGTAATCTGCTGACGA

TTTACAGTATCTTAGGCAACACTACCATCGAAGAACTGGAGGCGAAGTATGAGGGGAAAGG

CTATGGCGAGTTCAAAGGGGACCTGGCGGAAGTGGTTGTTAACGCTCTGAAACCTATTCAG

GATCGTTATTATGAGCTGATTGAGAGCGAAGAACTGGATCGTATTTTAGATGAGGGTGCCG

AACGTGCCAATCGGACCGCCAACAAAATGTTAAAAAAAATGGAAAACGCCATGGGTTTAG

GTCGCAAACGCC<u>AAGGGCGAGCTCAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGA</u>

<u>GTTGGCTGCTGCCACCGCTGAGCAATAACTAG</u>