THE ENGINEERING AND OPTIMIZATION OF PYRROLYSYL-SYNTHETASE AS A TOOL FOR NONCANONICAL AMINO ACID INCORPORATION

A Dissertation

by

VANGMAYEE SHARMA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe Liu
Committee Members,	Frank Raushel
	David Barondeau
	Mary Bryk
Head of Department,	Simon North

August 2017

Major Subject: Chemistry

ABSTRACT

Since the discovery of pyrrolysine as the 22nd amino acid, the field of chemical biology has expanded tremendously with important developments made in genetic noncanonical amino acid (ncAA) incorporation based on the pyrrolysine incorporation machinery. As first discovered in methanogenic archaea, pyrrolysine is incorporated using pyrrolysyl-tRNA and the cognate pyrrolysyl-tRNA synthetase which install this amino acid using an in-frame stop codon present on the mRNA. This archaeal synthetase and tRNA form an orthogonal pair in naïve organisms such as *Escherichia coli*, mammalian cells etc. and have been engineered over the past few years to incorporate several different noncanonical amino acids in proteins genetically. Although this system serves as an important chemical biology tool, it leaves much to be desired in terms of incorporation efficiency of these different unnatural amino acids.

In this dissertation, we have explored the method for optimizing this remarkable system to achieve high levels of ncAA incorporation by engineering its N-terminal domain. Further, we then examined its application in improving other tRNA/synthetase systems which have been created for incorporation of specific ncAAs by researchers in previous years. In addition, we have also probed the feasibility of incorporating several histidine derivatives in proteins in using an engineered C-terminal domain version of this synthetase and the cognate tRNA to be able to study mechanism of enzymes employing histidines in their catalytic activity. In particular, the mechanism of alanine racemase

enzyme was probed to validate the existence of a proton transfer chain near the catalytic site of the enzyme crucial to its activity.

Overall, our results can be translated to other orthogonal tRNA/synthetase pairs derived from pyrrolysyl-tRNA/synthetase to achieve improvement in existing ncAA incorporation systems and hence resolve the poor incorporation problem of these ncAAs. Also, the system developed by us can be used to probe several enzymes which employ histidine triads in their catalytic mechanism. It can also be used to modify or fine-tune the activity of these enzymes without compromising protein structure.

To my maa

Who has always been there through all my happiness and sorrow,

My accomplishments and failures,

And who has been the source of my strength and courage,

I dedicate this work.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Wenshe Liu, and my committee members, Dr. Raushel, Dr. Barondeau, and Dr. Bryk, for their guidance and support throughout the course of my graduate studies and research.

To all the Liu group members, past and present, it has been a great experience working with every one of you. A special thanks to Dr. Yu Zeng for being a wonderful mentor. To Dr. Bo Wu, Dr. Yan-Jiun Lee, Dr. Alfred Tuley, Dr. Xiaoyan Wang and Dr. Keturah Odoi, thanks for being such great peers and for all the advice and support. A special thanks to my friends Willie Hsu and Jeffery Tharp for being amazing colleagues and for creating a fun work environment. Thanks also to Wesley Wang and Zhipeng Wang for their help with experiments, to Xiaoshan Wang for letting me play with her amazing fish tank and to Erol Vatansever for his help with proofreading this work. Thank you to Dr. Prachi Joshi as well for her help in proofreading as well.

Special thanks to all the Chemistry Department staff, especially Sandy Horton and Julie Zercher for all there help and support and Lindsey Williams and Angie Stickley who saved my experiments by their prompt help in ordering all the critical instrument parts. I would also like to thank Dr. Joanna Goody-Pellois who has been great help and support throughout. Thanks also go to all the staff members of the First Year Program especially Veronica Ramirez, Kelley Brown and Travis McCartney and all the lab coordinators for the encouragement and support throughout these years and for making this journey easier

for me. A special thanks to the Dr. Edward Lee, Dr. Amber Schaefer and the late Dr. Tak Leung for their continued encouragement and the opportunity to teach.

The list of people I want to thank is endless for several people were part of this amazing journey, who supported, encouraged and helped me in several ways. To all my friends and family members, a huge shout out and thanks.

Finally, a very huge thank you to my mother and father for their love, patience, support and encouragement throughout this journey. To my sister, thank you for the abundant love, laughter and joy that you bring in our lives.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

Faculty committee recognition

This work was conducted under advisement of Dr. Wenshe Liu of the Department of Chemistry and supervised by a thesis committee consisting of Dr. Frank Raushel, Dr. David Barondeau of the Department of Chemistry and Dr. Mary Bryk of the Department of Biochemistry and Biophysics.

Student/collaborator contributions

The set of 2 primers used for EP-PCR for Chapter II were designed by Dr. Yu Zeng of Department of Chemistry and she also contributed in the first round of screening process. The SIRT1 assay for Chapter II was conducted by Wesley Wang of Department of Chemistry. The synthesis of compound AcdK used in studies for Chapter II was carried out by Dr. Yadagiri Kurra of Department of Chemistry. Yane-Shih Wang contributed in experiments conducted on sfGFP in chapter III. The ESI Mass Spectra and MS/MS data presented in Chapter III were analyzed in part by Dr. Yohannes Rezenom and Dr. Zhongjie Ren of the Department of Chemistry and were published in 2016.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was partially supported by a fellowship from Texas A&M University and partially by the National Institute of Health (Grants R01CA161158 and R01GM121584) and the Welch Foundation (Grant A-1715).

The funding agencies had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

NOMENCLATURE

2YT	Yeast extract tryptone
AcdK	N_ϵ-(4-azidobenzoxycarbonyl)-δ,ϵ-dehydrolysine
AcK	N_ϵ-(*tert*-butoxycarbonyl)-L-lysine
ADP	Adenosine diphosphate
Ala or A	Alanine
Amp	Ampicillin
Arg or R	Arginine
Asn or N	Asparagine
Boc	Butoxycarbonyl
CD	Circular dichroism
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid)
Cm	Chloramphenicol
DIBAL-H	Diisobutylaluminium hydride
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EF	Elongation factor
EPL	Expressed protein ligation

EP-PCR	Error-prone PCR
ESI-MS	Electrospray ionization mass spectrometry
EtOAc	Ethyl acetate
Fmoc	Fluoren-9-ylmethyloxycarbonyl
Glu or E	Glutamic acid
HDAC	Histone deacetylase
His or H	Histidine
IF	Initiation factor
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Lys or K	Lysine
Mh	3-Methyl histidine
mRNA	Messenger ribonucleic acid
MWCO	Molecular weight cut-off
NAD	Nicotinamide adenine dinucleotide
ncAA	Noncanonical amino acid
NCL	Native chemical ligation
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
Phe or F	Phenyl alanine
PLP	Pyridoxal 5'-Phosphate

PTM	Post-translational modification
PylRS	Pyrrolysyl-tRNA synthetase
PylT	Pyrrolysyl-tRNA
RF	Release factor
SAM	S-Adenosyl methionine
SDS	Sodium dodecyl sulphate
sfGFP	Super-folder green fluorescence protein
SPPS	Solid-phase peptide synthesis
SUMO	Small ubiquitin-related modifier
Ta	Thiazolyl alanine
TEV	*Tobacco etch virus*
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
tRNA	Transfer ribonucleic acid
Tyr or Y	Tyrosine
WT	Wild-type

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Protein Biosynthesis

As outlined by the central dogma of microbiology, genetic information gets transferred from DNA to RNA via a process known as transcription and from RNA to protein by a process known as translation.[1] This process of protein biosynthesis in living organisms usually involves 20 naturally occurring (canonical) amino acids and occurs on the ribosome. After the messenger RNA (mRNA) strand is transcribed from DNA, this mRNA is read by the ribosome and decoded based on its triplet nucleotide sequence, also known as codons (Figure I-1). As the first step of the protein synthesis, known as activation, transfer RNAs (tRNAs) are charged with the corresponding amino acid by their cognate aminoacyl-tRNA synthetases. The next step, also known as initiation, involves the smaller ribosomal subunit binding to the 5' end of the mRNA strand with the use of initiation factors (IFs).[2] Following initiation, elongation occurs as the next step in the process where elongation factors (EFs) assist in delivering charged tRNAs to the A-site of the ribosome. Peptidyl transferase activity ensues next, and the growing peptide chain is moved to the P-site of the ribosome along with movement of the ribosome towards the 3' end of the mRNA by the length of a codon. The uncharged tRNA is then released through the E-site of the ribosome. This entire process repeats itself until the ribosome reads one of the stop codons (UAA, UAG and UGA) which results in termination of the protein

synthesis. This final termination step utilizes release factors (RFs) to disassemble the protein translational machinery.

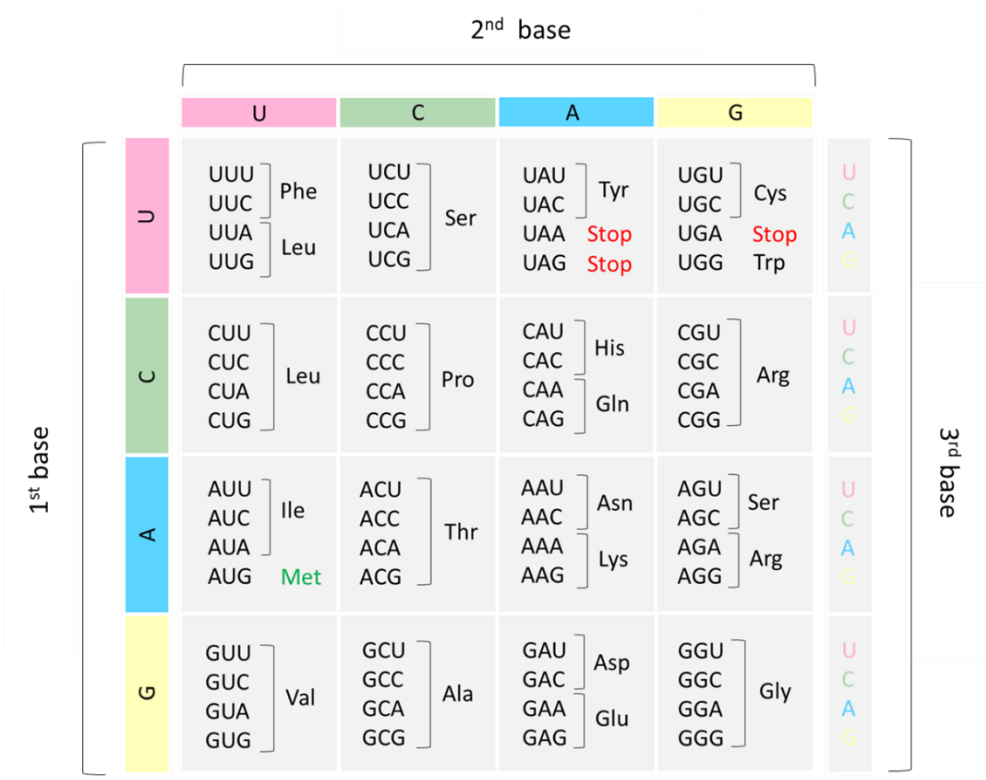


Figure I-1. The standard genetic code

Post-Translational Modifications

After the completion of the translation process, the synthesized proteins undergo various covalent modifications known as post-translational modifications (PTMs) to achieve their complete biological functionality.[3] There are more than 87,000 experimentally identified post-translational modifications so far and they broadly fall in one of the following PTMs, phosphorylation, methylation, glycosylation, acylation, ubiquitylation, citrullination, S-nitrosylation, hydroxylation, etc. (Figure I-2).[4-6]

All these PTMs serve diverse roles in altering the property, structure and function of proteins. Modifications such as phosphorylation, methylation and acetylation tweak the charged state of protein and play important roles in regulation of several cellular processes such as cell growth, intracellular signal transduction, apoptosis, etc.[5, 7-10] Glycosylation on the other hand changes protein's conformation and stability and regulates intercellular signaling pathway.[11] Acylation modifications such as palmitoylation and GlcNAcylation affect interaction of proteins with lipids and membranes and hence control protein trafficking, aggregation and stability.[12, 13] The ubiquitylation modification affixes polyubiquitin chains on proteins thereby marking them for degradation.[14] Apart from these few examples mentioned above, the other PTMs also serve important roles in regulating various cellular functions and misregulation of these PTMs is the cause of various diseases.[15]

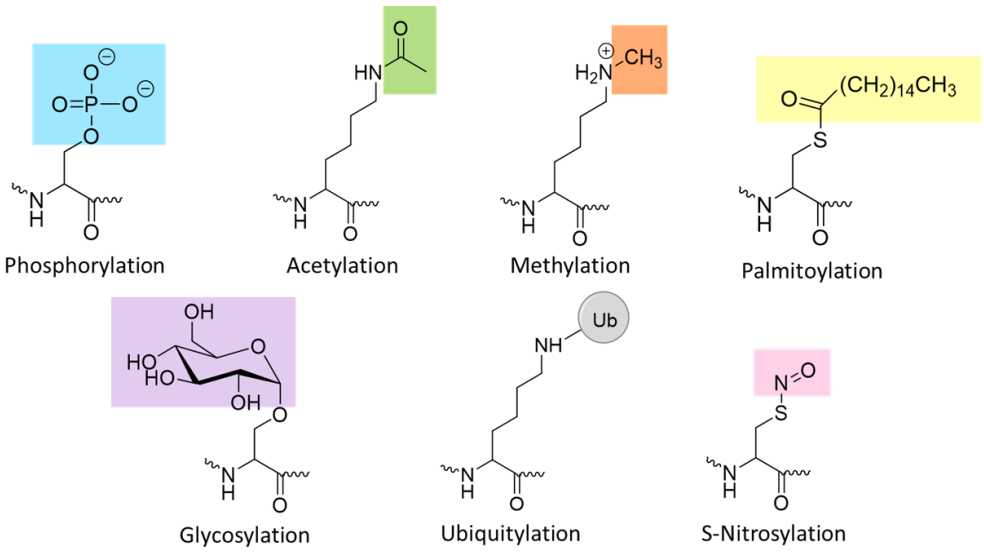


Figure I-2. A few common post-translational modifications.

The important roles played by these post-translationally modified proteins in maintaining normal cellular functioning of living organisms proves that nature requires more than the 20 naturally occurring amino acids to manage proper order for life. These observations resulted in a need to build a suitable method to experimentally install these PTMs in proteins in the form of noncanonical amino acids (ncAAs) to be able to study the structure and functions of these biological molecules in detail.

Histone Post-Translational Modifications

It is especially important to mention post-translational modifications in context of histone proteins. Histone proteins are evolutionarily conserved and are responsible for efficient packaging of genetic material in cells in form of chromatin. Chromatin fiber is a polymer consisting of repeating units of nucleosomes, where each nucleosome is composed of 146 base pairs of DNA wrapped around a histone octamer of two histone H2A, H2B heterodimers and a histone H3, H4 tetramer, linked together by H1 histone protein (Figure I-3).[16, 17]

The N-terminal tails of these histone proteins undergo several covalent post-translational modifications, such as acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ribosylation, etc. as mentioned in the section above, which dynamically regulate each other by a process termed as histone 'cross-talk'.[18] These modifications on histones are responsible for modulating several important cellular processes such as gene transcription, DNA repair, replication, etc. and form epigenetic markers.[18-20] Aberrant cross-talk results in detrimental overexpression of genes or

repression of essential genes and the resulting dysregulation of several of these PTMs have been implicated in diseases such as diabetes, cancer, Alzheimer's disease, etc. as the mechanism regulating the cell repair and expression/repression of various important genes is impaired in these diseases.[21, 22]

Among all these post-translational modifications, histone acetylation has been one of the most well studied PTMs. Several key lysine residues on all four histone proteins (H2A, H2B, H3 and H4) have been identified which undergo acetylation, and the modification is known to regulate transcriptional activation, DNA damage repair and chromatin assembly.[23] Since acetylation leads to neutralization of charge on the protein tail, it is assumed that this leads to loosening of the DNA-protein complex by reducing the electrostatic interactions between DNA and histone and leads to a 'euchromatin' state, the transcriptionally active state of chromatin (Figure I-3). In this form the DNA is accessible to transcriptional machinery, hence acetylation of N-terminal tail of histone regulates chromatin remodeling and controls gene transcription. Conversely, histone methylation results in a 'heterochromatin' state in which transcriptional machinery is inactive due to condensed chromatin form (Figure I-3). Methylation usually occurs on arginine & lysine residues of the histone tail. Methylation, along with acetylation, is known to regulate expression of various genes, overexpression of several of which results in a diseased cell state. Thus, all these PTMs form a transcription regulation code, often referred to as the 'histone code'.[24]

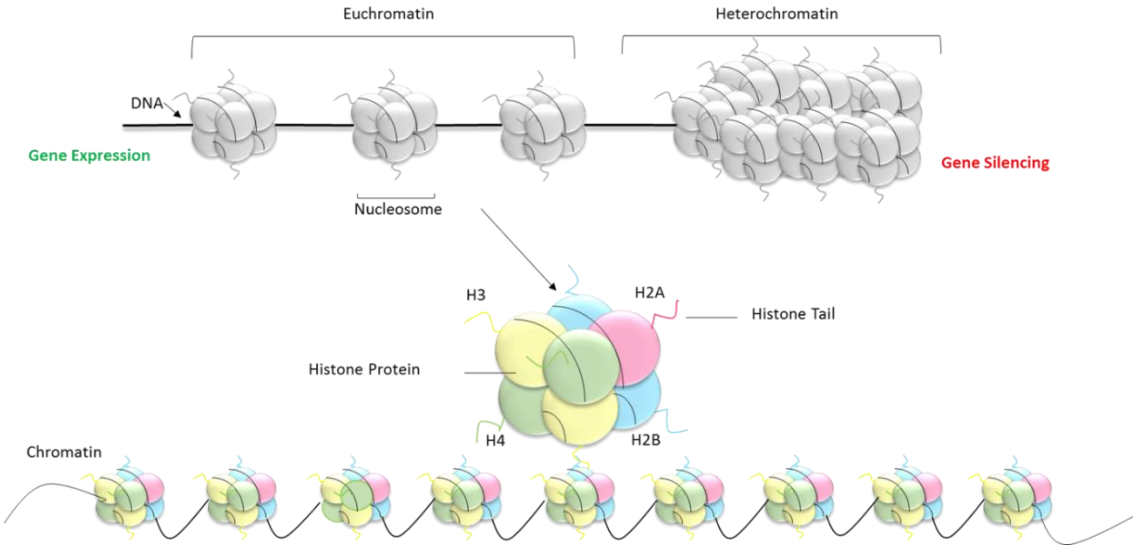


Figure I-3. Regulation of chromatin structure and gene transcription.

Other histone PTMs such as phosphorylation, ADP-ribosylation, ubiquitylation and SUMOylation along with acetylation and methylation form intricate cross-talk maps. So far, studies have shown histone H3S10 phosphorylation, which is a transcription activating mark, affects acetylation on the same histone tail (*cis* crosstalk) at histone H3K9 & H3K14, inhibiting one and promoting the other.[25] More recently, phosphorylation at H3S10 has been shown to trigger H4K16 acetylation (*trans* crosstalk) leading to transcription activation.[26] These and several other discovered cross-talks[27-30] (Figure I-4) have piqued interest of researchers all over the world to explore this field further due to their significant biological implications.

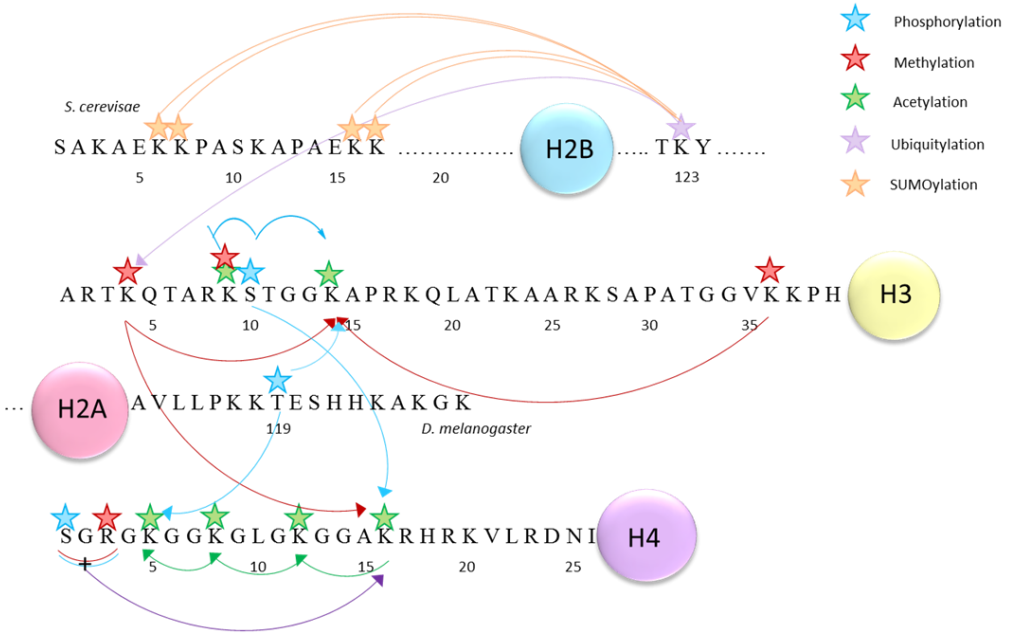


Figure I-4. Map showing some of the cross-talk interactions between different histones.

The development of efficient methodology to detect & map these cross-talks on a single nucleosome, and the importance of knowing the cross regulatory mechanism of these modifications has motivated us in trying to resolve several unanswered questions related to cross-talks. The following section discusses the known methods for obtaining post-translationally modified nucleosomes and proteins in general.

Noncanonical Amino Acid Incorporation

As mentioned above, suitable methodology is required for incorporation of ncAAs in proteins to be able to manipulate their chemical properties such as stability, structure and function and to be able to reconstitute nucleosomes with PTMs. The earliest methods devised involved chemical approaches to obtain these results. Techniques such as solid-phase peptide synthesis (SPPS)[31], native chemical ligation (NCL)[32, 33] and expressed protein ligation (EPL)[34] were used to chemically install non-native amino acids in peptides and proteins synthesized *in vitro*.

Synthetic Techniques

Pioneered by Bruce Merrifield, solid phase peptide synthesis (SPPS) technique was developed in 1963.[35] The general scheme of SPPS involves using a polystyrene resin as solid support which is modified by a linker to which the C-terminal of the first amino acid can be attached by use of activators (Figure I-5).[36] The amino acid so attached is protected on the N-terminal group by tert-butoxycarbonyl (Boc) group which is sensitive to trifluoroacetic acid (TFA) deprotection or fluoren-9-ylmethyloxycarbonyl (Fmoc)

group which is sensitive to deprotection in basic conditions.[37] Followed by peptide coupling and deprotection, this strategy can yield peptides and proteins less than 100 residues in length and has been one of the earliest methods used for incorporation of ncAAs. The main disadvantages of this technique lie in the limitation posed on the length of the peptide that can be synthesized, chemical groups that can be introduced in the form of non-canonical amino acid and cost ineffectiveness.[38]

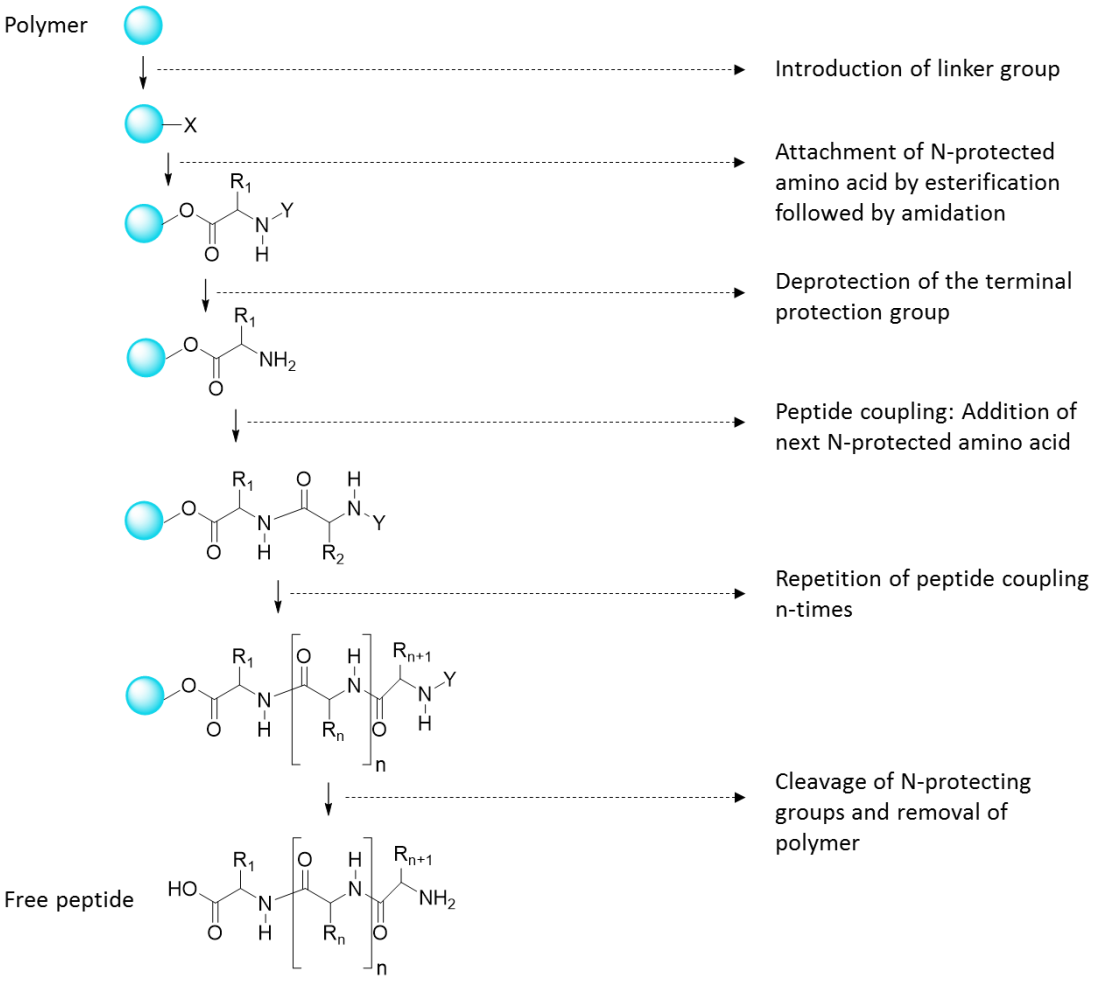


Figure I-5. General scheme of solid phase peptide synthesis.

To overcome the constraint on the length of the peptide that can be synthesized, a technique called as native chemical ligation (NCL) was devised in 1994.[39] This highly chemoselective strategy requires an N-terminal cysteine residue present on a peptide which spontaneously reacts with another peptide with α-thioester on its C-terminal, when mixed in mild denaturing conditions at neutral pH by formation of thioester linkage which rearranges in an intramolecular fashion to yield native peptide bond (Figure I-6). This reaction does not require protection of side chains or additional cysteine groups as only the N-terminal cysteine residues take part in such a reaction.[40]

This method has been used for synthesis of peptides containing ncAA longer than 100 residues in length. It is especially appealing to utilize this technique when synthesizing peptides incorporating ncAAs that can be toxic to cells.[41] Although elegant, this technique is performed in organic solvent such as dimethylformamide (DMF) which causes the protein to mis-fold, rendering the synthesized protein less useful for applications requiring native protein. Other disadvantages such as low yield of peptide obtained, inefficiency for synthesis of complex and longer proteins and being highly labor intensive render this technique unattractive.[42]

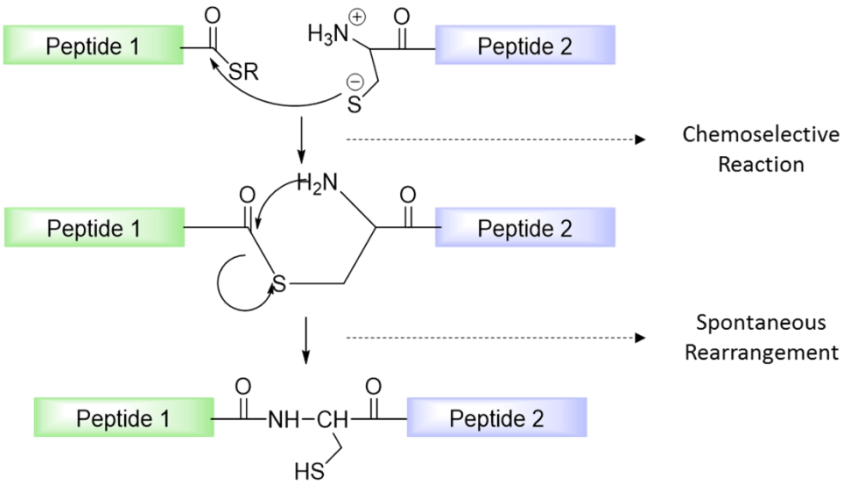


Figure I-6. General strategy of native chemical ligation.

Semi-synthetic Technique

A semi-synthetic technique known as expressed protein ligation (EPL) was devised next which combined native chemical ligation strategy with a protein expressed recombinantly.[43] The expressed protein generates an α-thioester as an intermediate on its C-terminal by protein splicing of mutant intein.[42, 43] The α-thioester generated then reacts with a peptide containing N-terminal cysteine to afford a semisynthetic product (Figure I-7).

This technique has been used for incorporation of several ncAAs at different sites in proteins to generate peptide/protein much longer than 100 residues in length.[44-46] However, like other chemical synthetic strategies, it requires presence of special chemical functionalities to be present for subsequent chemoselective reaction to take place and hence poses several restrictions on its application. The poor yield of product obtained using this strategy also makes it a cumbersome and expensive strategy to utilize for generation of ncAA incorporated protein.

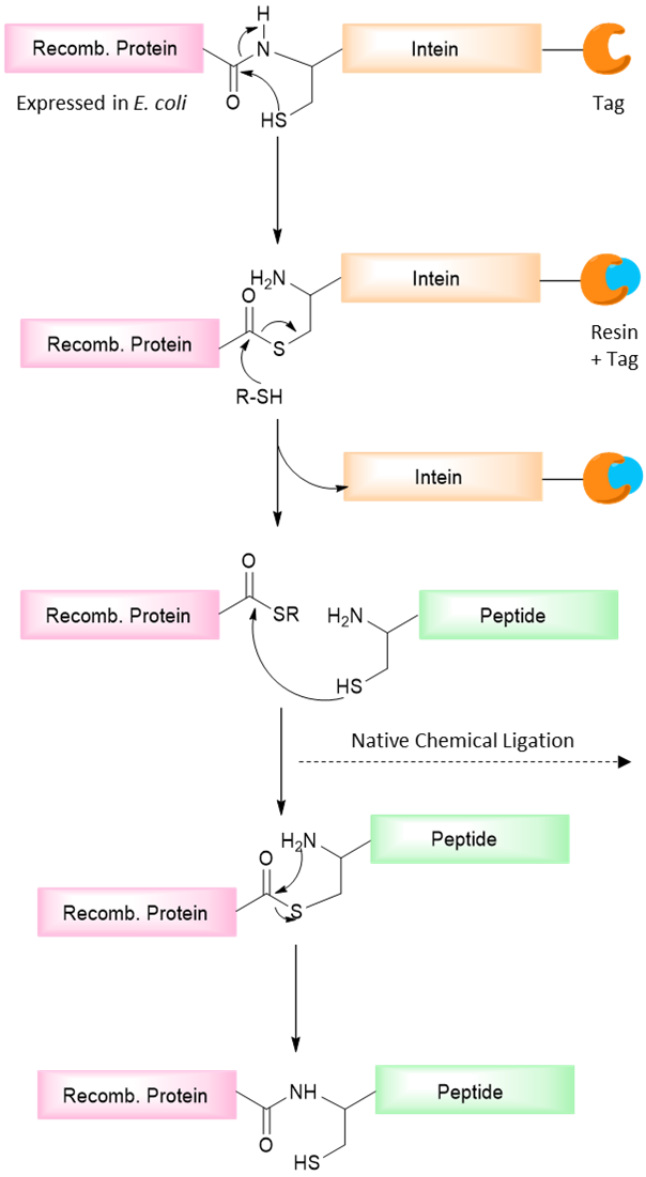


Figure I-7. Expressed protein ligation strategy.

Although, all these techniques pioneered progress in the field of ncAA incorporation, there are several limitations to these techniques. Drawbacks such as constraint on the position where modification is possible, length of the peptide chain that can be synthesized and design constraints render these techniques inefficient.[33, 47] Another disadvantage of these techniques is their inapplicability for *in vivo* protein and enzyme studies due to mis-folding of protein obtained using these methodologies.

Biosynthetic Techniques

Other biosynthetic *in vitro* techniques have been developed to incorporate ncAAs by use of protein synthesis machinery existing in nature. These techniques are based on adaptor hypothesis,[48] as per which the interaction between a mRNA codon and the anti-codon of tRNA is not dependent on the amino acid attached to the 3' terminal of the aminoacyl charged tRNA and therefore an aminoacyl tRNA (x-tRNA^x) can be charged with an ncAA (y-tRNA^x) to incorporate it into a peptide *in vitro*.[49, 50] The major drawback of this technique is non-site-specific incorporation of the desired ncAA in proteins as it gets installed throughout the protein sequence at the sites corresponding to the specific codon.[51] Also, this technique yields misincorporated protein as the ncAA is in competition with the canonical amino acid, resulting in the latter being incorporated too at a few of the corresponding codon sites.[52] Another drawback of this method is the limitation on the ncAAs that can be incorporated using this system based on their chemical reactivity and interaction with tRNAs as the tRNA cannot by enzymatically mischarged due to its high specificity and chemically mischarging the tRNA poses a challenge due to

several reactive sites on the tRNA.[53] Other disadvantages include high expense and low yield of protein obtained using this technique as it can usually be employed only on a small scale.

The next major advancement in this field occurred in 1989 with the development of a semisynthetic general technique for site-specific incorporation of ncAA by Noren *et al*.[54] This technique utilized UAG, a stop codon, to code for the desired ncAA instead of its usual function of terminating the polypeptide synthesis in the natural protein synthesis machinery. Since UAG is the most rarely used stop codon out of the three stop codons, namely amber, opal and ochre stop codons (UAG, UGA and UAA), it is seldom utilized in termination of essential genes in organisms. Based on this knowledge, the desired site for ncAA incorporation is mutated to TAG using site-directed mutagenesis technique and a chemically modified amino-acylated suppressor tRNA, directed specifically against UAG codon, is supplied to this system to get site-specific ncAA incorporation.[54, 55]

As mentioned by the authors, [54] the other scientific observations that contributed to the development of this technique are as follows: 1) Development of functionally efficient amber suppressor tRNAs,[56] 2) Experimental evidence of adaptor hypothesis as mentioned previously, according to which the interaction between mRNA codon & anti-codon region of tRNA is independent of the amino acid attached to the 3' terminus of the aminoacylated tRNA,[49, 57] 3) Broad substrate specificity demonstrated by peptidyltransferases and elongation factors, indicating that ribosomes can accommodate an array of different amino acids with modified side chains, 4) Chemical synthesis approach developed to obtain tRNAs aminoacylated with amino acids lacking free amino

group (N-protected amino acids) which could only be accommodated at the P-site of ribosome[58, 59] and 5) the subsequent development of strategy to synthesize tRNAs aminoacylated with N-deprotected amino acid derivatives, so that they could be accommodated at the A-site of the ribosome as well, thus overcoming the challenge of chemically mischarging tRNAs.[60]

In Vivo Incorporation Technique

The successful *in vitro* implementation of this technique prompted progress in development of strategy to achieve *in vivo* incorporation of ncAA by extending the approach used in this methodology. Earlier *in vivo* techniques involved the use of auxotrophic strains to incorporate ncAAs but possessed several disadvantages such as requirement of the ncAA being incorporated to be a close structural homolog of the canonical amino acids and limitations on applicability of this method due to the toxicity of the ncAA being incorporated to the biological system and misincorporation of the homolog canonical amino acid by this system.[61]

It was recognized that for efficient *in vivo* incorporation of ncAAs, the amino acid being incorporated requires to be soluble, stable and transportable across the cellular membrane. Apart from these specification requirements for the ncAA, a system would be required with an orthogonal tRNA and corresponding tRNA synthetase which utilize a nonsense codon to encode the ncAA. Orthogonality implies that this tRNA is not aminoacylated by any of the 20-canonical aminoacyl tRNA synthetases and correspondingly the orthogonal tRNA synthetase does not charge any of the 20-canonical

aminoacyl tRNAs. This is to ensure that the ncAA charged tRNA avoids detection by the cellular proofreading system which prevents misacylation of the tRNAs.[62] It was discerned that orthogonality could be achieved by utilizing tRNA/synthetase pair from a different organism, granted the inefficiency of cross-species aminoacylation.

The first successful implementation of such a system employed tyrosyl-tRNA/synthetase pair from archaea *Methanococcus jannaschii* for ncAA incorporation in *E. coli*.[63] This tRNA/synthetase pair was chosen since it does not recognize endogenous tyrosyl-tRNA /synthetase of *E. coli* due to the differences in recognition elements between them.[64] The tRNA acceptor stem plays a prominent role in recognition of the tRNA by the synthetase and the acceptor stems of *M. jannaschii* $tRNA^{Tyr}$ differs from that of *E. coli* $tRNA^{Tyr}$. The first base pairs on the acceptor stem of *M. jannaschii* $tRNA^{Tyr}$ are C1:G72, whereas in *E. coli* the first base pairs are G1:C72, hence they are orthogonal to each other. Also, the *M. jannaschii* tRNA synthetase lacks a proofreading mechanism, which implies that it does not edit out the non-canonical amino acid attached on the 3' terminus of the tRNA.[65] It has also been shown that $tRNA^{Tyr}_{CUA}$, which is the $tRNA^{tyr}$ with modified anticodon region, is not recognized by other *E. coli* synthetases but serves well in the *E. coli* protein translation machinery.[63, 64] This system has been evolved and utilized to incorporate several different ncAAs since its first application.[42]

Although great ncAA substrate diversity was achieved by engineering and evolving this *M. jannaschii* tRNA/synthetase pair, due to constraints posed by amino acids that can be accommodated at the catalytic domain of this tRNA synthetase, this system also has its limitations. Hence, other possible orthogonal tRNA/synthetase pairs were also

explored to increase the chemical diversity of amino acids that can be incorporated using similar design. Several orthogonal pairs were utilized to this end including aspartyl-tRNA/synthetase, tyrosyl-tRNA/synthetase, glutaminyl-tRNA/synthetase pairs from *Saccharomyces cerevisiae*.[66-68]

The next development occurred after the discovery of 22nd genetically encoded amino acid, pyrrolysine,[69, 70] which paved the way for utilization of pyrrolysyl-tRNA/synthetase pair for ncAA incorporation. This system was used in all the studies mentioned later and is explained in detail below.

Pyrrolysine Biosynthesis and Incorporation

After the discovery of selenocysteine as the 21st genetically encoded amino acid in 1986,[71] pyrrolysine was identified as the 22nd amino acid almost two decades later in 2002.[69, 70] These two amino acids are unique in the sense that they are naturally encoded by nonsense codons UGA and UAG respectively, however the mechanism of their encoding is vastly different from each other. Out of these two, mechanism of encoding pyrrolysine is of great interest since it is encoded by an in-frame amber stop codon (UAG) and has become an important chemical biology tool for ncAA incorporation.

As first observed in the archaeal *Methanosarcinaceae* family, pyrrolysine is biosynthesized from two molecules of lysine in the natural pathway. The gene products of *pylBCD* gene, namely PylB, PylC and PylD, implement the biosynthesis process as illustrated in Figure I-8.[72] The product PylB (radical SAM enzyme) converts one of the lysine molecules into a methylated D-ornithine derivative which then ligates to the other

lysine residue in a reaction catalyzed by PylC (member of carbomyl phosphate synthetase family). PylD (contains NAD-binding domain) then catalyzes the oxidation of the resultant dipeptide formed, leading to elimination of water and formation of pyrrolysine as the final product.[73, 74] The remaining *pyl* genes, *pylT* and *pylS* produce pyrrolysyl tRNA ($tRNA^{Pyl}$) and pyrrolysyl-tRNA synthetase (PylRS) as the gene products. The PylRS aminoacylates the $tRNA^{Pyl}$ with pyrrolysine which subsequently encodes this amino acid in the in-frame UAG codon present in MtmB, MtbB and MttB, which are the pyrrolysine containing proteins responsible for metabolism of monomethylamine, dimethylamine and trimethylamine respectively in the methanogens to generate methane.[75-77] Apart from methanogenic archaea, pyrrolysine and hence *pyl* genes are also found in certain bacteria, namely *Desulfitobacterium hafniense* and deltaproteobacteria found in *Olavius algarvensis* worm.[78, 79]

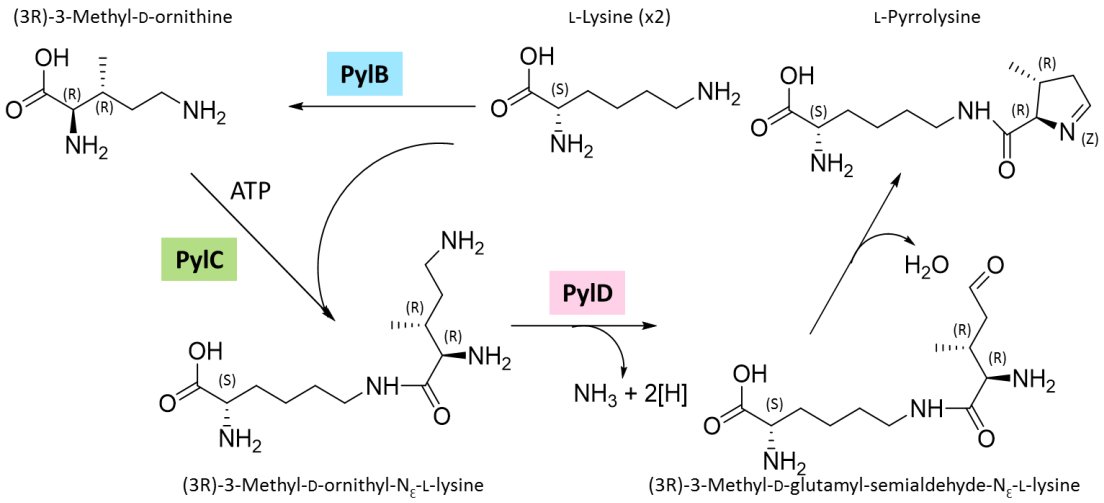


Figure I-8. Pyrrolysine biosynthesis pathway.

This archaeal pyrrolysyl-tRNA/synthetase pair was thus recognized as potential orthogonal pair for efficient incorporation of ncAAs in *E. coli*[80] and efforts were then directed in engineering this system to achieve substrate diversity of ncAA that can be recognized by the PylRS. In recent years, engineered $tRNA^{Pyl}$/ PylRS from *Methanosarcina barkeri* and *Methanosarcina mazei* have become dominantly used orthogonal incorporation pairs for genetic incorporation of ncAAs in *E. coli*.[81]

Pyrrolysyl-tRNA Synthetase Structure

The structure of pyrrolysyl-tRNA synthetase has been studied in detail since pyrrolysine was discovered as the 22^{nd} amino acid in order to determine strategies for engineering the synthetase for modifying its ncAA substrate recognition ability.

PylRS is a homodimer and is encoded by *pylS* gene in archaea with each monomer consisting of a distinct catalytic C-terminal domain and an N-terminal domain.[82] In bacteria, the PylRS is encoded by two different genes *pylSc* and *pylSn* which encode the C-terminal and the N-terminal domain of the synthetase respectively.[83] This tRNA synthetase belongs to class II of aminoacyl-tRNA synthetases as determined by the structural details of its ATP-binding domain.[82] Its catalytic domain consists of antiparallel β-sheets surrounded by α-helices which are typical characteristics of class II aminoacyl-tRNA synthetases.[84]

The crystal structures of C-terminal domains of *Methanosarcina mazei* PylRS and *Desulfitobacterium hafniense* PylRS have been resolved using X-ray crystallography and provide insight into substrate recognition elements of its catalytic domain (Figure I-9).[82,

[83, 85, 86] Hydrophobic interactions between the substrate and the PylRS determine the affinity of the synthetase for the substrate and the specificity of the substrate is determined by the hydrogen bonding interactions between the active site of the synthetase and the substrate.[82] Based on this knowledge it was predicted that the potential substrates for the synthetase should contain certain key structural elements such as the carboxylate group should be separated from the α-amino group by six atoms.[87] Since canonical amino acids lack this structural element, the pyrrolysyl-tRNA/synthetase pair acts as an excellent orthogonal pair. Also, it was determined that lysine derivatives act as suitable substrates for the wild-type synthetase without the need of much engineering of the active site, since the synthetase can tolerate alterations and substitutions to the pyrroline side chain of the substrate as long as the lysine derivatives have a bulky hydrophobic group that can be accommodated in the active site in place of pyrroline ring.[82, 88]

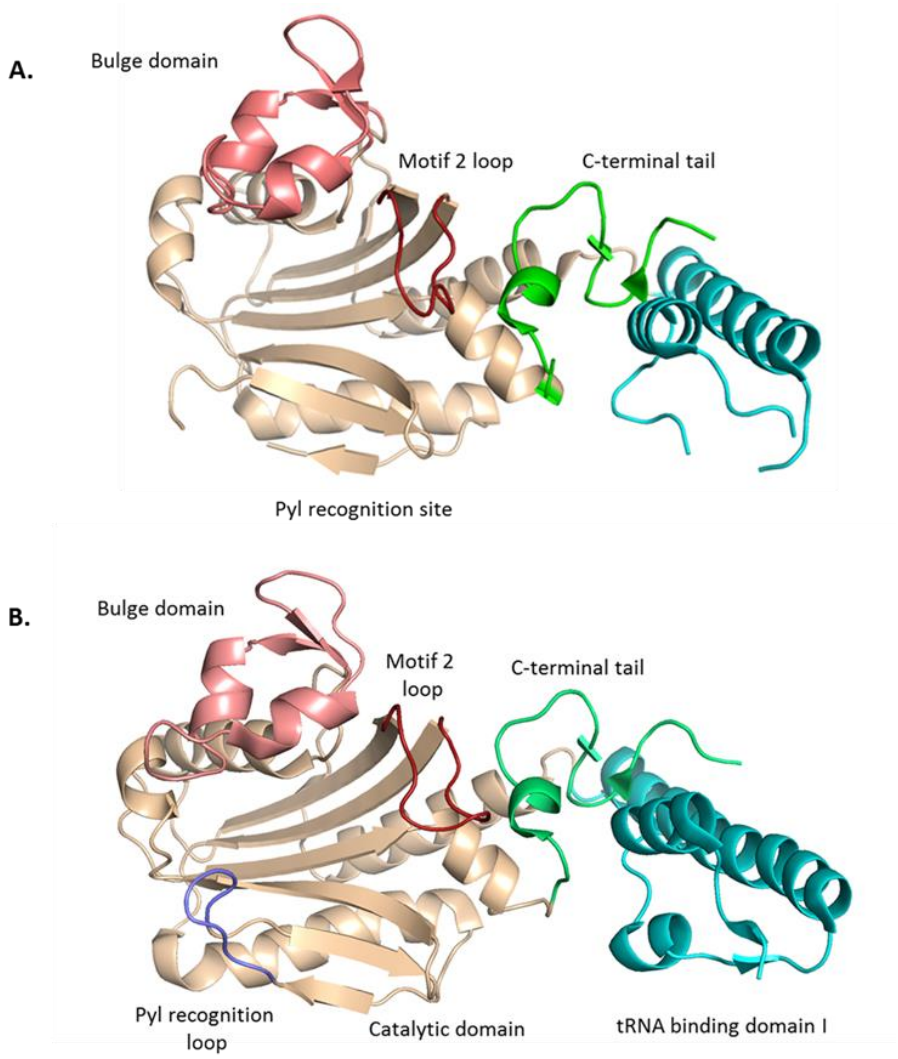


Figure I-9. A. Catalytic (C-terminal) domain of *Methanosarcina mazei* PylRS (PDB:2E3C). **B.** Catalytic (C-terminal) domain of *Desulfitobacterium hafniense* PylRS (PDB:2ZNJ)

As mentioned earlier, the substrate specificity is determined by the hydrophobic and hydrogen bonding interactions between the substrate and the active site of the synthetase. One of the key residues at the active site of the synthetase is Y384 present on the mobile loop, which plays an important role in orienting the substrate at the catalytic domain through hydrogen bonding interactions between its -OH group and the nitrogen atoms on the pyrrole ring and the α-amino group.[82, 89] The Y384 residue also greatly stabilizes the complex formed between the substrate and the active site of the synthetase through van der Waals interaction energy gained between substrate and itself. Another important residue determining the substrate specificity is N346 residue which forms an indirect hydrogen bond with the α-amino group of the pyrrolysine through water and a direct hydrogen bond with the carbonyl oxygen.[82] The hydrogen bonding interaction between R330 and the carbonyl group also helps in the proper orientation of the substrate.[82, 86] The C348 residue at the active site of the synthetase has also been shown to interact with the carbonyl group of the substrate in place of N346 residue in case of certain ncAAs.[90]

Based on the structural and functional knowledge of the key residues of the catalytic domain of the PylRS, the active site of *Methanosarcina barkeri* and *Methanosarcina mazei* PylRS have been modified by several researchers over the last decade for increasing the suitable ncAA substrate range that can be recognized by the synthetase and include several lysine and phenylalanine derivatives.[91] This dissertation explores the ability to further expand the suitable substrates that can be incorporated by this orthogonal pyrrolysyl-tRNA/synthetase system to several histidine derivatives as

mentioned in chapter III. The engineered pyrrolysine-tRNA/ synthetase system used for studies mentioned in the following chapters is based on the pyrrolysine incorporation machinery as illustrated in Figure I-10.

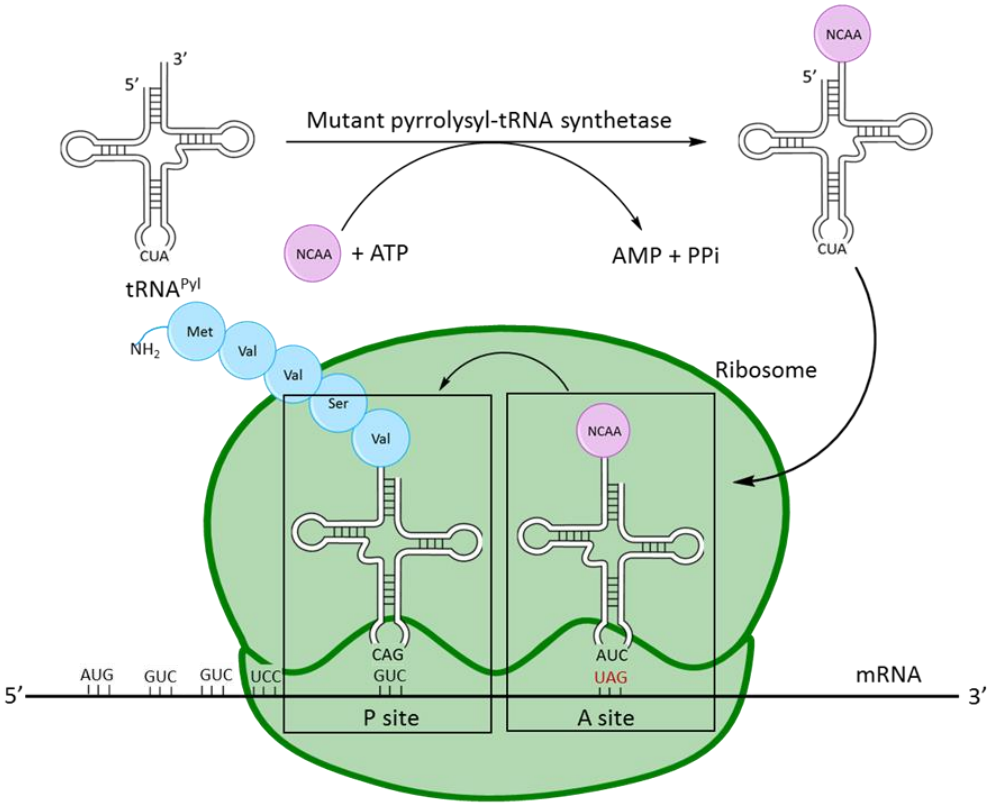


Figure I-10. Pyrrolysine machinery used for ncAA incorporation.

Although, so much is known about the catalytic C-terminal domain of the archaeal and bacterial pyrrolysyl-tRNA synthetase, the function of N-terminal domain of the synthetase remains relatively unexplored. This is due to the inability in crystallizing the N-terminal domain caused by its insolubility. Recent studies have shown that the N-terminal domain of the PylRS plays an important role in binding the $tRNA^{Pyl}$.[92] This dissertation explores effect of engineering this N-terminal domain on overall incorporation efficiency of the system as detailed in chapter II.

CHAPTER II

ENGINEERING THE N-TERMINAL DOMAIN OF PYRROLYSYL-TRNA SYNTHETASE

Introduction

The last two decades have seen tremendous growth and interest in the field of site-specific genetic incorporation of non-canonical amino acids (ncAAs) due to the applicability of this method to modify the structural and functional properties of various enzymes and proteins by altering the desired amino acid at the sites of interest.[93] The genetic ncAA incorporation system, as illustrated in Figure I-10, utilizes a rare codon suppression system, usually amber (TAG) codon and consists of an orthogonal tRNA and corresponding tRNA synthetase that can incorporate the required ncAA at the amber mutation site in the protein of interest.[94] For expression in the bacterial *Escherichia coli* system, orthogonality is achieved by utilizing the amber suppressor tRNA/synthetase pairs from archaea such as pyrrolysyl-tRNA/synthetase from *Methanocaldococcus jannaschii*, *Methanosarcina barkeri,* or *Methanosarcina mazei*, or the tyrosyl-tRNA/synthetase from *Methanocaldococcus jannaschii*.[80, 94, 95]

As covered in detail in previous chapter, site-specific genetic incorporation system offers several advantages over other commonly used techniques to achieve the same goal such as native chemical ligation (NCL), expressed protein ligation (EPL) and use of auxotrophic bacterial strains.[32, 33, 96] These methods have limitations on the possible

modifications, sites that can be modified, the length of the peptide that can be synthesized (EPL and NCL) and drawbacks such as misincorporation and limitations due to toxicity of the ncAA being incorporated (auxotrophic strains).[34, 47, 97] Although the site-specific ncAA genetic incorporation technique overcomes these major disadvantages and has evolved over the last few years to incorporate an extensive repertoire and diverse group of amino acids with various functional groups, its major shortcoming that remains to be addressed is the overall poor efficiency of these different ncAAs incorporation and low protein expression yield obtained with ncAA incorporated product.[98] The factors affecting the incorporation efficiency range from ncAA solubility to the efficiency of the PylRS system itself in generating sufficient charged ncAA-tRNA^{Pyl}. The aim of our study was to target the latter factor mentioned above by engineering the synthetase to improve the incorporation efficiency of the system. Previous approaches to resolve this issue have involved utilizing methods to evolve tRNA^{Pyl} to optimize the interaction between the pyrrolysyl-synthetase and tRNA^{Pyl}.[99]

For our study, we chose to focus on the *Methanosarcina mazei* pyrrolysyl-tRNA/synthetase (mm tRNA^{Pyl}/PylRS), which we have used to incorporate various phenylalanine, lysine and histidine derivatives, as described in the next chapter.[100-102] This archaeal synthetase is encoded by *pylS* gene and its catalytic C-terminal domain is well studied with resolved crystal structure (PDB: 2E3C) (Figure I-9).[73, 103] Unlike other synthetases, PylRS has high promiscuity and low selectivity towards substrates rendering it suitable for incorporation of a wide variety of ncAAs by mutating a few key residues on its catalytic domain.[104] The N-terminal domain of the synthetase, however, remains

relatively inadequately studied due to the hydrophobic nature of this domain which renders the protein unstable for crystallization.[85] To obtain higher amounts of charged ncAA-tRNA^{Pyl}, a simple solution would be to express larger amount of PylRS. However, due to the insoluble nature of the PylRS, it cannot be overexpressed in the system to improve the ncAA incorporation efficiency as it precipitates out in the expression media. It was shown in a recent study by R. Jiang et al. (2012) that the N-terminal domain of *Methanosarcina barkeri* PylRS plays a role in binding tRNA^{Pyl} thus elucidating its relatively unknown function. [92] In this study, they identified the key nucleotide residues on the D-stem, TψC loop and the variable loop of bacterial tRNA^{Pyl} (from *Desulfitobacterium hafniense*) which interact with bacterial homologs of the N-terminal and C-terminal domains of the mm pyrrolysyl-tRNA synthetase (Figure II-1). All these interactions between tRNA^{Pyl} and N-terminal domain of synthetase are present distal from the acceptor stem, which is responsible for aminoacylation with ncAA. We thus chose to target the N-terminal domain of *M. mazei* PylRS, as it is unlikely to compromise the aminoacylation of the tRNA^{Pyl}, to improve its interaction with the tRNA^{Pyl} and hence the ncAA incorporation efficiency.

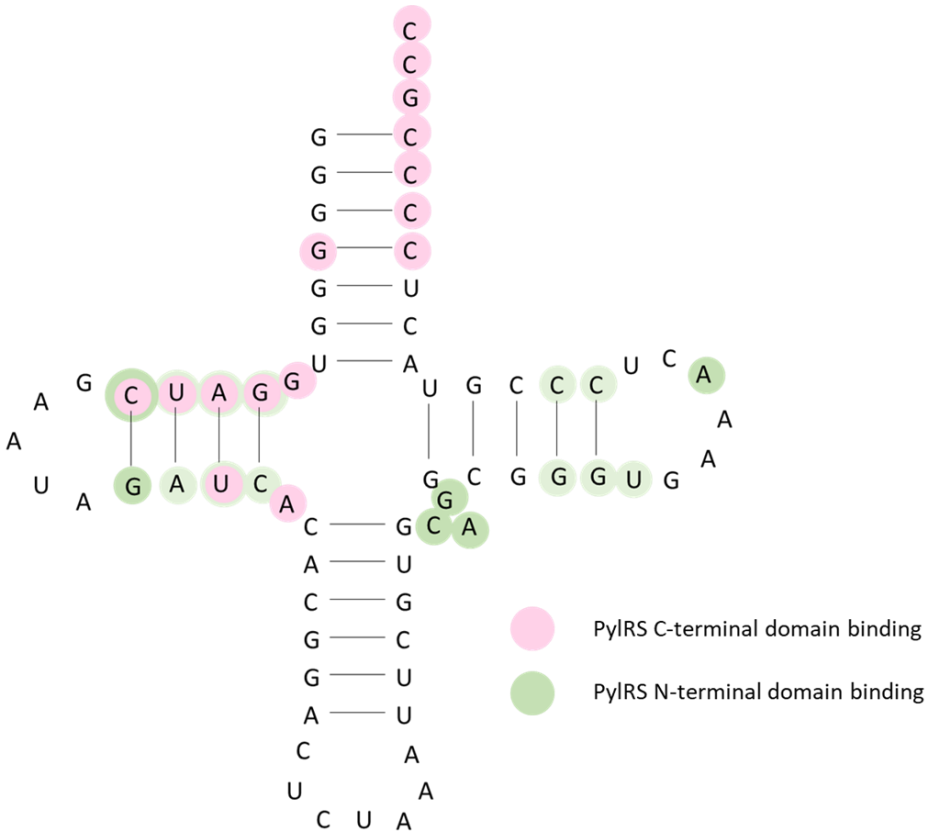


Figure II-1. *D. hafniense* $tRNA^{Pyl}$ showing key bases involved in binding the C-terminal domain and the N-terminal domain of the bacterial PylRS.

We employed a screening strategy to identify mutations on the N-terminal domain of the PylRS based on this scheme and is discussed in detail later. As most PylRS mutants developed for incorporation of specific ncAAs have their mutations on the C-terminal domain of the PylRS, can be simply transferred to other PylRS mutant synthetases for improving incorporation efficiency of their corresponding ncAAs. To test this, the mutations identified on the optimized system were then transposed to AcKRS, a pyrrolysyl-synthetase variant, used to incorporate the post-translational modification acetyl lysine into histone protein and tested the deacetylase activity of Sirtuin 1 (SIRT1) enzyme on histone tetramer. The SIRT1 enzyme belongs to NAD-dependent histone deacetylase (HDAC) family.[105] Among the 7 sirtuins (SIRT1-7) belonging to this HDAC family, SIRT1 is localized in nucleus and has been shown to target acetylated histone sites such as histones H3K9, H3K56 and H4K16 *in vitro*.[106, 107] As discussed later, we tested the activity of this enzyme on reconstituted H3-H4 tetramer with acetylated H3K23 using the optimized AcKRS.

Experimental Details

Error-Prone PCR (EP-PCR) Conditions

All the dNTPs and buffer used for the PCR reaction were purchased from Thermo Fischer Scientific®, the oligo primers were purchased from Integrated DNA Technology®. The reactions were performed on C1000™ Thermo Cycler from Bio-Rad® with the following thermocycling conditions: Initial denaturation- 95°C for 30 seconds, 25 cycles

with 56°C for 30s, 68°C for 90s, followed by final extension- 68°C for 5 minutes. Each PCR reaction consisted of following contents listed below:

10x Thermo® Buffer	10 µl
50x dNTP mix	2 µl
100 mM dCTP	1 µl
100 mM dTTP	1 µl
50 mM $MgCl_2$	14 µl
50 mM $MnCl_2$	1 µl
DNA template	2 µl
Primer- f (100 µM)	1 µl
Primer- r (100 µM)	1 µl
Taq Polymerase (50)	1 µl
H_2O	66 µl
	100 µl

Library Construction and Selection

Codon optimized *M. mazei* PylRS sequence (in pEVOL vector) was used as the template for construction of library using EP-PCR.[108] The mm PylRS sequence gene was amplified from pEVOL-mm PylRS-PylT plasmid by primers, Opt- mmPylRS-SpeI-F and Opt- mmPylRS-BstXI-R using the EP-PCR reaction conditions as mentioned above to introduce random mutations on the N-terminal domain of the synthetase sequence. The primers were designed to amplify the N-terminal tail. The pEVOL vector was amplified using the primers pEVOL-Opt- mmPylRS-BstXI-F and pEVOL-Opt- mmPylRS-SpeI-R to introduce restriction digestion cutting sites in the vector (Figure II-2). The gene product was then digested by *SpeI* and *BstXI* restriction enzymes (purchased from NEB®), purified and ligated into the amplified pEVOL vector digested by *SpeI* and *BstXI* to generate plasmid pEVOL-R1-mm PylRS-PylT plasmid. The generated library was directly used to transform *E. coli* Top10 pBAD sfGFP N134TAG cells using electroporation.

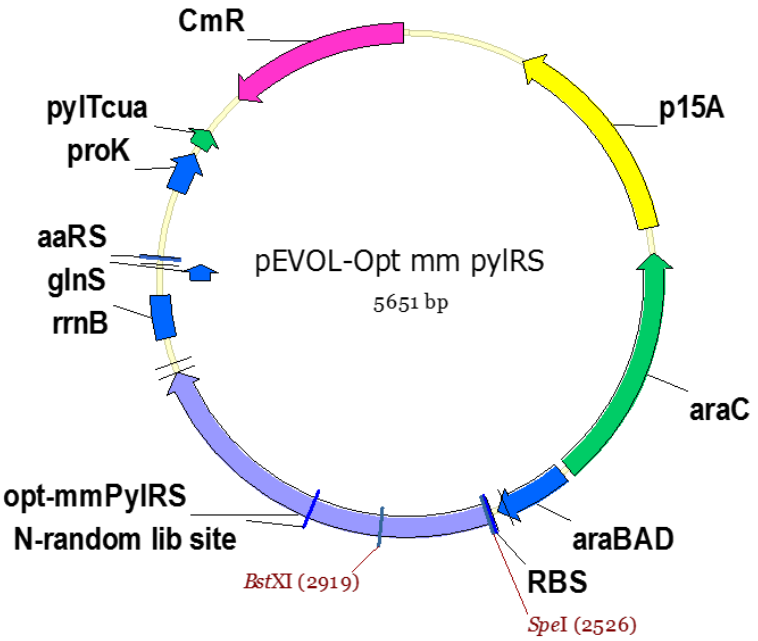


Figure II-2. Plasmid map of pEVOL mm PylRS-PylT showing the site on N-terminal tail for library generation.

The transformed cells were plated on screening plates containing LB agar, supplemented with Chloramphenicol (Cm, 34 μg/mL), Ampicillin (Amp, 100 μg/mL), 0.2% Arabinose and ncAA (0.2 mM BocK). Plates were incubated at 37°C overnight. The cell phenotypic differences of viability and fluorescence intensity were analyzed and photographed (Figure II-3). The clones with strong fluorescence on the ncAA containing plate were selected as the active clones and were amplified. The pEVOL plasmid of the active clones was isolated and sequenced.

The entire process of library construction and selection was repeated 3 times with the best clone from each round selected as template for the next round.

All the plasmid sequences were confirmed by DNA sequencing (Table II-1). All oligonucleotide primers were purchased from Integrated DNA Technologies, Inc.

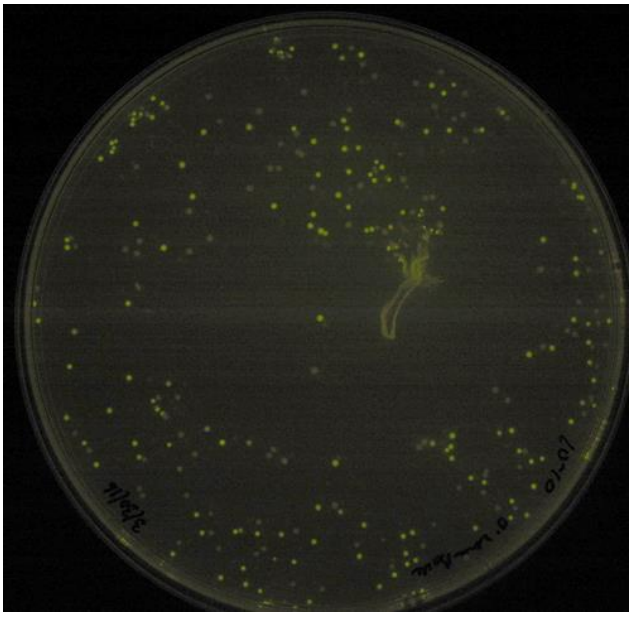


Figure II-3. Screening plate (Amp, Cm, 0.2% arabinose and 0.2 mM BocK) with 10^{-8} dilution of cloned library plated (Round 3).

Table II-1. Sequencing results of few clones from Round I, II and III of selection.

Clone Name	Mutations
R1-6 mm PylRS	N121R
R1-7 mm PylRS	R19H
R2-13 mm PylRS	R19H, T56S, K109E, T121S
R2-23 mm PylRS	R19H, H29R
R2-26 mm PylRS	R19H, I36N, M40T, R113H, Q125R
R3-1 mm PylRS	R19H, H29R, K67Q, K106E, M108K, E123V
R3-3 mm PylRS	R19H, H29R, S111P
R3-7 mm PylRS	R19H, H28R, H29R, V31A, S32N, K105E, T122S
R3-8 mm PylRS	R19H, K27Q, H29R, T122S
R3-10 mm PylRS	R19H, K25E, H28Q, H29R, D78H, T122S, S160T, D206H, S209T
R3-11 mm PylRS	R19H, H29R, T122S
R3-12 mm PylRS	R19H, K25D, H29R, R33L, N87D, T122S, Q194H
R3-14 mm PylRS	R19H, H28R, H29R, K96E, T122S

Primer List

Opt- mmPylRS-SpeI-F, 5’-GGA ATT ACT AGT ATG GAT AAG AAA CCG C-3’

Opt- mmPylRS-BstXI-R, 5’-TTA GAA CCA GAC GGC TGG GCC TGT GC-3’

pEVOL-Opt- mmPylRS-BstXI-F, 5’- ACA GGC CCA GCC GTC TGG TTC TAA G -3’

pEVOL-Opt- mmPylRS-SpeI-R, 5’- ATC CAT ACT AGT AAT TCC TCC TGT TA -3’

PylRS L301M Y306L L309A F, 5’- ATG GCA CCG AAC CTG CTG AAT TAC GCA CGT AAA -3’

PylRS L301M Y306L L309A R, 5’- CAT CGG ACG CAG GCA GAA GTT CTT GTC CAC -3’

PylRS C348F F, 5’- CAG ATG GGT TCT GGC TGC AC -3’

PylRS C348F R, 5’- AAA AAA GTT CAG CAT AGT GAA CTC TTC C -3’

Opt mmpylRS L309T F, 5’- CGT AAA CTG GAT CGT GCA CTG CCG -3’

Opt mmpylRS L309T R, 5’- GGT GTA ATT GTA CAG GTT CGG TGC CAG C -3’

Opt mmpylRS C348G F, 5’- CAG ATG GGT TCT GGC TGC AC -3’

Opt mmpylRS C348G R, 5’- ACC AAA GTT CAG CAT AGT GAA CTC TTC C -3’

Opt mmpylRS Y384F F, 5’ - GGT GAC ACT CTG GAC GTT ATG CAT -3’

Opt mmpylRS Y384F R, 5’- GAA AAC CAT GCA GGA ATC ACC AAC -3’

pET28a-SUMO-NcoI-F, 5’ - GGA GAT ATA CCA TGG GCA GCA GCC -3’

pET28a-SUMO-BamHI-R, 5’ - GAC GGG ATC CAC CAC CAA TCT GTT C -3’

TEV-hH4-del-F, 5’- GAA AAT CTG TAC TTC CAG TCT GGT -3’

TEV-hH4-del-R, 5’- GGA TCC TGG CTG TGG TGA T -3’

Plasmid Construction

Construction of pEVOL mm AcKRS-PylT and pEVOL mm AcdKRS-PylT:

To construct pEVOL mm AcKRS-PylT (R3-11), mutations were generated on pEVOL mm PylRS-PylT (R3-11) by site-directed mutagenesis, using the following primers: PylRS L301M Y306L L309A F, PylRS L301M Y306L L309A R, PylRS C348F F and PylRS C348F R. The sequence of the generated plasmid was confirmed using DNA sequencing.

Similarly, for construction of pEVOL mm AcdKRS-PylT (R3-11), mutations were generated on pEVOL mm PylRS-PylT (R3-11) by site-directed mutagenesis, using the following primers: Opt mmpylRS L309T F, Opt mmpylRS L309T R, Opt mmpylRS C348G F, Opt mmpylRS C348G R, Opt mmpylRS Y384F F and Opt mmpylRS Y384F R. The plasmid sequence of the plasmid generated was also confirmed by DNA sequencing. All the primers used in this experiment were purchased from Integrated DNA Technologies Inc.

Construction of pETDuetI-His6x-SUMO-TEV-H4 plasmid:

For construction of pETDuetI-His6x-SUMO-TEV-H4 plasmid (Figure II-4), His6x-SUMO fragment was amplified using pET28a-SUMO-NcoI-F and pET28a-SUMO-BamHI-R primers to introduce *NcoI* and *BamHI* restriction digestion sites on the amplified fragment. TEV-hH4-del-F and TEV-hH4-del-R primers were used for deleting a nucleotide from pETDuetI-sfGFP-TEV-hH4 plasmid to account for frame shift.

Both the plasmid and amplified fragment were digested using *NcoI* and *BamHI* restriction enzymes and ligated to generate histone H4 plasmid with SUMO attached on the N-terminal tail of the histone. The plasmid sequence was confirmed by DNA sequencing. All the primers used were purchased from Integrated DNA Technologies Inc.

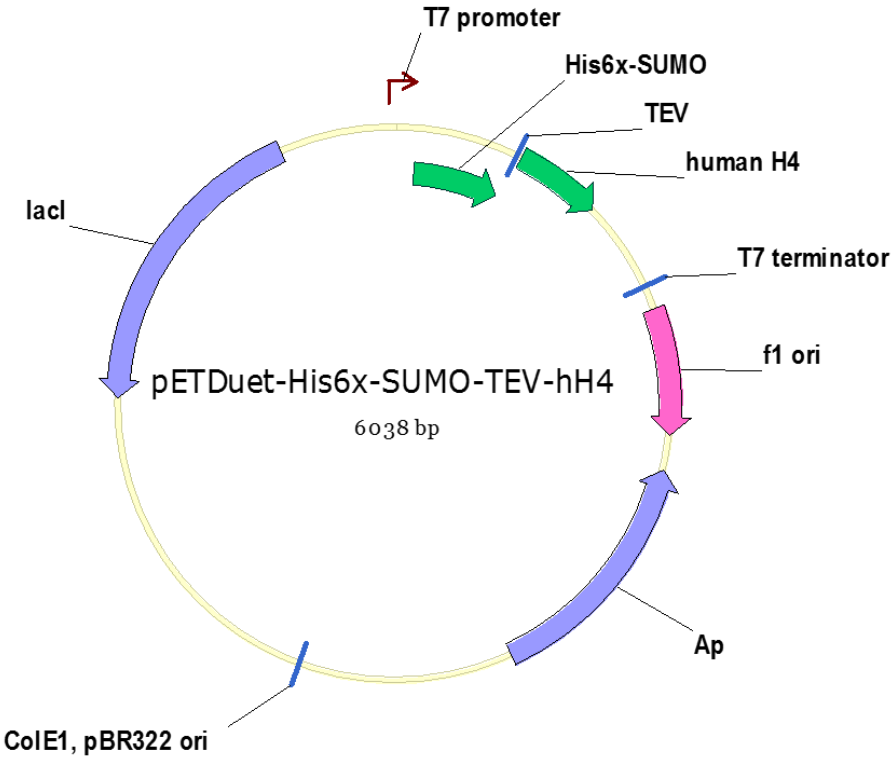


Figure II-4. Plasmid map of pETDuetI His6x-SUMO-TEV-hH4 constructed.

DNA Sequences

M. mazei PylRS:

atggataagaaaccgctgaatactctgatttctgcaactggtctgtggatgagccgtaccggcaccatccacaagatcaaacac
cacgaggtttcccgtagcaaaatctacatcgaaatggcgtgcggtgaccacctggtggtaaacaactcccgttcttctcgtactg
cacgtgctctgcgccaccacaagtaccgtaagacctgcaagcgctgtcgcgtgtctgatgaagacctgaacaaattcctgacta
aagcgaacgaagatcagacttctgtgaaggtgaaagttgtttctgccccaacccgcaccaagaaagcgatgccgaagtccgtt
gcacgcgctccgaaaccgctggagaacaccgaagccgcacaggcccagccgtctggttctaagttttctccggcaatcccgg
tttctactcaggagtctgtgtctgtgccagcttctgttagcacttctatttcctctatcagcactggtgcgactgcgtccgctctggta
aaaggtaacactaacccgatcaccagcatgtctgctccggttcaggcttctgcaccggcactgactaaaagccagactgaccgt
ctggaggttctgctgaacccgaaagatgaaatcagcctgaactctggcaaaccgttccgtgaactggaatccgaactgctgtct
cgtcgtaagaaagacctgcaacaaatctatgctgaagagcgtgaaaactacctgggtaaactggaacgtgaaatcacccgtttc
tttgtggaccgtggtttcctggaaatcaagtctccgatcctgatcccgctggaatacatcgagcgcatgggtattgataacgacac
cgaactgtccaagcagattttccgtgtggacaagaacttctgcctgcgtccgatgctggcaccgaacctgtacaattacctgcgt
aaactggatcgtgcactgccggacccgatcaaaatctttgaaatcggtccatgctatcgtaaggagagcgacggtaaagaaca
cctggaagagttcactatgctgaacttttgtcagatgggttctggctgcacccgtgaaaatctggaatctatcatcaccgacttcct
gaaccacctgggcattgacttcaaaatcgttggtgattcctgcatggtttacggtgacactctggacgttatgcatggtgatctgga
actgagcagcgctgttgtgggtccgattccgctggatcgtgaatggggtatcgataaaccgtggattggtgctggcttcggtctg
gaacgtctgctgaaagttaagcacgactttaagaacatcaaacgtgctgcgcgttccgagtcctattacaacggcattagcacta
acctgtaa

M. mazei AcKRS:

atggataagaaaccgctgaatactctgatttctgcaactggtctgtggatgagccgtaccggcaccatccacaagatcaaacac
cacgaggtttcccgtagcaaaatctacatcgaaatggcgtgcggtgaccacctggtggtaaacaactcccgttcttctcgtactg
cacgtgctctgcgccaccacaagtaccgtaagacctgcaagcgctgtcgcgtgtctgatgaagacctgaacaaattcctgacta
aagcgaacgaagatcagacttctgtgaaggtgaaagttgtttctgccccaacccgcaccaagaaagcgatgccgaagtccgtt
gcacgcgctccgaaaccgctggagaacaccgaagccgcacaggcccagccgtctggttctaagttttctccggcaatcccgg
tttctactcaggagtctgtgtctgtgccagcttctgttagcacttctatttcctctatcagcactggtgcgactgcgtccgctctggta
aaaggtaacactaacccgatcaccagcatgtctgctccggttcaggcttctgcaccggcactgactaaaagccagactgaccgt
ctggaggttctgctgaacccgaaagatgaaatcagcctgaactctggcaaaccgttccgtgaactggaatccgaactgctgtct
cgtcgtaagaaagacctgcaacaaatctatgctgaagagcgtgaaaactacctgggtaaactggaacgtgaaatcacccgtttc
tttgtggaccgtggtttcctggaaatcaagtctccgatcctgatcccgctggaatacatcgagcgcatgggtattgataacgacac
cgaactgtccaagcagattttccgtgtggacaagaacttctgcctgcgtccgatgatggcaccgaacctgctgaatctggcgcgt
aaactggatcgtgcactgccggacccgatcaaaatctttgaaatcggtccatgctatcgtaaggagagcgacggtaaagaaca
cctggaagagttcactatgctgaacttttttcagatgggttctggctgcacccgtgaaaatctggaatctatcatcaccgacttcctg
aaccacctgggcattgacttcaaaatcgttggtgattcctgcatggtttacggtgacactctggacgttatgcatggtgatctggaa
ctgagcagcgctgttgtgggtccgattccgctggatcgtgaatggggtatcgataaaccgtggattggtgctggcttcggtctgg
aacgtctgctgaaagttaagcacgactttaagaacatcaaacgtgctgcgcgttccgagtcctattacaacggcattagcactaa
cctgtaa

M. mazei AcdKRS:

atggataagaaaccgctgaatactctgatttctgcaactggtctgtggatgagccgtaccggcaccatccacaagatcaaacac
cacgaggtttcccgtagcaaaatctacatcgaaatggcgtgcggtgaccacctggtggtaaacaactcccgttcttctcgtactg

cacgtgctctgcgccaccacaagtaccgtaagacctgcaagcgctgtcgcgtgtctgatgaagacctgaacaaattcctgacta

aagcgaacgaagatcagacttctgtgaaggtgaaagttgtttctgccccaacccgcaccaagaaagcgatgccgaagtccgtt

gcacgcgctccgaaaccgctggagaacaccgaagccgcacaggcccagccgtctggttctaagttttctccggcaatcccgg

tttctactcaggagtctgtgtctgtgccagcttctgttagcacttctatttcctctatcagcactggtgcgactgcgtccgctctggta

aaaggtaacactaacccgatcaccagcatgtctgctccggttcaggcttctgcaccggcactgactaaaagccagactgaccgt

ctggaggttctgctgaacccgaaagatgaaatcagcctgaactctggcaaaccgttccgtgaactggaatccgaactgctgtct

cgtcgtaagaaagacctgcaacaaatctatgctgaagagcgtgaaaactacctgggtaaactggaacgtgaaatcacccgtttc

tttgtggaccgtggtttcctggaaatcaagtctccgatcctgatcccgctggaatacatcgagcgcatgggtattgataacgacac

cgaactgtccaagcagattttccgtgtggacaagaacttctgcctgcgtccgatgctggcaccgaacctgtacaattacacccgt

aaactggatcgtgcactgccggacccgatcaaaatctttgaaatcggtccatgctatcgtaaggagagcgacggtaaagaaca

cctggaagagttcactatgctgaactttggtcagatgggttctggctgcacccgtgaaaatctggaatctatcatcaccgacttcct

gaaccacctgggcattgacttcaaaatcgttggtgattcctgcatggttttcggtgacactctggacgttatgcatggtgatctgga

actgagcagcgctgttgtgggtccgattccgctggatcgtgaatggggtatcgataaaccgtggattggtgctggcttcggtctg

gaacgtctgctgaaagttaagcacgactttaagaacatcaaacgtgctgcgcgttccgagtcctattacaacggcattagcacta

acctgtaa

PylT:

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

sfGFP N134TAG:

atggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaatttag

cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttcc

gtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttta

aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtt

aaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaa**tag**ggcaacattctgggtcataaact

ggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcaca

acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggat

aatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccgc

cgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattaa

His6x-TEV-hH3:

catcaccatcatcaccacagccaggatccggaaaatctgtacttccaggctcgcaccaaacagactgctcgtaagtccactggc

ggtaaagcgccgcgtaaacagctggcaaccaaggcagcgcgtaaaagcgctccagctactggcggcgtgaagaagccgca

ccgttatcgcccgggtactgtggctctgcgtgaaatccgccgctaccagaaaagcaccgaactgctgattcgcaaactgccattt

caacgtctggttcgcgaaattgctcaggatttcaaaaccgacctgcgcttccagtctagcgctgtgatggcactgcaagaggcgt

ctgaggcatatctggttggcctgttcgaagataccaacctgtgcgcaatccatgcaaagcgtgtaaccattatgccgaaagacat

ccaactggctcgtcgtatccgtggtgagcgtgcgtga

His6x-SUMO-TEV-hH4:

Atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatgtcggactcagaagtcaatc

aagaagctaagccagaggtcaagccagaagtcaagcctgagactcacatcaatttaaaggtgtccgatggatcttcagagatct

tcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaaaagacagggtaaggaaatggactccttaa

gattcttgtacgacggtattagaattcaagctgatcagacccctgaagatttggacatggaggataacgatattattgaggctcac

agagaacagattggtggtggatccgaaaatctgtacttccagtctggtcgtggtaaaggtggtaaaggcctgggtaaaggtggt

gctaagcgtcaccgtaaagtgctgcgcgacaacatccagggtatcaccaaaccagctattcgccgtctggcacgtcgcggtgg
tgtgaaacgcatcagcggtctgatctatgaagaaacccgtggtgttctgaaagtatttctggagaacgttatccgcgatgcggtg
acctacaccgaacacgcgaaacgtaagaccgttactgctatggatgttgtgtacgctctgaaacgccagggtcgtactctgtac
ggtttcggtggctga

SIRT1:

gcggacgaggcggccctcgcccttcagcccggcggctccccctcggcggcgggggccgacagggaggccgcgtcgtccc
ccgccggggagccgctccgcaagaggccgcggagagatggtcccggcctcgagcggagcccgggcgagcccggtggg
gcggccccagagcgtgaggtgccggcggcggccaggggctgcccgggtgcggcggcggcggcgctgtggcgggaggc
ggaggcagaggcggcggcggcaggcggggagcaagaggcccaggcgactgcggcggctggggaaggagacaatggg
ccgggcctgcagggcccatctcgggagccaccgctggccgacaacttgtacgacgaagacgacgacgacgagggcgagg
aggaggaagaggcggcggcggcggcgattgggtaccgagataaccttctgttcggtgatgaaattatcactaatggttttcattc
ctgtgaaagtgatgaggaggatagagcctcacatgcaagctctagtgactggactccaaggccacggataggtccatatacttttt
gttcagcaacatcttatgattggcacagatcctcgaacaattcttaaagatttattgccggaaacaatacctccacctgagttggat
gatatgacactgtggcagattgttattaatatccttttcagaaccaccaaaaaggaaaaaaagaaaagatattaatacaattgaaga
tgctgtgaaattactgcaagagtgcaaaaaaattatagttctaactggagctggggtgtctgtttcatgtggaatacctgacttcag
gtcaagggatggtatttatgctcgccttgctgtagacttcccagatcttccagatcctcaagcgatgtttgatattgaatatttcagaa
aagatccaagaccattcttcaagtttgcaaaggaaatatatcctggacaattccagccatctctctgtcacaaattcatagccttgtc
agataaggaaggaaaactacttcgcaactatacccagaacatagacacgctggaacaggttgcgggaatccaaaggataattc
agtgtcatggttcctttgcaacagcatcttgcctgatttgtaaatacaaagttgactgtgaagctgtacgaggagatatttttaatcag
gtagttcctcgatgtcctaggtgcccagctgatgaaccgcttgctatcatgaaaccagagattgtgtttttttggtgaaaatttaccag
aacagtttcatagagccatgaagtatgacaaagatgaagttgacctcctcattgttattgggtcttccctcaaagtaagaccagtag

cactaattccaagttccataccccatgaagtgcctcagatattaattaatagagaacctttgcctcatctgcattttgatgtagagctt
cttggagactgtgatgtcataattaatgaattgtgtcataggttaggtggtgaatatgccaaactttgctgtaaccctgtaaagctttc
agaaattactgaaaaacctccacgaacacaaaaagaattggcttatttgtcagagttgccacccacacctcttcatgtttcagaag
actcaagttcaccagaaagaacttcaccaccagattcttcagtgattgtcacacttttagaccaagcagctaagagtaatgatgatt
tagatgtgtctgaatcaaaaggttgtatggaagaaaaaccacaggaagtacaaacttctaggaatgttgaaagtattgctgaaca
gatggaaaatccggatttgaagaatgttggttctagtactggggagaaaaatgaaagaacttcagtggctggaacagtgagaaa
atgctggcctaatagagtggcaaaggagcagattagtaggcggcttgatggtaatcagtatctgtttttgccaccaaatcgttaca
ttttccatggcgctgaggtatattcagactctgaagatgacgtcttatcctctagttcttgtggcagtaacagtgatagtgggacatg
ccagagtccaagtttagaagaacccatggaggatgaaagtgaaattgaagaattctacaatggcttagaagatgagcctgatgtt
ccagagagagctggaggagctggatttgggactgatggagatgatcaagaggcaattaatgaagctatatctgtgaaacagga
agtaacagacatgaactatccatcaaacaaatcatag

Protein Sequences

<u>*M. mazei* PylRS</u>:

MDKKPLNTLI SATGLWMSRT GTIHKIKHHE VSRSKIYIEM ACGDHLVVNN SRSSRTARAL RHHKYRKTCK RCRVSDEDLN KFLTKANEDQ TSVKVKVVSA PTRTKKAMPK SVARAPKPLE NTEAAQAQPS GSKFSPAIPV STQESVSVPA SVSTSISSIS TGATASALVK GNTNPITSMS APVQASAPAL TKSQTDRLEV LLNPKDEISL NSGKPFRELE SELLSRRKKD LQQIYAEERE NYLGKLEREI TRFFVDRGFL EIKSPILIPL EYIERMGIDN DTELSKQIFR VDKNFCLRPM LAPNLYNYLR KLDRALPDPI KIFEIGPCYR KESDGKEHLE EFTMLNFCQM

GSGCTRENLE SIITDFLNHL GIDFKIVGDS CMVYGDTLDV MHGDLELSSA VVGPIPLDRE WGIDKPWIGA GFGLERLLKV KHDFKNIKRA ARSESYYNGI STNL

M. mazei AcKRS:

MDKKPLNTLI SATGLWMSRT GTIHKIKHHE VSRSKIYIEM ACGDHLVVNN SRSSRTARAL RHHKYRKTCK RCRVSDEDLN KFLTKANEDQ TSVKVKVVSA PTRTKKAMPK SVARAPKPLE NTEAAQAQPS GSKFSPAIPV STQESVSVPA SVSTSISSIS TGATASALVK GNTNPITSMS APVQASAPAL TKSQTDRLEV LLNPKDEISL NSGKPFRELE SELLSRRKKD LQQIYAEERE NYLGKLEREI TRFFVDRGFL EIKSPILIPL EYIERMGIDN DTELSKQIFR VDKNFCLRPM MAPNLLNLAR KLDRALPDPI KIFEIGPCYR KESDGKEHLE EFTMLNFFQM GSGCTRENLE SIITDFLNHL GIDFKIVGDS CMVYGDTLDV MHGDLELSSA VVGPIPLDRE WGIDKPWIGA GFGLERLLKV KHDFKNIKRA ARSESYYNGI STNL

M. mazei AcdKRS:

MDKKPLNTLI SATGLWMSRT GTIHKIKHHE VSRSKIYIEM ACGDHLVVNN SRSSRTARAL RHHKYRKTCK RCRVSDEDLN KFLTKANEDQ TSVKVKVVSA PTRTKKAMPK SVARAPKPLE NTEAAQAQPS GSKFSPAIPV STQESVSVPA SVSTSISSIS TGATASALVK GNTNPITSMS APVQASAPAL TKSQTDRLEV LLNPKDEISL NSGKPFRELE SELLSRRKKD LQQIYAEERE NYLGKLEREI

TRFFVDRGFL EIKSPILIPL EYIERMGIDN DTELSKQIFR VDKNFCLRPM LAPNLYNYTR KLDRALPDPI KIFEIGPCYR KESDGKEHLE EFTMLNFGQM GSGCTRENLE SIITDFLNHL GIDFKIVGDS CMVFGDTLDV MHGDLELSSA VVGPIPLDRE WGIDKPWIGA GFGLERLLKV KHDFKNIKRA ARSESYYNGI STNL

sfGFP N134X:

MVSKGEELFT GVVPILVELD GDVNGHKFSV RGEGEGDATN GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERTIS FKDDGTYKTR AEVKFEGDTL VNRIELKGID FKEXGNILGH KLEYNFNSHN VYITADKQKN GIKANFKIRH NVEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSVLSK DPNEKRDHMV LLEFVTAAGI THGMDELYKG SHHHHHH

X represents ncAA

His6x-TEV-hH3:

HHHHHHSQD PENLYFQART KQTARKSTGG KAPRKQLATK AARKSAPATG GVKKPHRYRP GTVALREIRR YQKSTELLIR KLPFQRLVRE IAQDFKTDLR FQSSAVMALQ EASEAYLVGL FEDTNLCAIH AKRVTIMPKD IQLARRIRGE RA

His6x-SUMO-TEV-hH4:

MGSSHHHHHH SSGLVPRGSH MSDSEVNQEA KPEVKPEVKP ETHINLKVSD GSSEIFFKIK KTTPLRRLME AFAKRQGKEM DSLRFLYDGI RIQADQTPED

LDMEDNDIIE AHREQIGGGS ENLYFQSGRG KGGKGLGKGG AKRHRKVLRD NIQGITKPAI RRLARRGGVK RISGLIYEET RGVLKVFLEN VIRDAVTYTE HAKRKTVTAM DVVYALKRQG RTLYGFGG

SIRT1:

ADEAALALQ PGGSPSAAGA DREAASSPAG EPLRKRPRRD GPGLERSPGE PGGAAPEREV PAAARGCPGA AAAALWREAE AEAAAAGGEQ EAQATAAAGE GDNGPGLQGP SREPPLADNL YDEDDDDEGE EEEEAAAAAI GYRDNLLFGD EIITNGFHSC ESDEEDRASH ASSSDWTPRP RIGPYTFVQQ HLMIGTDPRT ILKDLLPETI PPPELDDMTL WQIVINILSE PPKRKKRKDI NTIEDAVKLL QECKKIIVLT GAGVSVSCGI PDFRSRDGIY ARLAVDFPDL PDPQAMFDIE YFRKDPRPFF KFAKEIYPGQ FQPSLCHKFI ALSDKEGKLL RNYTQNIDTL EQVAGIQRII QCHGSFATAS CLICKYKVDC EAVRGDIFNQ VVPRCPRCPA DEPLAIMKPE IVFFGENLPE QFHRAMKYDK DEVDLLIVIG SSLKVRPVAL IPSSIPHEVP QILINREPLP HLHFDVELLG DCDVIINELC HRLGGEYAKL CCNPVKLSEI TEKPPRTQKE LAYLSELPPT PLHVSEDSSS PERTSPPDSS VIVTLLDQAA KSNDDLDVSE SKGCMEEKPQ EVQTSRNVES IAEQMENPDL KNVGSSTGEK NERTSVAGTV RKCWPNRVAK EQISRRLDGN QYLFLPPNRY IFHGAEVYSD SEDDVLSSSS CGSNSDSGTC QSPSLEEPME DESEIEEFYN GLEDEPDVPE RAGGAGFGTD GDDQEAINEA ISVKQEVTDM NYPSNKS

Expression and Purification of sfGFP-N134X

For expression of sfGFP N134X, where X is BocK, AcK or AcdK, the following expression and purification protocol was followed. To express sfGFP N134BocK, the plasmid pEVOL mm PylRS (WT, R1-7, R2-23 and R3-11) was used together with pBAD sfGFP N134TAG to co-transform *E. coli* Top10 cells. The cells were grown in a 5 mL LB medium containing 100 µg/mL ampicillin (Amp) and 34 µg/mL chloramphenicol (Cm), overnight. The overnight culture was further used to inoculate into a 200 mL 2YT medium with the same concentration of Amp and Cm. Cells were grown in a 37°C shaker (250 r.p.m.) to OD_{600} 0.4 - 0.6. The protein expression was induced by the addition of 0.2% arabinose and 0.2 mM BocK as the final concentration. After 12 h expression, the cells were harvested (4k 20 min, 4°C), washed, and fully resuspended with a lysis buffer (300 mM NaCl, 50 mM NaH_2PO_4, 10 mM imidazole, pH 7.5). The sonication of cells underwent under ice/water incubation for 3 times (2 min each time with an interval of 5 min). Then the cell lysate was centrifuged (10k 40 min, 4°C), and the supernatant was collected. 1 mL Ni Sepharose^{TM} 6 Fast Flow column (GE Healthcare) was added to the supernatant and incubated for 1 h at 4°C. The mixture was loaded to a column and the flow-through was removed. 15 mL of wash buffer (300 mM NaCl, 50 mM NaH_2PO_4, 20 mM imidazole, pH 7.5) was used to wash the protein-bound resin in three batches followed by 5 mL of elution buffer (300 mM NaCl, 50 mM NaH_2PO_4, 250 mM imidazole, pH 7.5) to elute the target protein. All the eluted fractions were collected and concentrated by Amicon Ultra-15 Centrifugal Filter Device (10k MWCO, Millipore), and dialyzed in 20 mM phosphate pH 8.5 buffer 3 times. FPLC Q-Sepharose anion exchange column was

used to further purify the sfGFP N134BocK protein. The elution samples were collected and again concentrated by Amicon Ultra-15 Centrifugal Filter Device (10k MWCO, Millipore). For expression of sfGFP N134AcK, similar procedure was followed, except for addition of 5 mM nicotinamide and 5 mM AcK instead of BocK at the induction stage. For expression of sfGFP N134AcdK, 1 mM AcdK was added instead of BocK and same procedure was followed for purification. The protein expressions were analyzed on 12% SDS-PAGE and quantified using BCA Assay.

Expression and Purification of H3K23AcK

To express H3K23AcK protein, pEVOL mm AcKRS-PylT (WT or R3-11) and pETDuet-1 His6x-TEV-H3K23TAG were used to transform *E. coli* BL21 (DE3) cells. Cells were grown in 2YT media supplemented with Cm (34 μg/mL) and Amp (100 μg/mL), to OD_{600} 0.4−0.6, and the expression of H3K23AcK was induced with the addition of 5 mM N_{ε}-Acetyl-L-Lysine (Chem-Impex), 5 mM nicotinamide, 0.2% arabinose, and 1 mM IPTG for 12 h at 37 °C. Cells were then collected by centrifugation; resuspended in a lysis buffer containing 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 0.1% NaN_3, and 0.1% Triton X-100; and then lysed by sonication. The cell lysate was centrifuged, and the pellet was collected. The collected pellet was then dissolved in a urea buffer containing 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 6 M urea and cleared by centrifugation. The supernatant was then collected and further purified on Ni-NTA column (Ni-Sepharose 6 Fast Flow column from GE Healthcare). The column was washed with a wash buffer containing 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 20 mM imidazole, and

6 M urea and then eluted with a second buffer containing 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 300 mM imidazole, and 6 M urea. Following purification, the proteins were analyzed on 15% SDS-PAGE and quantified using BCA Assay.

Histone H4 Expression and Tetramer Refolding

For tetramer construction, histone H4 was also expressed. Fusion gene coding an N-terminal His-tagged SUMO-linked to histone H4 with a TEV protease cleavage site between them was cloned into a pETduet-1 plasmid to increase H4 expression yield as mentioned above. The plasmid was used to transform BL21 (DE3) cells. The transformed cells were then grown in 2YT media supplemented with Cm (34 μg/mL) and Amp (100 μg/mL), to OD_{600} 0.4−0.6 and then supplemented with 1 mM IPTG to induce expression for 4 h at 37 °C. For protein purification, the same protocol was used as mentioned for H3K23AcK expression above.

Histone His6x-SUMO-TEV-H4 protein was re-dissolved in 6M guanidinium chloride buffer (6 M guanidinium chloride, 20 mM Tris, 250 mM NaCl, pH 8.0) and mixed with Histone His-TEV-H3 K23AcK solution (in histone elution buffer: 50 mM Tris-HCl, 100 mM NaCl, 250 mM imidazole, 6 M Urea, pH 7.8) in a 1:1 molar ratio. Total protein concentration was adjusted to 2 μg/μl. His-SUMO-TEV-H4 and His-TEV-H3 K23AcK mixture were dialyzed sequentially at 4°C in 2 M NaCl buffer (2 M NaCl, 20 mM Tris, 1 mM EDTA, pH 7.5), 1 M NaCl (1 M NaCl, 20 mM Tris, 1 mM EDTA, pH 7.5) and 0.25 M NaCl buffer (250 mM NaCl, 20 mM Tris, 1 mM EDTA, pH 7.5). Suspension was

centrifuged (13000g 10 min, Eppendorf AG minispin) at 4°C and supernatant was collected as the tetramer product.

SIRT1 Deacetylation Assay

The histone H3 K23AcK/H4 tetramer concentration was adjusted to 0.4 μM by adding 0.25 M NaCl buffer (250 mM NaCl, 20 mM Tris, 1 mM EDTA, pH 7.5). NAD^+ (1mM), DTT (1 mM) and His-SIRT1 (0.2 μM) (expressed using published protocol[109]) were also added to tetramer reaction solutions. After being incubated at 37°C for 3 h, the reaction mixture was directly subjected to 18% SDS-PAGE gel and transferred to nitrocellulose membrane with standard semi dry Western blot protocol (Bio-Rad Trans-Blot Turbo transfer system). The membrane was coated with 5% fat-free milk (10 mL) for 2 h at room temperature and then treated with pan-acetylation antibody from PTM bio-lab (#PTM-101) overnight at 4°C (1: 2000, 5 mL). The membrane then was washed with PBST (PBS with 0.1% Tween-20, 10 mL) on the shaker six times with 2 min intervals. The membrane was then treated with secondary antibody (1: 10000, 5 mL) from Jackson Immuno Research (West Grove, PA) at room temperature for 1 h. The membrane was then washed with PBST on the shaker three times with 2 min intervals. The results were then visualized with Pierce ECL Western Blotting Substrate (#32106). Images were taken by ChemiDoc XRS^+ system from Bio-Rad (Hercules, CA).

AcdK Synthesis

The following synthesis scheme[110] (Figure II-5) was followed for synthesis of N_ϵ-(4-azidobenzoxycarbonyl)-δ,ϵ-dehydrolysine (AcdK) by Dr. Yadagiri Kurra:

To a 500-mL round bottom flask was added L-Glutamic acid (30 g, 0.2 mol), phthalic anhydride (29.6 g, 0.2 mol), triethylamine (2.79 mL, 0.02 mol) and toluene (250 mL). The mixture was heated to reflux and the water formed in the reaction was collected using a Dean-Stark apparatus. After 24 h, the reaction was cooled to room temperature and solvent was removed under reduced pressure. Ethyl acetate (EtOAc) was added to dilute the reaction and pH was adjusted to 2 with HCl (3.0 M). Then the mixture was extracted with EtOAc (300 mL × 3) and the organic layers were dried, concentrated. The resulting solid was washed with EtOAc and then filtered to collect the solid. Drying under high vacuum of the solid afforded the N-protected L-Glutamic acid **1** (22.5 g, 41% yield).

1 (22.5 g, 0.081 mol) was dissolved in MeOH (200 mL) and cooled to 0°C. $SOCl_2$ was added dropwise. The mixture was warmed to room temperature naturally and stirred overnight. All the volatiles were removed under reduced pressure and then diluted with EtOAc. The organic layer was washed with $NaHCO_3$ aqueous and brine. Then dried and concentrated to give methyl ester **2** (25.4 g, 100%).

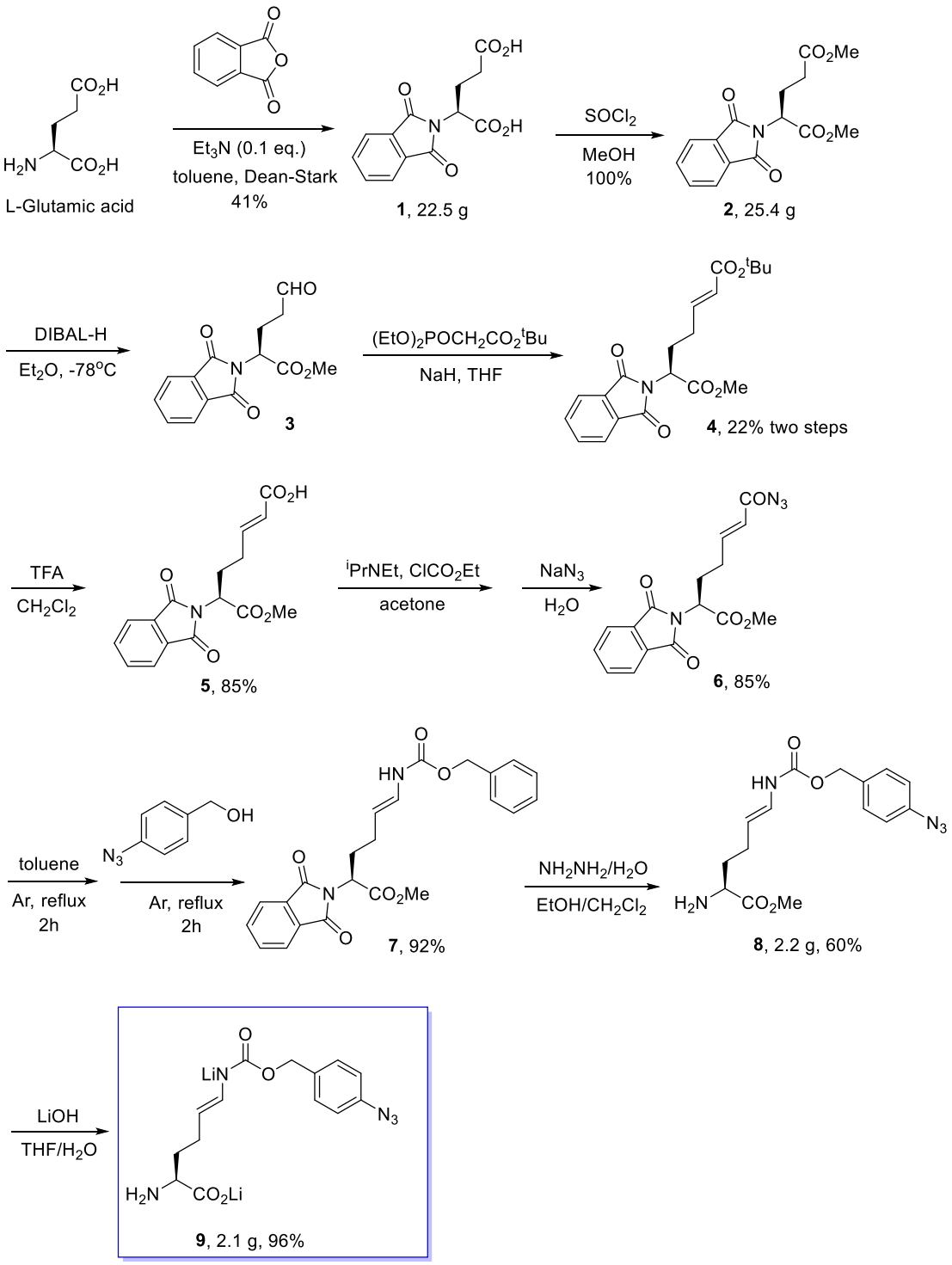


Figure II-5. Synthesis scheme of AcdK (Compound 9)[110]

2 (25.4 g, 0.083 mol) was dissolved in anhydrous diethyl ether (300 mL), cooled to -78^{o}C. DIBAL-H (99.8 mL, 1.0 M in hexane, 1.2 eq.) was added dropwise under Argon protection. Stirring continued for another 1h. Water (10 mL) was added, warmed to room temperature, and continue stirring for 1 h. The precipitate was removed through Celite and the filtrate was concentrated to produce the crude aldehyde **3**, which was used directly to the next step without further purification.

Anhydrous THF (80 mL) was added to a round-bottom flask, and cooled to 0^{o}C. NaH (1.9 g, 60% dispersion in mineral oil, 1.2 eq.) was added carefully. Then tert-butyl 2-(diethoxyphosphoryl) acetate (11.99 g, 1.2 eq.) was added dropwise. After bubble ceased to release, **3** (~ 0.0396 mol) in THF (20 mL) was added and the mixture was warmed to room temperature. After 2 h, NH_4Cl aqueous was added to quench the reaction. After diethyl ether extraction, the organic phase was dried and concentrated. The residue was purified via column chromatography with hexanes/EtOAc (3:1 v/v) as eluent to give product **4** (6.9 g, 22% yield, two steps)

4 (6.9 g, 0.0184788 mol) was dissolved in CH_2Cl_2 (70 mL), cooled to 0 ^{o}C. Trifluoroacetic acid (14.15 mL, 10 eq.) was added dropwise. Then the solution was warmed to room temperature and stirred for 1.5 h. Diluted with CH_2Cl_2, washed with H_2O, then dried and concentrated. The residue was purified via column chromatography with hexanes/ethyl acetate (1:1 v/v) as eluent to afford pure acid **5** (6.1 g, 85% yield).

5 (5.0 g, 0.015758 mol) was dissolved in acetone (150 mL), cooled to 0 ^{o}C and then added diisopropylethylamine (6.587 mL, 2.4 eq.). Ethyl chloroformate (3.3 mL, 2.2 eq.) was added dropwise and the solution was stirred at 0 ^{o}C for 1 h. Sodium azide (5.12

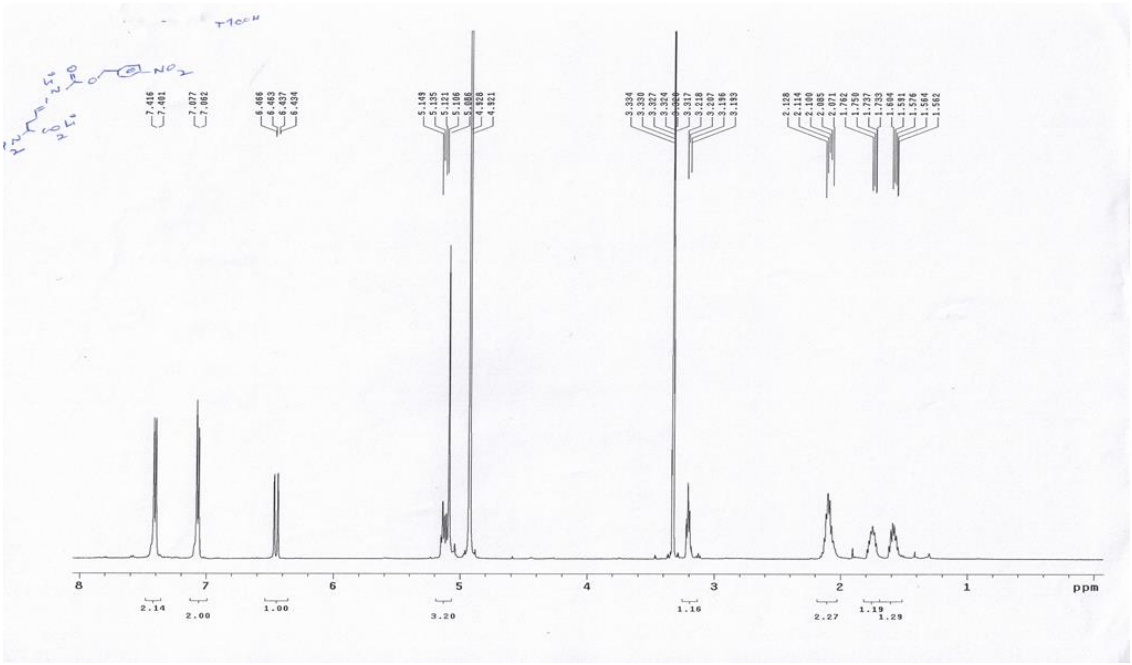
g, 5 eq.) in water (25 mL) was added and the reaction mixture was warmed to room temperature. After 1h with stirring, brine was added and extracted with ethyl acetate. The organic layer was dried and concentrated. The residue was purified via column chromatography with hexanes/ethyl acetate (3:1 v/v) as eluent to afford pure acid **6** (4.6 g, 85% yield).

6 (4.6 g, 0.01344 mol) was dissolved in toluene, and heated to reflux for 2 h under Argon protection. Then (2-azidophenyl) methanol (2 g, 1.0 eq.) was added and the mixture continued refluxing for another 2 h under Argon protection. Cooled to room temperature, and the solvent was removed under reduced pressure and the residue was purified via column chromatography with hexanes/ethyl acetate (3:1 to 1:1 v/v) as eluent to afford product **7** (5.1 g, 92% yield).

7 (5.1 g, 0.011 mol) was dissolved in ethanol/dichloromethane (16 mL/25 mL) and hydrazine monohydrate (0.85 mL, 1.6 eq.) was added. The mixture was stirred at room temperature overnight. The white solid formed in the reaction was filtered away and the filtrate was concentrated. The residue was purified via column chromatography with methanol/ammonium hydroxide/dichloromethane (15:50:450 v/v/v) as eluent to afford amine **8** (2.2 g, 60% yield).

8 (2.2 g, 0.0066 mol) was dissolved in THF/H_2O (25 mL/12.5 mL) and LiOH (0.3168 g, 2 eq.) was added. After stirring for 1 h, the undissolved impurities were filtered away and the filtrate was concentrated under vacuum to afford amino acid **9** (2.1 g, 96% yield). ^{1}H NMR (CD_3OD, 300 MHz) 7.40 (d, 2H, *J* = 4.5 Hz), 7.07 (d, 2H, *J* = 4.5 Hz), 6.43 (dd, 1H, *J* = 0.9, 8.7 Hz), 5.12 (dd, 1H, *J* = 4.5, 8.7 Hz), 5.08 (s, 2H), 3.20 (dd, 1H, *J*

= 0.9, 4.2 Hz), 2.12-2.07 (m, 2H), 1.76-1.73 (m, 1H), 1.60-1.56 (m, 1H). ^{13}C NMR (CD_3OD, 75 MHz) 181.4, 154.6, 139.8, 133.6, 129.3, 123.9, 118.6, 110.4, 65.5, 55.7, 36.0, 26.1.



[110]

Figure II-6. ^{1}H-NMR data of compound 9.

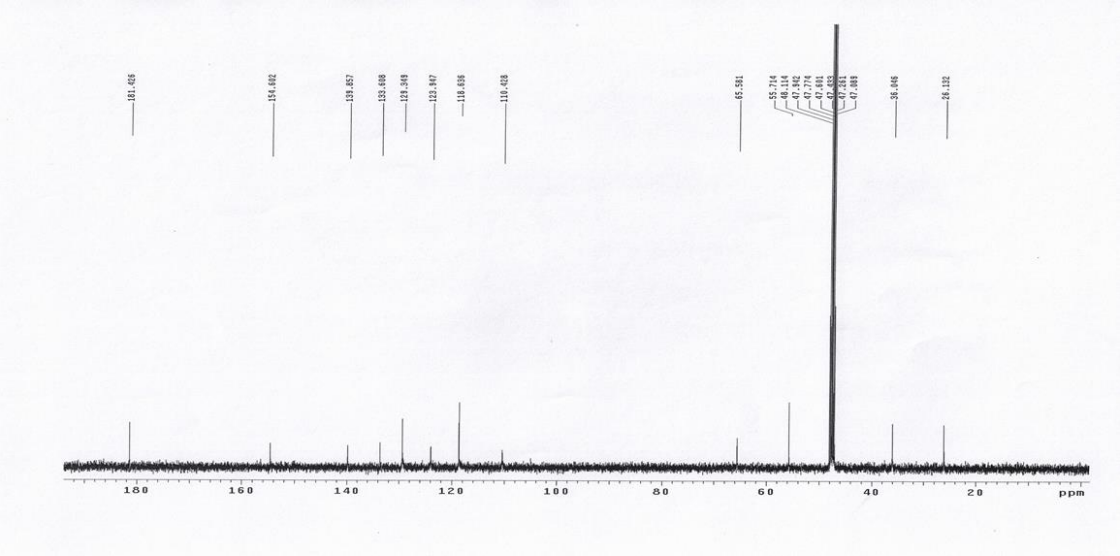


Figure II-7. ^{13}C-NMR data of compound 9. [110]

Results and Discussion

To target the N-terminal domain of the PylRS, a library of *Methanosarcina mazei* pyrrolysyl-synthetase was constructed with random mutations generated on the N-terminal domain of the synthetase. A plasmid pEVOL-PylRS derived from a pEVOL vector developed by Schultz *et. al.*[111] was constructed to contain genes coding $tRNA^{Pyl}$ under the control of the proK promoter and *M. mazei* PylRS under the control of the pBAD promoter. The PylRS gene was codon-optimized for expression in *E. coli*. Two restriction sites SpeI and BstXI were introduced around coding sites for the N-terminus methionine and residue S130, respectively. In order to create this library of *M. mazei* PylRS mutants, we utilized a technique known as error-prone PCR (EP-PCR)[108] which introduces random mutations in an amplified DNA sequence by taking advantage of *Taq* DNA polymerase's low fidelity and by modulating the Mg^{2+}, Mn^{2+} and dNTPs concentration in the PCR reaction. EP-PCR generated random mutations between M1 and S130 residues of the *M. mazei* PylRS. The mutants so generated were then co-transformed with super-folder green fluorescence (sfGFP) reporter protein with mutation at N134TAG and screened in presence of an ncAA, N_ϵ-(*tert*-butoxycarbonyl)-L-lysine (BocK) (Figure II-8.), for selecting the mutants with high ncAA incorporation efficiency (screening strategy shown in Figure II-9). A previous test indicated that *E. coli* Top10 cells transformed with the original pEVOL-PylRS and pBAD-sfGFP134TAG plasmids displayed visible but low green fluorescence when grown in the presence of 0.2 mM BocK. We expected that mutant PylRS clones more active than the wild-type would lead to higher fluorescence of expressed sfGFP, leading to more fluorescently visible colonies on plates and several

clones were identified using this strategy. All the mutants were sequenced and the results compared with that of wild-type PylRS sequence. The mutant showing the highest efficiency (R1-7 mm PylRS), from this first round of selection had a mutation at R19H and was chosen as a template for generating a second library using EP-PCR and it was subjected to the same screening strategy. The DNA sequencing results showed that the mutations converge in each round of selection.The mutant from the second round of screening had an additional mutation at H29R. The entire cloning and screening process was repeated three times to obtain the most efficient mm PylRS mutant (R3-11 mm PylRS) from the third round of selection. This clone has an additional mutation at T122S along with the mutations present on R19H and H29R. The sequencing results from each round of the convergent sequences are as summarized in Table II-2.

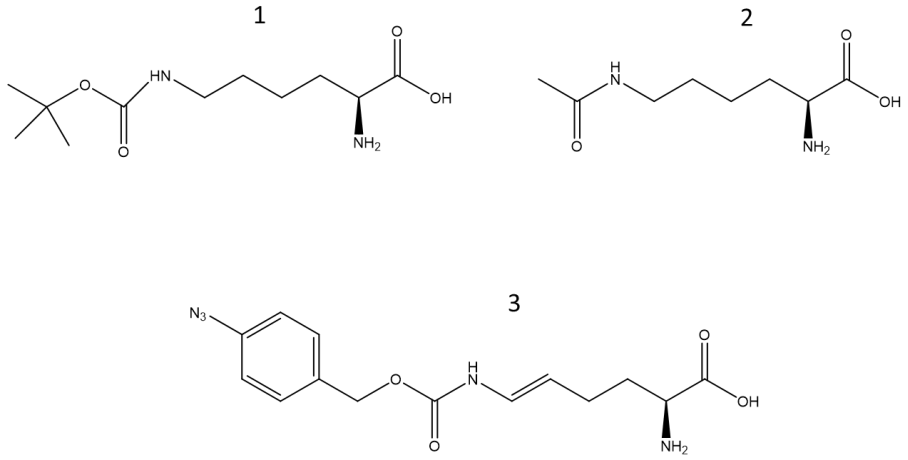


Figure II-8. ncAAs used in the study **1.** N$_\epsilon$-(*tert*-butoxycarbonyl)-L-lysine (BocK) **2.** N$_\epsilon$-acetyl-L-lysine (AcK) **3.** N$_\epsilon$-(4-azidobenzoxycarbonyl)-δ,ϵ-dehydrolysine (AcdK).

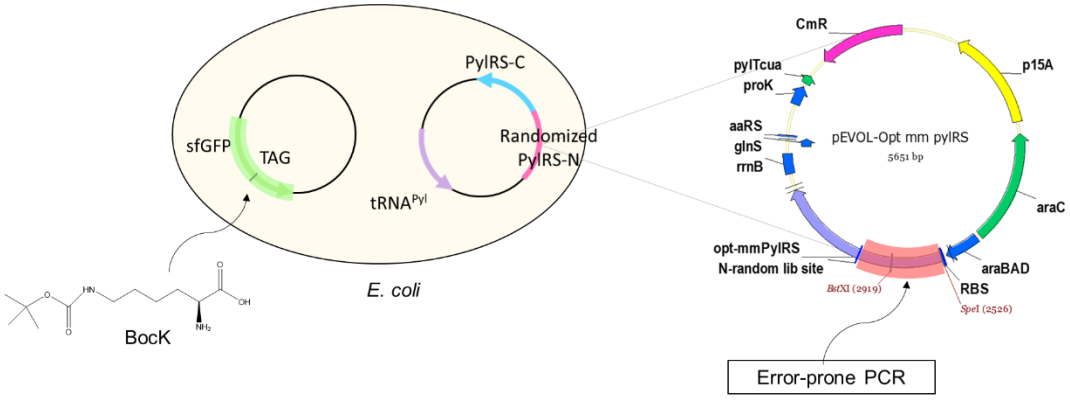


Figure II-9. Strategy used for screening efficient N-terminal domain modified PylRS mutants in library generated by error prone PCR.

Table II-2. Sequencing results of the best mutants identified in each round of selection.

Mutant Name		Mutations	
Wild-type mm PylRS[a]		--	
R1-7 mm PylRS[b]	R19H		
R2-23 mm PylRS[c]	R19H	H29R	
R3-11 mm PylRS[d]	R19H	H29R	T122S

[a] codon optimized sequence. [b] 1st round of selection. [c] 2nd round of selection. [d] 3rd round of selection.

The efficiency of all three mutant clones (R1-7, R2-23 and R3-11 mm PylRS) and wild type PylRS for the genetic incorporation of BocK was further tested and compared on a larger scale. *E. coli* Top10 cells containing both pBAD-sfGFP134TAG and either of the four pEVOL plasmids containing a wild type or mutant PylRS gene (R1-7, R2-23 or R3-11 mm PylRS) were grown in LB medium supplemented with 0.2 mM BocK and the expression of full-length sfGFP was induced by the addition of 0.2% arabinose. The expressed sfGFP was then affinity-purified using the Ni-NTA resin and analyzed by SDS-PAGE. Expression levels of sfGFP in four different cells were then quantified. As the results indicate, R3-11 mutant shows ~4 times increase in ncAA incorporation efficiency whereas mutants R1-2 and R2-23 show ~2 times and ~3 times increase in incorporation efficiency respectively. These results demonstrate our initial hypothesis of obtaining a better performing evolved PylRS clone by mutating N-terminal domain of PylRS that directly interacts with $tRNA^{Pyl}$. These results are summarized in Table II-3, Figure II-10.

These observations encouraged us to further test if the mutations generated on the N-terminal domain on the PylRS mutant obtained from the final round of selection (R3-11 mm PylRS) also increase the efficiency of the synthetase in incorporating other ncAAs. We tested this by replicating these mutations on the N-terminal domain of AcKRS[112, 113] which is a *M. mazei* PylRS derivative with a modified catalytic domain used for incorporating N_ϵ-acetyl-L-lysine (AcK) (Figure II-8.) in proteins. The incorporation efficiency of the mutant mm AcKRS (Opt AcKRS) was first compared with that of the original AcKRS (Non-opt AcKRS) for incorporation of 5 mM AcK in sfGFP N134TAG and the protein expression levels quantified (Table II-3, Figure II-11). The mutations were

also generated on another *M. mazei* PylRS, AcdKRS[110] which has been used by us to incorporate N_ϵ -(4-azidobenzoxycarbonyl)-δ,ϵ-dehydrolysine (AcdK) (Figure II-8.), a precursor for generating dimethyl lysine, in proteins. The optimized AcdKRS was compared with that of original AcdKRS in incorporating 1 mM AcdK in sfGFP N134TAG and the proteins were purified and the expression level quantified. The results are as summarized in Table II-3, Figure II-12.

The introduction of R19H, H29R and T122S mutations to these two synthetases is not expected to interrupt their recognition of corresponding ncAAs. On the contrary, improved incorporation of ncAAs was foreseen. As predicted, in comparison to a pEVOL-mm AcKRS plasmid (Non-opt), the newly constructed pEVOL R3-11 mm AcKRS plasmid led to five times increase in production level of sfGFP in the presence of 5 mM AcK. Although the results of sfGFP production for cells transformed with plasmids pEVOL R3-11 mm AcdKRS and pBAD-sfGFP134TAG were not dramatically higher than cells transformed with original pEVOL mm AcdKRS and pBAD-sfGFP134TAG for 1 mM AcdK, the yield is still improved significantly by about 20%.

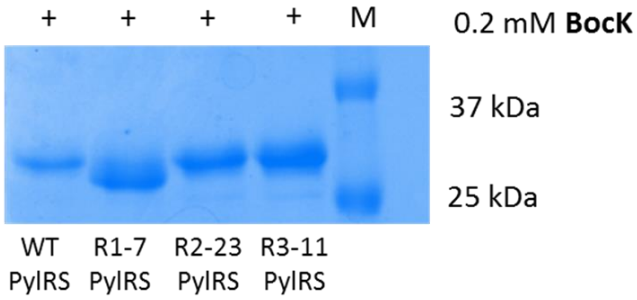


Figure II-10. Incorporation of 0.2 mM BocK in sfGFP N134TAG using WT mm PylRS, R1-7 mm PylRS, R2-23 mm PylRS and R3-11 mm PylRS.

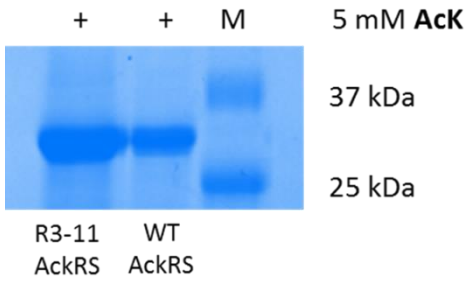


Figure II-11. Incorporation of 5 mM AcK in sfGFP N134TAG using WT mm AcdKRS and R3-11 mm AcKRS.

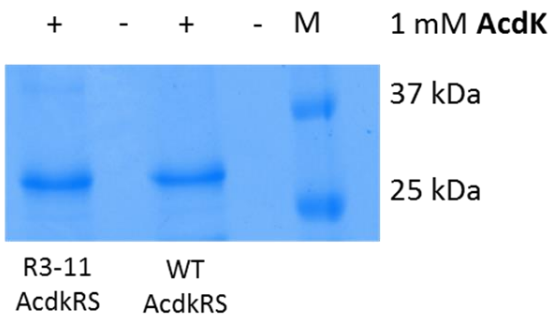


Figure II-12. Incorporation of 1 mM AcdK in sfGFP N134X using WT mm AcdKRS and R3-11 mm AcdKRS.

Table II-3. sfGFP N134X[a] protein expression yields for BocK, AcK and AcdK ncAAs.

Synthetase	Yield with BocK[b] (0.2mM) (mg/l)	Yield with AcK[c] (5mM) (mg/l)	Yield with AcdK[d] (5mM) (mg/l)
Wild-type synthetase	9.7	4.4	15.5
R3-11 synthetase	33.8	23.6	18.9

[a] X=ncAA. [b] synthetase used mm PylRS [c] synthetase used mm AcKRS [d] synthetase used mm AcdKRS

An important application of the genetic incorporation of AcK is to generate acetylated histones for functional investigation of histone acetylation in the regulation of chromatin.[114] Therefore, we then tested the improvement in incorporation efficiency of the R3-11 mm AcKRS synthetase in incorporating N_ϵ-acetyl-L-lysine in histone proteins. As explained earlier, histones undergo various post-translational modifications (PTMs) on their N-terminal tail which regulate several important cellular processes such as regulation of gene expression, cell repair and cell signaling, etc.[22] Acetylation of lysine on the N-terminal tail of histones is one of the key post-translational modification.[115] This PTM is regulated by a set of enzymes known as histone deacetylases (HDACs).[116] Our aim was to study the role of one such HDAC, SIRT1[117] to test its deacetylation activity on the H3K23AcK site.

To test this, TAG mutation was introduced in H3 histone at K23 position and the WT mm AcKRS and the pEVOL R3-11 mm AcKRS plasmid was coupled with the pETDuet H3 K23TAG plasmid. *E. coli* BL21 cells transformed with these two plasmids produced H3 K23AcK with a yield of 9.4 mg/l in LB medium supplemented with 5 mM AcK. A control experiment that used the original pEVOL mm AcKRS instead of pEVOL R3-11 mm AcKRS only provided a yield of 3.7 mg/l (Figure II-13). As is indicated by the results, the R3-11 mutant synthetase significantly improves the amount of acetylated histone protein obtained. The acetylated H3 K23 histone was then refolded with H4 histone to generate a tetramer using H4 histone expressed with SUMO tag for its increased production (Figure II-14 showing refolded tetramer with 1:1 H3 K23AcK and H4). This

tetramer complex was then subjected to a SIRT1-catalyzed deacetylation assay and then probed by a pan anti-acK antibody by Western blot analysis (Figure II-15).

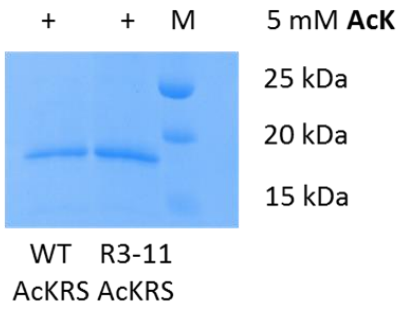


Figure II-13. Incorporation of 5 mM AcK in histone H3K23TAG using WT mm AcKRS (yield: 3.7 mg/l) and R3-11 mm AcKRS (yield: 9.4 mg/l).

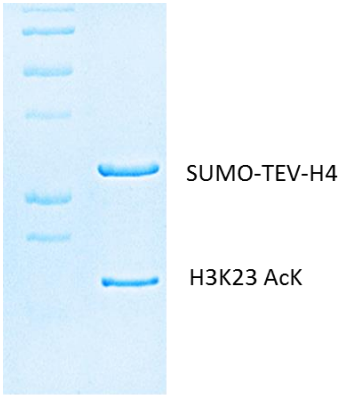


Figure II-14. Refolded acetylated histone tetramer of H2A, H2B, H3K23AcK and H4

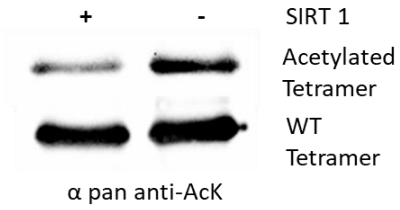


Figure II-15. SIRT1 deacetylation assay on WT tetramer and acetylated (H3 K23AcK) tetramer.

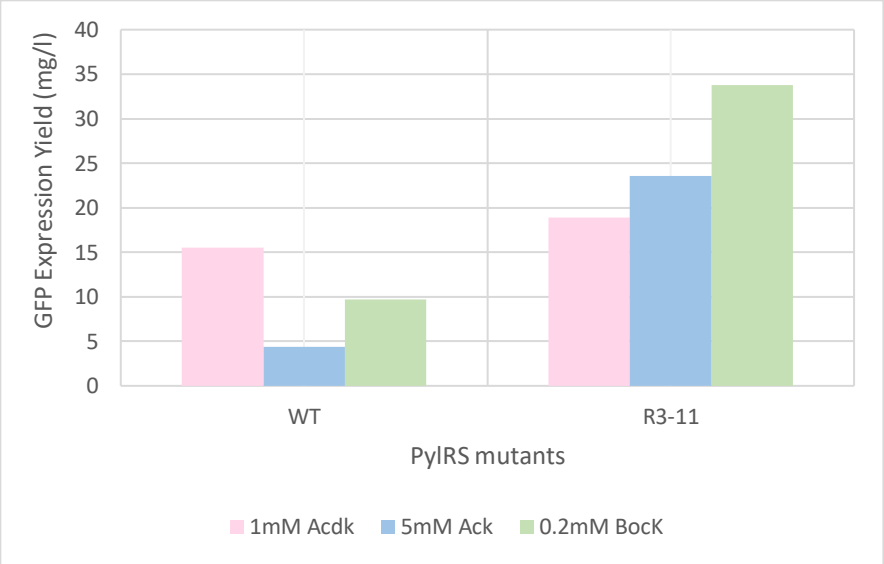


Figure II-16. Comparison of incorporation efficiency of WT PylRS synthetase and R3-11 PylRS synthetase for BocK, AcK and AcdK ncAAs.

Conclusion

In summary, as the results indicate (Figure II-16), the mutations identified increase the efficiency of the synthetase in incorporation of various ncAAs by a significant extent and thus, we have developed a general method to improve ncAA incorporation efficiency for the pyrrolysine incorporation system. As indicated by the results, by evolving the N-terminal region of *M. mazei* PylRS that directly interacts with $tRNA^{Pyl}$ but doesn't participate in the aminoacylation catalysis, we have shown that the three mutations identified, R19H, H29R and T122S can significantly improve the incorporation of BocK at amber codon. The direct transfer of these three mutations to two other PylRS mutants that were previously evolved for the genetic incorporation of AcK and AcdK also improved the incorporation efficiency of these two ncAAs, implying that these results can be extended to various synthetases to improve incorporation efficiency of several different ncAAs with the potential to solve the problem of low yield of expressed protein obtained. We hypothesize that this observation could either be a result of increased tRNA synthetase solubility or improved tRNA binding to the tRNA synthetase since it has been implicated in previous studies[92] that the role of N-terminal domain of archaeal PylRS is responsible for binding tRNA. Given the increasing adoption of the pyrrolysine system for basic and advanced research, we anticipate broad application of this current method reported by us.

CHAPTER III

GENETICALLY ENCODING HISTIDINE MIMETICS IN PROTEINS USING C-TERMINAL DOMAIN ENGINEERED PYRROLYSYL SYNTHETASE [*]

Introduction

Among the 20 canonical amino acids, histidine is unique in many ways. It is the only amino acid with an aromatic but hydrophilic side chain. As illustrated in Figure III-1, the imidazole side chain can attain two possible neutral forms, in which either of the two nitrogen atoms can serve as a hydrogen bond donor or a hydrogen bond acceptor. When protonated, the positive charge can oscillate between the two nitrogen atoms for initiating favorable charge-charge/dipole interactions in proteins. Due to this unique feature histidine has been found in many catalytic charge relay systems in enzymes, coordinating charge or proton transfer during catalysis, most commonly observed in catalytic triads and catalytic diads in proteases and other hydrolytic enzymes.[118, 119]

[*] Reprinted (adapted) with permission from "Probing the Catalytic Charge-Relay System in Alanine Racemase with Genetically Encoded Histidine Mimetics." by V. Sharma, Y.-S. Wang, W. R. Liu, *ACS Chemical Biology*, (**2016**), 11, 3305-3309. Copyright © 2016 American Chemical Society.

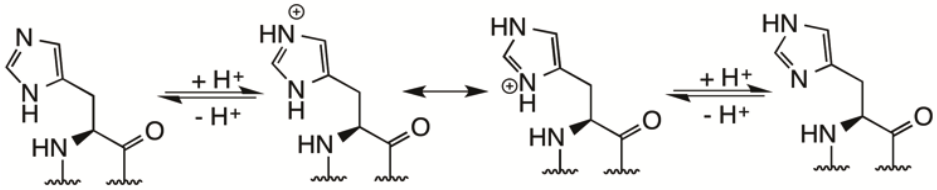


Figure III-1. Neutral and charged states of histidine in proteins.

To probe the role of histidine in enzymatic reactions, it is usually mutated to another residue such as alanine, asparagine or phenylalanine. Although useful for the investigation of histidine functions, none of these amino acids can accurately imitate the unique characteristics of histidine. Genetic incorporation of structurally similar histidine analogs can prove to be a very effective tool in studying biological roles of histidine residues in enzymes by fine tuning hydrogen bonding potentials of the histidine side chain. As demonstrated by Hsieh et al. (1979), to ascertain the role of a histidine residue in a hydrogen-bonding network in an enzyme, substitution of this residue by histidine derivatives so as to deprive the availability of proton on pyrrole (-NH) nitrogen or to block the proton acceptor site of pyridine (-N) nitrogen of histidine residue can greatly help in determining its function as a proton donor/acceptor or a general acid-base in the enzyme.[120] Besides genetic incorporation of histidine mimetics, techniques such as native chemical ligation (NCL)[32, 33, 96] or the use of auxotrophic bacterial strains[121, 122] can also be utilized for studying a system involving histidine mimetics. However, limitations such as the modifications possible[34] and the length of the peptide[47] that can be synthesized *via* NCL, toxicity of noncanonical amino acid (ncAA) being incorporated in cellular system and undesirable misincorporation *via* the use of auxotrophic strains[97] render these techniques inadequate for studying complex enzymatic systems, consequently directing the use of genetic incorporation of ncAAs as the preferred technique. In this chapter, we show we show that two close mimetics of histidine, 3-methyl-histidine (Mh) and thiazole alanine (Ta), can be genetically encoded using engineered pyrrolysine incorporation

machinery. We further show its application in probing the mechanism of action of an enzyme known as alanine racemase.

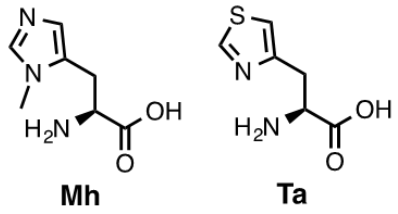


Figure III-2. Histidine derivatives 3-methyl histidine (Mh) and thiazole alanine (Ta) used in the study.

Experimental Details

General Experimental Procedure

E. coli Top10 were co-transformed with a pEVOL-pylT-PylRS(N346A/C348A) plasmid containing $tRNA^{Pyl}_{CUA}$ and two copies of PylRS N346A/C348A mutant (previously reported)[90] and a pBAD plasmid containing the gene encoding alanine racemase (AlaR). QuickChange® mutagenesis was performed to obtain three different alanine racemase mutants with the following mutations: H166TAG, H200TAG and H127TAG. The transformed cells were grown overnight at 37°C, 250 rpm, in auto-induction media[123] and supplemented with 2 mM of the ncAA to obtain six corresponding mutant enzymes.

To study the activity of the alanine racemase enzyme and its derivatives, the enzyme kinetic activity assay was employed. The assay consisted of coupled reactions between conversion of D-alanine to L-alanine and subsequent reaction of L-alanine with L-alanine dehydrogenase to form NADH which is detectable by its absorbance at 340 nm. All the reactions were carried out at room temperature in a buffer containing 50 µM PLP (Pyridoxal Phosphate), 0.1 M KCl, 0.1 M CHES (N-Cyclohexyl-2-aminoethanesulfonic acid) Buffer (pH 9.0), 10 mM NAD^+, and 0.1 – 10 mM D-alanine and 2 units/mL of L-alanine dehydrogenase.[124] The extinction coefficient of NADH used for calculations was 6220 M^{-1} cm^{-1}. The enzyme concentrations were 0.11 nM for WT, 0.09 µM for H166Ta, 0.16 µM for H166Mh, 0.17 µM for H200Ta, 0.12 µM for H200Mh, 0.17 µM for H127Ta and 0.37 µM for H127Mh. The kinetic parameters k_{cat} and k_{cat}/K_M values were then calculated.

Primer List

AlaR- KpnI-F, 5’- GCC GGT GGT ACC TCA GTG GTG GTG -3’

AlaR- NcoI-R, 5’- GCC TGG CCA TGG ATG AAC GAC TTT CAT C -3’

Del-pBAD-alaR-F, 5’- ATG AAC GAC TTT CAT CGC GAT ACG TGG GCG -3’

Del-pBAD-alaR-R, 5’- GGT TAA TTC CTC CTG TTA GCC CAA AAA ACG GG -3’

His 125A-R, 5’ - CGC AAT AGG AAA AGG GCC -3’

His 125-TAG-F, 5’ - TTC CAT TTG AAA ATG GAC ACC -3’

His 125-TAG-R, 5’ - CTA AAT AGG AAA AGG GCC GCT -3’

His 127A-R, 5’ - CGC GAA ATG AAT AGG AAA AGG GCC GCT G -3’

His 127-QC-F, 5’- TTG AAA ATG GAC ACC GGC ATG GGA CG -3’

His 127-QC-R, 5’- CTA GAA ATG AAT AGG AAA AGG GCC GCT G -3’

H166A-QC-R, 5’ - CGC CGT GTA CAC CCC TTC AAG CAC AA -3’

H166F-QC-R, 5’ - AAA CGT GTA CAC CCC TTC AAG CAC AA -3’

His 166-QC-F, 5’- TTT GCG ACT GCG GAT GAG GTG AAC AC -3’

His 166-QC-R, 5’- CTA CGT GTA CAC CCC TTC AAG CAC AA -3’

His 200A-R, 5’ - CGC GAC GAG CGG CGG GCG CG -3’

His 200-QC-F, 5’- TGC GCC AAC AGC GCA GCG TCG CTC CG -3’

His 200-QC-R, 5’- CTA GAC GAG CGG CGG GCG CGA CGG CAG -3’

His 200A-R, 5’ - CGC GAC GAG CGG CGG GCG CG -3’

Plasmid Construction

For construction of the plasmid for expression of alanine racemase, the AlaR gene was amplified from pET23a-wtAlaR by flanking primers, AlaR- KpnI-F and AlaR- NcoI-R. The gene product was digested by the restriction enzymes *KpnI* and *NcoI*, purified and ligated into pBad vector digested by *KpnI* and *NcoI* restriction enzymes to generate plasmid pBAD-AlaRwt-His6x as shown in Figure III-3. The *NcoI* restriction digestion site was then removed from the pBAD vector by primers Del-pBAD-alaR-F and Del-pBAD-alaR-R as listed above.

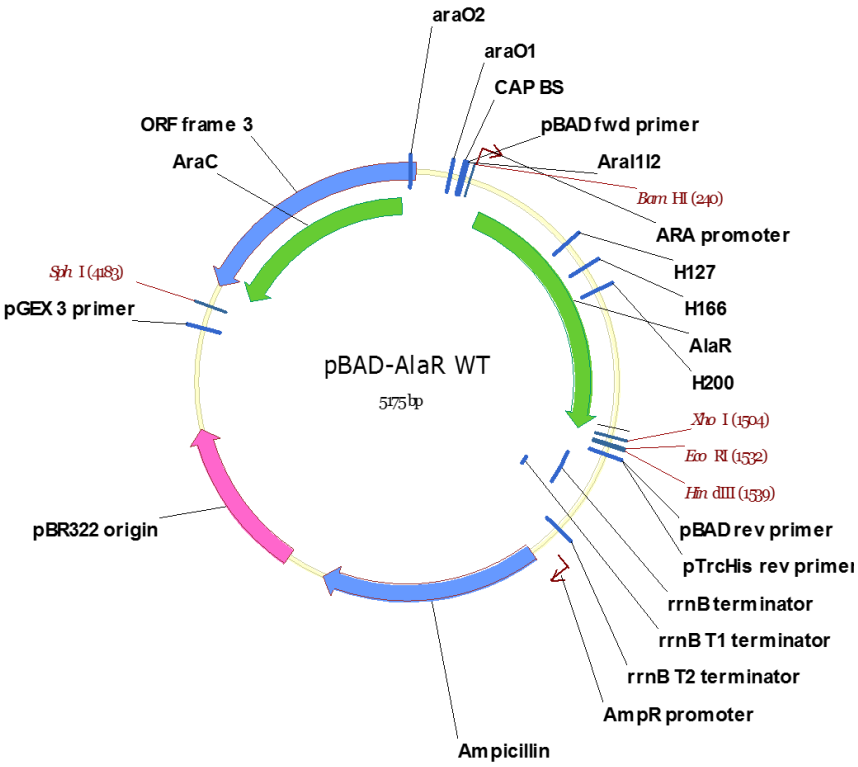


Figure III-3. Map showing pBAD AlaRwt-His6x plasmid.

The plasmids (i) pBAD-AlaR H125A-His6x, (ii) pBAD-AlaR H125TAG-His6x, (iii) AlaR H127TAG-His6x, (iv) pBAD-AlaR H166A-His6x, (v) AlaR H166F-His6x, (vi) AlaR H166TAG-His6x, (vii) pBAD-AlaR H200TAG-His6x and (viii) pBAD-AlaR H127A H166A H200A-His6x were all derived from the plasmid pBAD-AlaRwt-His6x by introducing mutations by site-directed mutagenesis using the following primers respectively: (i) His 125-TAG-F and His 125A-R, (ii) His 125-TAG-F and His 125-TAG-R, (iii) His 127-QC-F and His 127-QC-R (iv) His 166-QC-F and H166A-QC-R (v) His 166-QC-F and H166F-QC-R, (vi) His 166-QC-F and His 166-QC-R, (vii) His 200-QC-F and His 200-QC-R, (viii) His 127-QC-F and His 127A-R, His 166-QC-F and H166A-QC-R & His 200-QC-F and His 200A-R.

All the plasmid sequences were confirmed by DNA sequencing. All oligonucleotide primers were purchased from Integrated DNA Technologies, Inc.

DNA Sequences

B. stearothermophilus Alanine Racemase:

atgaacgactttcatcgcgatacgtgggcggaagtggatttggacgccatttacgacaatgtggcgaatttgcgccgt
ttgctgccggacgacacgcacattatggcggtcgtgaaggcgaacgcctatggacatggggatgtgcaggtggcaaggaca
gcgctcgaagcgggggcctcccgcctggcggttgccttttggatgaggcgctcgctttaagggaaaaaggaatcgaagcgc
cgattctagttctcggggcttcccgtccagctgatgcggcgctggccgcccagcagcgcattgccctgaccgtgttccgctccg
actggttggaagaagcgtccgccctttacagcggccctttcctattcatttc**cat**ttgaaaatggacaccggcatgggacggctt
ggagtgaaagacgaggaagagacgaaacgaatcgtagcgctgattgagcgccatccgcattttgtgcttgaaggggtgtaca

cgcatttgcgactgcggatgaggtgaacaccgattatttttcctatcagtatacccgtttttgcacatgctcgaatggctgccgtc

gcgcccgccgctcgtccattgcgccaacagcgcagcgtcgctccgtttccctgaccggacgttcaatatggtccgcttcggcat

tgccatgtatgggcttgccccgtcgcccggcatcaagccgctgctgccgtatccattaaaagaagcattttcgctccatagccgc

ctcgtacacgtcaaaaaactgcaaccaggcgaaaaggtgagctatggtgcgacgtacactgcgcagacggaggagtggatc

gggacgattccgatcggctatgcggacggcgtccgccgcctgcagcactttcatgtccttgttgacggacaaaaggcgccgat

tgtcggccgcatttgcatggaccagtgcatgatccgcctgcctggtccgctgccggtcggcacgaaggtgacactgattggtc

gccaaggggacgaggtaatttccattgatgatgtcgctcgccatttggaaacgatcaactacgaagtgccttgcacgatcagtta

tcgagtgccccgtatttttttccgccataagcgtataatggaagtgagaaacgccattggccgcggggaaagcagtgcacacca

ccaccaccaccactga

H127

H166

H200

sfGFP S2TAG:

atggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaatttag

cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttcc

gtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttta

aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtt

aaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaaaatggcaacattctgggtcataaact

ggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcaca

acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggat

aatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgctaccg

ccgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattaa

Protein Sequences

B. stearothermophilus Alanine Racemase:

MNDFHRDTWA EVDLDAIYDN VANLRRLLPD DTHIMAVVKA NAYGHGDVQV ARTALEAGAS RLAVAFLDEA LALREKGIEA PILVLGASRP ADAALAAQQR IALTVFRSDW LEEASALYSG PFPIHFHLKM DTGMGRLGVK DEEETKRIVA LIERHPHFVL EGVYTHFATA DEVNTDYFSY QYTRFLHMLE WLPSRPPLVH CANSAASLRF PDRTFNMVRF GIAMYGLAPS PGIKPLLPYP LKEAFSLHSR LVHVKKLQPG EKVSYGATYT AQTEEWIGTI PIGYADGVRR LQHFHVLVDG QKAPIVGRIC MDQCMIRLPG PLPVGTKVTL IGRQGDEVIS IDDVARHLET INYEVPCTIS YRVPRIFFRH KRIMEVRNAI GRGESSAHHH HHH

sfGFP S2X:

MVXKGEELFT GVVPILVELD GDVNGHKFSV RGEGEGDATN GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERTIS FKDDGTYKTR AEVKFEGDTL VNRIELKGID FKENGNILGH KLEYNFNSHN VYITADKQKN GIKANFKIRH NVEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSVLSK DPNEKRDHMV LLEFATAAGI THGMDELYKG SHHHHHH

X represents ncAA

Expression and Purification of sfGFP-S2X

For expression of sfGFP S2X, where X is Mh or Ta, the following expression and purification protocol was followed. *E. coli* Top10 cells were co-transformed with a pEVOL-pylT-PylRS(N346A/C348A) plasmid containing $tRNA_{CUA}^{Pyl}$ and two copies of PylRS N346A/C348A mutant and pBAD sfGFP S2TAG plasmid. The transformed cells were grown overnight at 37°C, 250 rpm, in minimal media and supplemented with 2 mM of the ncAA. The culture was inoculated in the auto-induction media (50 mL) containing 100 mg/mL ampicillin and 34 mg/mL chloramphenicol and Mh or Ta were added to final concentration of 2 mM after growing the cells in the media for 30 min. The protein expression was induced when the culture reached OD_{600} of 0.5 by adding arabinose to a final concentration of 0.2 %. After 24 h expression, the cells were harvested (4k 20 min, 4°C), washed, and fully resuspended with a lysis buffer (300 mM NaCl, 50 mM NaH_2PO_4, 10 mM imidazole, pH 7.5). The sonication of cells underwent under ice/water incubation for 3 times (2 min each time with an interval of 5 min). Then the cell lysate was centrifuged (10k 40 min, 4°C), and the supernatant was collected. 1 mL Ni Sepharose™ 6 Fast Flow column (GE Healthcare) was added to the supernatant and incubated for 1 h at 4°C. The mixture was loaded to a column and the flow-through was removed. 15 mL of wash buffer (300 mM NaCl, 50 mM NaH_2PO_4, 20 mM imidazole, pH 7.5) was used to wash the protein-bound resin in three batches followed by 5 mL of elution buffer (300 mM NaCl, 50 mM NaH_2PO_4, 250 mM imidazole, pH 7.5) to elute the target protein. All the eluted fractions were collected and concentrated by Amicon Ultra-15 Centrifugal Filter Device (10k MWCO, Millipore), and dialyzed in 20 mM phosphate pH 8.5 buffer 3 times. FPLC

Q-Sepharose anion exchange column was used to further purify the sfGFP S2X protein. The elution samples were collected and again concentrated by Amicon Ultra-15 Centrifugal Filter Device (10k MWCO, Millipore). The protein expressions were analyzed on 12% SDS-PAGE and quantified using BCA Assay.

Expression and Purification of alaR mutants

For expression of alanine racemase mutant proteins, *E. coli* Top10 cells were co-transformed with a pEVOL-pylT-PylRS(N346A/C348A) plasmid containing $tRNA^{Pyl}_{CUA}$ and two copies of PylRS N346A/C348A mutant[90] and corresponding pBAD-alaRHxxx-His6x plasmid with TAG mutation at the required histidine position. The transformed cells were grown overnight at 37°C, 250 rpm, in auto-induction media (with tyrosine, cysteine, histidine, phenylalanine and tryptophan left out of the media)[123] and supplemented with 2 mM of the ncAA. The culture was inoculated in the auto-induction media (50 mL) containing 100 mg/mL ampicillin and 34 mg/mL chloramphenicol and Mh or Ta were added to final concentration of 2 mM after growing the cells in the media for 30 min. The protein expression was induced when the culture reached OD_{600} of 0.5 by adding arabinose to a final concentration of 0.2 %. After 24 h induction, cells were harvested, resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) supplemented with 20 µM pyridoxal-5'-phosphate (PLP) and sonicated. The cell lysate was clarified by centrifugation (30 min, 11,000 g, 4ºC). The supernatant was injected into a 30 mL Ni^{2+}-NTA column (Qiagen) and washed with 60 to 90 mL lysis buffer. Protein was finally eluted out by running an imidazole gradient from 10 mM to 250 mM in lysis buffer. Pure

fractions were collected and concentrated. The eluted protein was dialyzed against 1X PBS buffer with 20 µM PLP concentrated using an Amicon Ultra-15 Centrifugal Filter (10,000 MWCO). The purified proteins were analyzed by 12% SDS-PAGE. AlaRHxxx-His6X mutant proteins incorporated with Mh and Ta were expressed and purified with the yields as shown in Figure III-4.

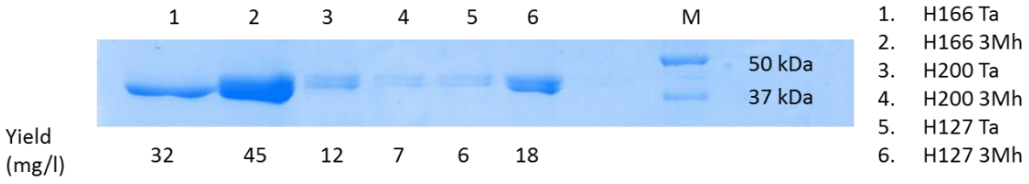
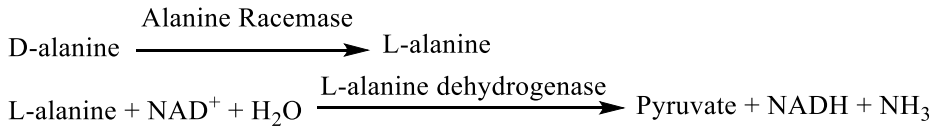


Figure III-4. Expression of alanine racemase mutants supplied with 2mM ncAA.

Kinetic Assay for studying enzymatic activity



The kinetic assay employed to study the enzymatic activity of WT alanine racemase and the mutant enzymes consisted of coupled reactions between conversion of D-alanine to L-alanine and subsequent reaction of L-alanine with L-alanine dehydrogenase to form NADH which is detectable by its absorbance at 340 nm. All the reactions were carried out at room temperature in a buffer containing 50 µM PLP (Pyridoxal Phosphate), 0.1 M KCl, 0.1 M CHES (N-Cyclohexyl-2-aminoethanesulfonic acid) Buffer (pH 9.0), 10 mM NAD^+, and 0.1 – 10 mM D-alanine and 2 units/mL of L-alanine dehydrogenase.[124] The extinction coefficient of NADH used for calculations was 6220 M^{-1} cm^{-1}. The enzyme concentrations were 0.11 nM for WT, 0.09 µM for H166-Ta, 0.16 µM for H166-Mh, 0.17 µM for H200-Ta, 0.12 µM for H200-Mh, 0.17 µM for H127-Ta and 0.37 µM for H127-Mh. All the measurements were made at 340 nm on UV-1800 UV-Spectrophotometer, Shimadzu®.

Circular Dichroism Analysis

All the mutant protein samples were further purified on FPLC (Bio-Rad, NGC™ Chromatography System) and dialyzed in 20 mM borate buffer (pH 9.1) to prepare their samples for measurement. Circular dichroism (CD) measurements were made on AVIV-62DS spectrometer using quartz cuvette of 1 mM path-length. CD spectra was recorded

at 25°C in 20 mM borate buffer (pH 9.1) purified by filtration before use. Data was collected from 190 nm to 280 nm in increments of 1 nm with averaging time of 10s. Concentrations of all protein samples were normalized to 0.022 mg/mL for helicity calculations. Fraction helicity (f_h) was calculated using the ellipticity data at 222 nm obtained according to the method described by Rohl et al. (1997).[125] The following formula was used for the calculations: $f_h = \frac{\theta_{222\,nm}(obs) - \theta_{222\,nm}(coil)}{\theta_{222\,nm}(\alpha\text{- helix}) - \theta_{222\,nm}(coil)}$; $= \frac{\theta_{222} - \theta_C}{\theta_H - \theta_C}$. In this formula, θ = mean residue ellipticity, θ_C = 2220 – 53 * T (in °C) = 2220 – 53 * 25= 895 deg cm^2 d mol^{-1}, θ_H = (-44000+ 250 * T (in °C) $\left[1 - \frac{3}{N_{res}}\right]$ = (-44000 + 250 * 25) (1-$\frac{3}{393}$) = -37461.8 deg cm^2 d mol^{-1} and N_{res} = Number of residues. The calculated % helicity of mutant samples as compared to wild-type alanine racemase (22.7 %) are listed in the Table II-1 below:

Table III-1. Calculated % helicity of alanine racemase and its mutants.

Enzyme	% Helicity
WT	22.7
H166Mh	22.9
H166Ta	20.4
H200Mh	21.6
H200Ta	19.9
H127Mh	21.7
H127Ta	20.7

Results and Discussion

In order to create a mutagenic system that genetically incorporates ncAAs that closely mimic histidine for fine tuning of histidine-involved hydrogen bonding interactions, we referred to the engineering of the pyrrolysine incorporation machinery.[126] We set our sights on the incorporation of histidine analogs such as 3-methyl-histidine (Mh) and thiazole alanine (Ta), as both are close histidine mimetics. We have previously shown that a rationally designed pyrrolysyl-tRNA synthetase variant with the mutations N346A/C348A (PylRS(N346A/C348A)) can recognize a large number of phenylalanine derivatives and, together with its cognate $tRNA_{CUA}^{Pyl}$, mediate their incorporation into proteins at amber (TAG) codons in *Escherichia coli* and mammalian cells.[100, 101] Given the similarity between some phenylalanine derivatives and histidine analogs we predicted that this same enzyme can also recognize these ncAAs and direct their incorporation into proteins at amber codons. To test this proposition, *E. coli* BL21 cells coding PylRS(N346A/C348A), $tRNA_{CUA}^{Pyl}$, and superfolder green fluorescent protein (sfGFP) with an amber mutation at the S2 position were grown in minimal media supplemented with 5 mM Mh or Ta. For both ncAAs, their presence in the growth media promoted the suppression of the amber mutation in sfGFP and the subsequent full-length protein expression. Without providing a ncAA, no full-length sfGFP was expressed (Figure III-5).

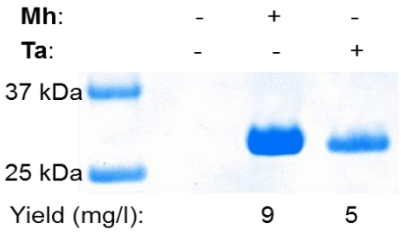


Figure III-5. The selective incorporation of Mh and Ta into sfGFP at its S2 position.

The electrospray ionization mass spectrometry (ESI-MS) analysis of the purified proteins detected molecular weights of 27,732 Da for the Mh-incorporated protein and 27,735 Da for the Ta-incorporated protein that agree well with the theoretical molecular weights (27,733 and 27,736 Da, respectively), confirming the selective incorporation of these two ncAAs and the high fidelity of the incorporation system (Figure III-6 and Figure III-7). Mh has also been genetically encoded using an alternative system.[127]

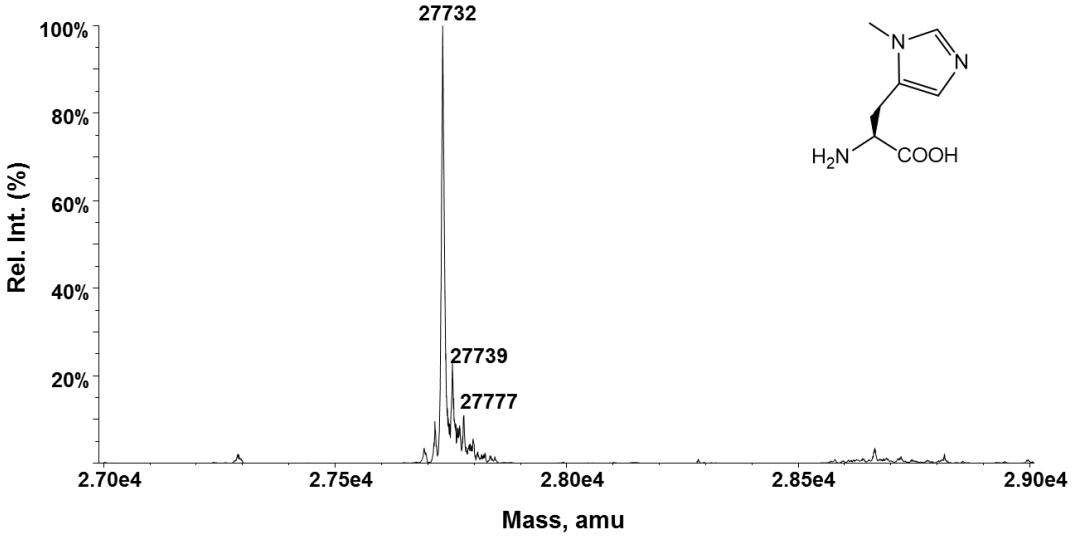


Figure III-6. ESI-MS analysis of sfGFP S2Mh (theoretical molecular weight: 27, 733 Da)

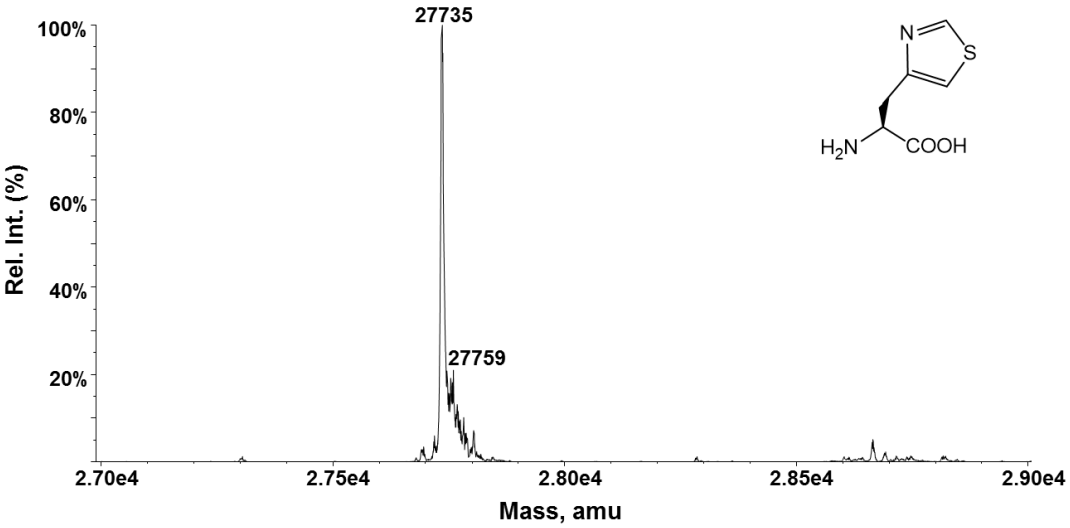


Figure III-7. ESI-MS analysis of sfGFP S2Ta (theoretical molecular weight: 27, 736 Da)

We envisioned that the incorporation of both Mh and Ta can be applied to probe fundamental roles of histidine in charge-relay systems of enzymes. Instead of probing well studied catalytic triads and diads, we chose to investigate histidine functions in a potentially extended charge-relay system in alanine racemase (EC 5.1.1.1). Alanine racemase has become a promising drug target since the discovery of its key role in the formation of peptidoglycan in bacterial cell walls.[128-132] Mechanistic illustration of this enzyme may facilitate the identification of inhibitors[112, 133-138] that can serve as novel antibiotics. This Pyridoxal 5'-Phosphate (PLP)-dependent enzyme catalyzes the interconversion of L-alanine to D-alanine.[139-141] Although it is one of the most well studied enzymes, information on several aspects of its mechanism remains elusive. It is known that alanine racemase employs a two-base catalytic mechanism in which K39 and Y265 residues act as general acid/base catalysts[142] (Figure III-8).

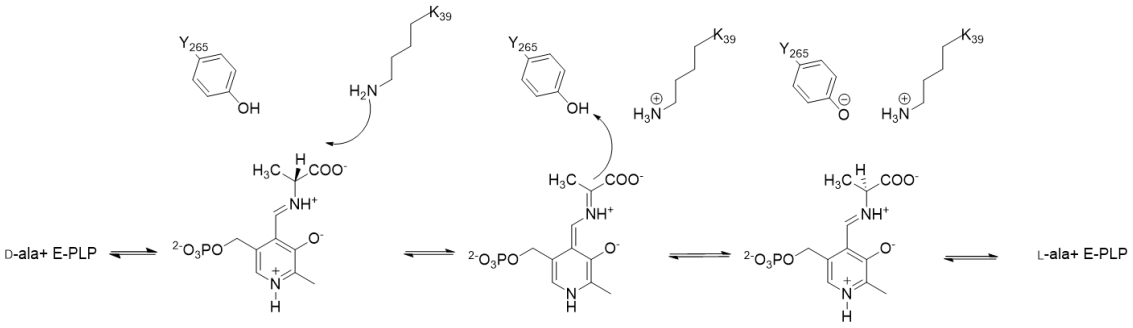


Figure III-8. The catalytic mechanism of alanine racemase.

In the racemization reaction from D-alanine to L-alanine, D-alanine first reacts with PLP to form an external aldimine. K39 then abstracts the α proton from this external aldimine to generate a quinonoid intermediate that receives a proton from Y265 to form a second external aldimine. L-alanine is then released from this external aldimine as the final product. The whole catalytic process is fully reversible. pH profile studies have indicated that the phenolic hydroxide of Y265 has a pK_a value around 7.1-7.4, contributing to a very high catalytic rate of wild-type alanine racemase with a k_{cat}/K_M of 1.6 (0.2)×10^6 $M^{-1}s^{-1}$.[124] We predicted the existence of an extended charge-relay system and hydrogen bonding network involving residues E161, H127, H200, R219, H166, and Y265 in alanine racemase that are potentially involved in bringing down the pK_a of the Y265 side chain hence affecting the catalytic activity of the enzyme (Figure III-9). E161 is solvent exposed as determined by the active site structure of alanine racemase from *Bacillus stearothermophilus* (PDB: 1SFT).[143]

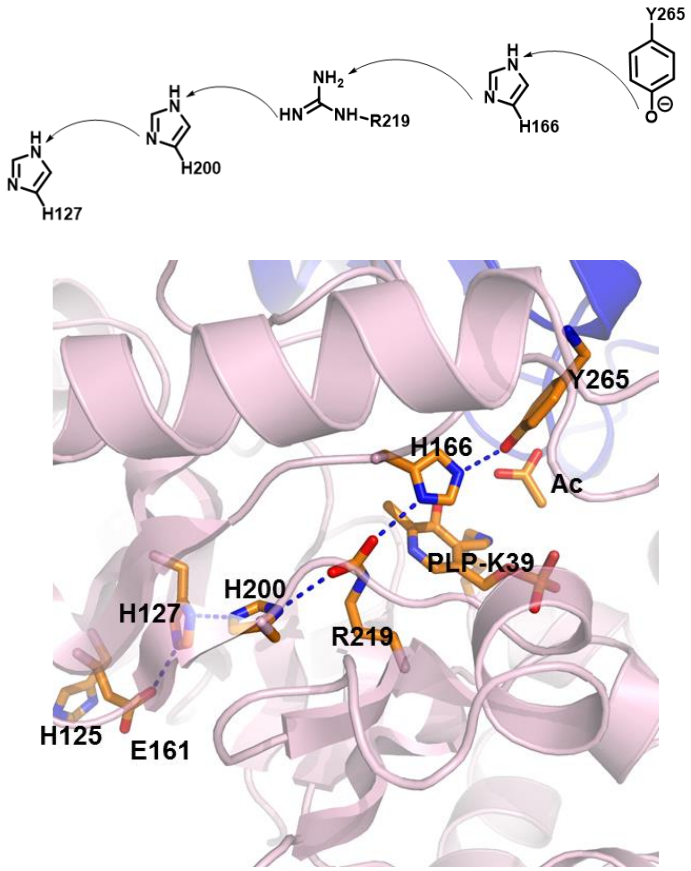


Figure III-9. The active site structure of *B. stearothermophilus* alanine racemase (PDB:1SFT) showing predicted charge relay system.

There are three histidine residues involved in this proposed charge-relay system. It has been shown that H166 coordinates the activation of Y265 by R219.[124] Although H127 and H200 are distal from Y265, they potentially serve critical roles in the activation of Y265 through this charge-relay system with the extended hydrogen bonding network. To test this proposition and directly observe how hydrogen bonding interactions contribute to the final activation of Y265, we decided to substitute these two residues in addition to H166 with either Mh or Ta using our established system for the incorporation of these two ncAAs. In total, six alanine racemase mutants containing these ncAAs were expressed and purified to homogeneity. Circular dichroism analysis of purified proteins indicates that they all fold like the wild-type enzyme (Figure III-10). This is an expected observation given that both Mh and Ta are highly analogous to histidine with Ta being almost spatially identical to histidine. Two mutants H166Ta and H166Mh were also characterized by the mass spectrometry analysis to confirm the incorporation of Ta and Mh (Figure III-11 and Figure III-12). Catalytic activity of these mutant enzymes along with wild-type alanine racemase enzyme was characterized according to the methods of Sun et al.[124] and the determined values are presented in Table III-2.

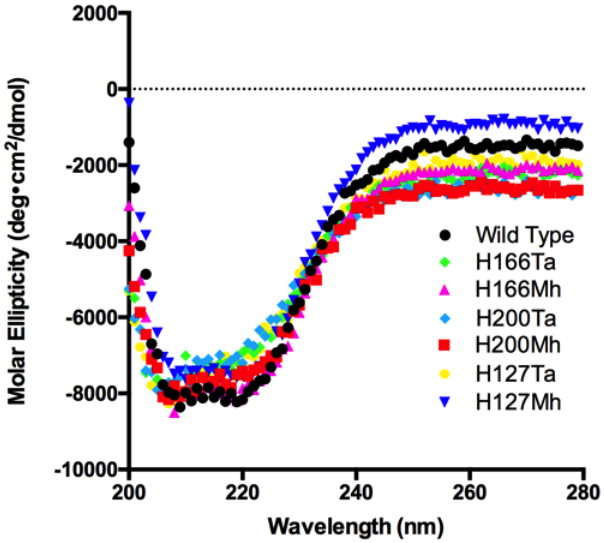


Figure III-10. Circular dichroism spectra of wild-type alanine racemase and its six mutants.

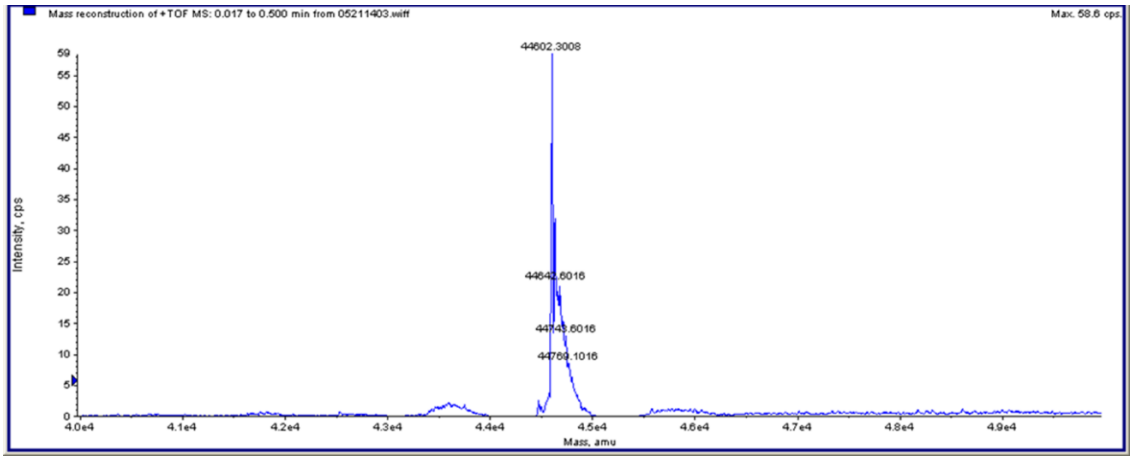


Figure III-11. ESI-MS of alanine racemase H166-Mh mutant. Theoretical molecular weight is 44,601 Da which corresponds well with the major peak observed in the spectra of 44,602 Da.

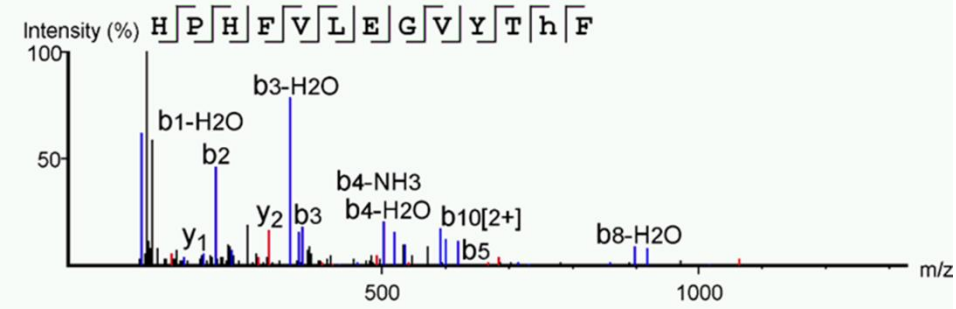


Figure III-12. MS/MS analysis of alanine racemase H166-Ta mutant. The b1 ion indicates presence of thiazole alanine at the 166 position of the alanine racemase mutant sample.

Table III-2. Catalytic parameters of wild-type alanine racemase and its mutants

Enzyme	k_{cat} (s^{-1})	Rel. k_{cat}	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	Rel. k_{cat}/K_M
wild type	$3.5\ (0.1) \times 10^{3}$	1	2.2	$1.6\ (0.2) \times 10^{6}$	1
H166Ta	$1.7\ (0.1) \times 10^{1}$	0.005	0.5	$3.3\ (0.2) \times 10^{4}$	0.02
H166Mh	$8.4\ (0.3) \times 10^{0}$	0.002	0.5	$1.8\ (0.2) \times 10^{4}$	0.01
H200Ta	$1.2\ (0.5) \times 10^{1}$	0.003	1.6	$7.3(0.8) \times 10^{3}$	0.005
H200Mh	$5.2\ (0.2) \times 10^{0}$	0.002	1.6	$3.2\ (0.4) \times 10^{3}$	0.002
H127Ta	$1.2\ (0.4) \times 10^{1}$	0.003	1.2	$1.0\ (0.1) \times 10^{4}$	0.006
H127Mh	$2.2\ (0.1) \times 10^{1}$	0.006	1.8	$1.2\ (0.2) \times 10^{4}$	0.007
H125Ta	$4.1\ (0.2) \times 10^{3}$	0.8	4.7	$8.7\ (0.4) \times 10^{5}$	0.5
H125A	$4.4\ (0.2) \times 10^{2}$	0.08	0.5	$8.2\ (0.4) \times 10^{5}$	0.5
H166A	$3.5\ (0.2) \times 10^{0}$	0.001	1.6	$2.2\ (0.3) \times 10^{3}$	0.001
H166F	$3.9\ (0.1) \times 10^{0}$	0.001	3.9	$1.0\ (0.1) \times 10^{3}$	0.0006
H127A/H166A/H200A	$2.8\ (0.4) \times 10^{-1}$	0.00008	0.8	$3.6(0.5) \times 10^{2}$	0.0002

Substituting any of the three histidines predicted to be involved in the charge-relay system with either Mh or Ta has a dramatic effect on the enzyme's catalytic activity. All of the generated mutants displayed diminished activities with more than 100-fold decrease in k_{cat} in comparison to their wild-type counterpart. Given the established role of H166 in the direct activation of Y265, low activities of mutants H166Mh and H166Ta are expected. However, H127 and H200 are distal from Y265 with no direct interaction possible between the tyrosine and histidine residues. Both of these residues are more than 10Å away from the catalytic site and more than 15Å away from the phenolic hydroxide group on the side chain of Y265. Therefore, such a drastic decrease in activity observed for H127 and H200 mutants strongly suggests that the side chains on these histidine residues are involved in activation of Y265 residue. Although the methyl group of Mh potentially disrupts the orientation of vicinal amino acid side chains of H127 and H200 when these two sites are substituted with Mh, Ta is spatially similar to histidine and therefore its substitution at H127 and H200 is expected to have minimal impact on the local protein structure around H127 and H200 residues. Unlike histidine whose two side chain nitrogen atoms can both serve as donor and acceptor roles in hydrogen bonding interactions, the low pK_a (~2) of the thiazole side chain of Ta dictates that both nitrogen and sulfur in the Ta side chain can only serve as hydrogen bond acceptors. This implies that the proton transfer chain as suggested by us, where the phenolic group of Y265 accepts a proton from the side chain of H166 which in turn accepts a proton from R219 and then extending this charge relay system, the side chain of H200 donates a proton to R219 while accepting a proton from the side chain of H127, would be interrupted on changing the proton donating

ability of all the histidine residues involved. This is illustrated by the diminished activity of enzyme observed by replacing H127, H200, and H166 with Ta, suggesting that the hydrogen bonding network in the predicted extended charge-relay system is disrupted and subsequently prohibits the direct activation of Y265 by this charge-relay system. All the mutant enzymes retain similar basal level activities possibly due to the access of Y265 to water molecule which acts as a substitute to the charge-relay system in activating the Y265 residue.

Contrary to their dramatic decreased k_{cat} values in comparison to the wild-type enzyme, H200Ta, H200Mh, H127Ta, and H127Mh have K_M values almost identical to that of the wild-type enzyme. This is expected since both H200 and H127 are distal to the active site and their replacement with Ta and Mh minimally affects the active site structure. However, both H166Ta and H166Mh have K_M values 4-fold lower than that of the wild-type enzyme. In comparison to histidine, both Ta and Mh are more hydrophobic. Their replacement of H166 may create a more hydrophobic active site environment that favors the substrate reaction with the bound PLP. For comparison purpose, we also expressed and determined catalytic activities of H166A and H166F mutants. Although both mutants show k_{cat} values not much lower than H166Ta and H166Mh, they have much higher K_M values, leading to overall k_{cat}/K_M values more than 10-fold lower than those of H166Ta and H166Mh. This observation may be because Ta and Mh mimic histidine better than ala and phe and therefore minimally disrupt the substrate binding when replacing H166. When H166, H200, and H127 are all replaced with Ala, the afforded mutant has a very low activity as expected. We also mutated H125, a histidine not part of the potential

charge relay system, to Ta and determined the activity of the afforded H125Ta mutant to serve as a control. As expected we observe no significant change in k_{cat} by introducing this mutation, the result indicates that H125Ta is solvent exposed and does not play any significant role in the activity of the enzyme as this surface-exposed residue is far away from the active site of the enzyme. A similar mutation at H125 to alanine also caused insignificant change to the catalytic activity of the enzyme. CD spectra of the wild-type enzyme and two H125 mutants are very similar, indicating no significant structural differences between them.

Conclusion

In summary, we have shown that the two close mimetics of histidine, Mh and Ta, can be genetically incorporated into proteins at amber codons using an engineered pyrrolysine incorporation machinery. Given their spatial resemblance and contrasting hydrogen bonding characteristics, replacing a histidine in a protein with Mh or Ta is expected to introduce minimal impact on the protein secondary structure, allowing for the fine tuning of the hydrogen bonding interactions critical in their enzymatic activity. By tweaking the hydrogen bonding interactions in a potentially extended charge-relay system by replacement of histidine residues involved in this system with Mh and Ta, we have demonstrated the possibility that all the histidine residues in this charge-relay system, H166, H200 and H127, coordinate the activation of the Y265 residue at the active site for catalysis, therefore suggesting the existence of this extended charge-relay system. The reported method can be easily adapted to study functional roles of histidine in many

proteins and enzymes and is expected to have profound impact in the biochemistry research area. One drawback of the reported method is the disruption of the hydrogen bonding system. An alternative solution is to the genetically incorporate fluorohistidines that allows fine-tuning histidine pK_a while maintaining all the hydrogen bonding potential.

CHAPTER IV

CONCLUDING REMARKS AND FUTURE OUTLOOK

In this dissertation, we accomplished improvement in an existing genetic incorporation tool, pyrrolysyl-tRNA/synthetase, which has been widely used over the past several years in incorporating several different ncAAs. This was achieved by evolving the N-terminal domain of *M. mazei* pyrrolysyl-tRNA synthetase that directly interacts with $tRNA^{Pyl}$, and a mutant clone that displayed improved amber suppression efficiency for the genetic incorporation of N_ϵ-(*tert*-butoxycarbonyl)-L-lysine three-fold more than the wild type synthetase.

Direct transfer of mutations R19H/H29R/T122S identified in this clone to two other PylRS mutants that were previously evolved for the genetic incorporation of N_ϵ-acetyl-L-lysine and N_ϵ-(4-azidobenzoxycarbonyl)-δ,ϵ-dehydrolysine also improved the incorporation efficacy of these two noncanonical amino acids. Since the three identified mutations are in the N-terminal domain that is separated from the catalytic domain for charging $tRNA^{Pyl}$ with a noncanonical amino acid, they can be potentially introduced to all other PylRS mutants for improving incorporation efficiency of their corresponding noncanonical amino acids, representing a general strategy to optimize pyrrolysine system-based noncanonical amino acid mutagenesis.

We also show in this dissertation, a method to incorporate histidine derivatives in proteins. Histidine, a unique amino acid with an imidazole side chain in which both the

nitrogen atoms can serve as a proton donor and proton acceptor in hydrogen bonding interactions, plays an important role in activity of several enzymes. Traditional methods are not ideal to probe the functional role of these histidines which are part of hydrogen bonding networks as fine-tuning the hydrogen bonding potential of the imidazole side chain is required for such a study but is infeasible due to limitations of the conventional mutagenesis methods.

Here, we showed incorporation technique for genetically encoding two histidine mimetics, 3-methyl-histidine and thiazole alanine, using engineered pyrrolysyl-tRNA/synthetase incorporation machinery. Our results verified the existence of an extended charge-relay system in alanine racemase with the alanine racemase mutants installed with 3-methyl-histidine or thiazole alanine at the predicted key histidine sites showing a dramatic loss in the enzyme's catalytic efficiency. This hinted at the role of this predicted charge-relay system in activating the Y265 residue present at the active site of the enzyme. Thus, this method can potentially serve as a useful tool in discerning function of key histidine residues, part of catalytic diads/triads of several enzymes and provide crucial information in enzyme mechanism which can be useful for design and development of inhibitors for these enzymes.

REFERENCES

[1] F. Crick, Central dogma of molecular biology. *Nature*, (1970), 227, 561-563.

[2] W. F. Anderson, L. Bosch, F. Gros, M. Grunberg-Manago, S. Ochoa, A. Rich, T. Staehelin, Initiation of protein synthesis in prokaryotic and eukaryotic systems. Summary of EMBO Workshop. *FEBS Lett.*, (1974), 48, 1-6.

[3] S. C. B. Yan, B. W. Grinnell, F. Wold, Post-translational modifications of proteins: some problems left to solve. *Trends Biochem. Sci.*, (1989), 14, 264-268.

[4] G. A. Khoury, R. C. Baliban, C. A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.*, (2011), 1, 90.

[5] S. Prabakaran, G. Lippens, H. Steen, J. Gunawardena, Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *WIREs. Syst. Biol. Med.*, (2012), 4, 565-583.

[6] O. N. Jensen, Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.*, (2006), 7, 391-403.

[7] L. N. Johnson, R. J. Lewis, Structural Basis for Control by Phosphorylation. *Chem. Rev.*, (2001), 101, 2209-2242.

[8] C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, M. Mann, Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science (New York, N.Y.)*, (2009), 325, 834.

[9] S. Zhao, W. Xu, W. Jiang, W. Yu, Y. Lin, T. Zhang, J. Yao, L. Zhou, Y. Zeng, H. Li, Y. Li, J. Shi, W. An, S. M. Hancock, F. He, L. Qin, J. Chin, P. Yang, X. Chen, Q. Lei, Y. Xiong, K.-L. Guan, Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science (New York, N.Y.)*, (2010), 327, 1000.

[10] M. A. Erce, C. N. Pang, G. Hart-Smith, M. R. Wilkins, The methylproteome and the intracellular methylation network. *Proteomics*, (2012), 12, 564-586.

[11] N. Haines, K. D. Irvine, Glycosylation regulates Notch signalling. *Nat. Rev. Mol. Cell Biol.*, (2003), 4, 786-797.

[12] M. E. Linder, R. J. Deschenes, Palmitoylation: policing protein stability and traffic. *Nat. Rev. Mol. Cell Biol.*, (2007), 8, 74-84.

[13] L. Wells, S. A. Whelan, G. W. Hart, O-GlcNAc: a regulatory post-translational modification. *Biochem. Biophys. Res. Commun.*, (2003), 302, 435-441.

[14] J. D. Schnell, L. Hicke, Non-traditional Functions of Ubiquitin and Ubiquitin-binding Proteins. *J. Biol. Chem.*, (2003), 278, 35857-35860.

[15] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem. Int. Ed.*, (2005), 44, 7342-7372.

[16] R. D. Kornberg, Chromatin structure: a repeating unit of histones and DNA. *Science (New York, N.Y.)*, (1974), 184, 868-871.

[17] R. D. Kornberg, Y. Lorch, Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, (1999), 98, 285-294.

[18] J. A. Latham, S. Y. R. Dent, Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.*, (2007), 14, 1017-1024.

[19] A. Groth, W. Rocha, A. Verreault, G. Almouzni, Chromatin challenges during DNA replication and repair. *Cell*, (2007), 128, 721-733.

[20] B. Li, M. Carey, J. L. Workman, The role of chromatin during transcription. *Cell*, (2007), 128, 707-719.

[21] A. Portela, M. Esteller, Epigenetic modifications and human disease. *Nat. Biotechnol.*, (2010), 28, 1057-1068.

[22] R. G. Urdinguio, J. V. Sanchez-Mut, M. Esteller, Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *The Lancet. Neurology*, (2009), 8, 1056-1072.

[23] M. D. Shahbazian, M. Grunstein, Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.*, (2007), 76, 75-100.

[24] S. R. Bhaumik, E. Smith, A. Shilatifard, Covalent modifications of histones during development and disease pathogenesis. *Nat. Struct. Mol. Biol.*, (2007), 14, 1008-1016.

[25] T. Ito, Role of histone modification in chromatin dynamics. *J. Biochem.*, (2007), 141, 609-614.

[26] A. Zippo, R. Serafini, M. Rocchigiani, S. Pennacchini, A. Krepelova, S. Oliviero, Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation. *Cell*, (2009), 138, 1122-1136.

[27] D. Nathan, K. Ingvarsdottir, D. E. Sterner, G. R. Bylebyl, M. Dokmanovic, J. A. Dorsey, K. A. Whelan, M. Krsmanovic, W. S. Lane, P. B. Meluh, E. S. Johnson, S. L. Berger, Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.*, (2006), 20, 966-976.

[28] I. Ivanovska, T. Khandan, T. Ito, T. L. Orr-Weaver, A histone code in meiosis: the histone kinase, NHK-1, is required for proper chromosomal architecture in Drosophila oocytes. *Genes Dev.*, (2005), 19, 2571-2582.

[29] Y. Dou, T. A. Milne, A. J. Tackett, E. R. Smith, A. Fukuda, J. Wysocka, C. D. Allis, B. T. Chait, J. L. Hess, R. G. Roeder, Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, (2005), 121, 873-885.

[30] K. Zhang, K. E. Williams, L. Huang, P. Yau, J. S. Siino, E. M. Bradbury, P. R. Jones, M. J. Minch, A. L. Burlingame, Histone Acetylation and Deacetylation: Identification of Acetylation and Methylation Sites of HeLa Histone H4 by Mass Spectrometry. *Mol. Cell. Proteomics*, (2002), 1, 500-508.

[31] W. Liu, A. S. Chan, H. Liu, S. A. Cochrane, J. C. Vederas, Solid supported chemical syntheses of both components of the lantibiotic lacticin 3147. *J. Am. Chem. Soc.*, (2011), 133, 14216-14219.

[32] P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, Synthesis of proteins by native chemical ligation. *Science (New York, N.Y.)*, (1994), 266, 776-779.

[33] T. M. Hackeng, J. H. Griffin, P. E. Dawson, Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.*, (1999), 96, 10068-10073.

[34] D. Schwarzer, P. A. Cole, Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.*, (2005), 9, 561-569.

[35] R. B. Merrifield, Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.*, (1963), 85, 2149-2154.

[36] J. M. Palomo, Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *RSC Adv.*, (2014), 4, 32658-32672.

[37] S. Kent, S. B. H. Kent, Chemical Synthesis of Peptides and Proteins. *Annu. Rev. Biochem.*, (1988), 57, 957-989.

[38] V. Made, S. Els-Heindl, A. G. Beck-Sickinger, Automated solid-phase peptide synthesis to obtain therapeutic peptides. *Beilstein J. Org. Chem.*, (2014), 10, 1197-1212.

[39] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, Synthesis of proteins by native chemical ligation. *Science (New York, N.Y.)*, (1994), 266, 776-779.

[40] P. E. Dawson, S. B. Kent, Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.*, (2000), 69, 923-960.

[41] T. N. Schumacher, L. M. Mayr, D. L. Minor, Jr., M. A. Milhollen, M. W. Burgess, P. S. Kim, Identification of D-peptide ligands through mirror-image phage display. *Science (New York, N.Y.)*, (1996), 271, 1854-1857.

[42] L. Wang, P. G. Schultz, Expanding the Genetic Code. *Angew. Chem. Int. Ed.*, (2005), 44, 34-66.

[43] T. W. Muir, D. Sondhi, P. A. Cole, Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.*, (1998), 95, 6705-6710.

[44] J. R. Sydor, M. Mariano, S. Sideris, S. Nock, Establishment of Intein-Mediated Protein Ligation under Denaturing Conditions: C-Terminal Labeling of a Single-Chain Antibody for Biochip Screening. *Bioconjugate Chem.*, (2002), 13, 707-712.

[45] B. Ayers, U. K. Blaschke, J. A. Camarero, G. J. Cotton, M. Holford, T. W. Muir, Introduction of unnatural amino acids into proteins using expressed protein ligation. *Biopolymers*, (1999), 51, 343-354.

[46] L. Berrade, J. A. Camarero, Expressed Protein Ligation: A Resourceful Tool to Study Protein Structure and Function. *Cell. Mol. Life Sci.*, (2009), 66, 3909-3922.

[47] H. P. Hemantha, N. Narendra, V. V. Sureshbabu, Total chemical synthesis of polypeptides and proteins: chemistry of ligation techniques and beyond. *Tetrahedron*, (2012), 68, 9491-9537.

[48] F. H. Crick, On protein synthesis. *Symp. Soc. Exp. Biol.*, (1958), 12, 138-163.

[49] F. Chapeville, F. Lipmann, G. Von Ehrenstein, B. Weisblum, W. J. Ray, Jr., S. Benzer, On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. U. S. A.*, (1962), 48, 1086-1092.

[50] A. E. Johnson, W. R. Woodward, E. Herbert, J. R. Menninger, N,iε-Acetyllysine transfer ribonucleic acid: a biologically active analogue of aminoacyl transfer ribonucleic acids. *Biochemistry*, (1976), 15, 569-575.

[51] M. Ibba, H. Hennecke, Towards engineering proteins by site-directed incorporation in vivo of non-natural amino acids. *Nat. Biotechnol.*, (1994), 12, 678-682.

[52] T. Kohno, D. Kohda, M. Haruki, S. Yokoyama, T. Miyazawa, Nonprotein amino acid furanomycin, unlike isoleucine in chemical structure, is charged to isoleucine tRNA by isoleucyl-tRNA synthetase and incorporated into protein. *J. Biol. Chem.*, (1990), 265, 6931-6935.

[53] B. Lemeignan, P. Sonigo, P. Marliere, Phenotypic suppression by incorporation of an alien amino acid. *J. Mol. Biol.*, (1993), 231, 161-166.

[54] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science (New York, N.Y.)*, (1989), 244, 182-188.

[55] J. A. Ellman, D. Mendel, P. G. Schultz, Site-specific incorporation of novel backbone structures into proteins. *Science (New York, N.Y.)*, (1992), 255, 197-200.

[56] A. G. Bruce, J. F. Atkins, N. Wills, O. Uhlenbeck, R. F. Gesteland, Replacement of anticodon loop nucleotides to produce functional tRNAs: amber suppressors derived from yeast tRNAPhe. *Proc. Natl. Acad. Sci. U. S. A.*, (1982), 79, 7127-7131.

[57] D. Grunberger, I. B. Weinstein, K. B. Jacobson, Codon Recognition by Enzymatically Mischarged Valine Transfer Ribonucleic Acid. *Science (New York, N.Y.)*, (1969), 166, 1635-1637.

[58] J. Roesser, M. Chorghade, S. Hecht, Ribosome-catalyzed formation of an abnormal peptide analog. *Biochemistry*, (1986), 25, 6361-6365.

[59] T. G. Heckler, J. R. Roesser, C. Xu, P. I. Chang, S. M. Hecht, Ribosomal binding and dipeptide formation by misacylated tRNAPhe's. *Biochemistry*, (1988), 27, 7254-7262.

[60] G. Baldini, B. Martoglio, A. Schachenmann, C. Zugliani, J. Brunner, Mischarging Escherichia coli tRNAPhe with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analog of phenylalanine. *Biochemistry*, (1988), 27, 7951-7959.

[61] N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. Dong, L. Moroder, R. Huber, Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. *FASEB J.*, (1999), 13, 41-51.

[62] M. Ibba, Strategies for in vitro and in vivo translation with non-natural amino acids. *Biotechnol. Genet. Eng. Rev.*, (1996), 13, 197-216.

[63] L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *J. Am. Chem. Soc.*, (2000), 122, 5010-5011.

[64] B. A. Steer, P. Schimmel, Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *J. Biol. Chem.*, (1999), 274, 35601-35606.

[65] H. Jakubowski, E. Goldman, Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.*, (1992), 56, 412-429.

[66] M. Pastrnak, T. J. Magliery, P. G. Schultz, A New Orthogonal Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. *Helv. Chim. Acta*, (2000), 83, 2277-2286.

[67] D. R. Liu, T. J. Magliery, M. Pastrnak, P. G. Schultz, Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.*, (1997), 94, 10092-10097.

[68] R. Furter, Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Sci.*, (1998), 7, 419-426.

[69] G. Srinivasan, C. M. James, J. A. Krzycki, Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science (New York, N.Y.)*, (2002), 296, 1459-1462.

[70] B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, M. K. Chan, A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science (New York, N.Y.)*, (2002), 296, 1462-1466.

[71] F. Zinoni, A. Birkmann, T. C. Stadtman, A. Bock, Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate

dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, (1986), 83, 4650-4654.

[72] D. G. Longstaff, R. C. Larue, J. E. Faust, A. Mahapatra, L. Zhang, K. B. Green-Church, J. A. Krzycki, A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl. Acad. Sci. U. S. A.*, (2007), 104, 1021-1026.

[73] M. A. Gaston, R. Jiang, J. A. Krzycki, Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.*, (2011), 14, 342-349.

[74] M. A. Gaston, L. Zhang, K. B. Green-Church, J. A. Krzycki, The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature*, (2011), 471, 647-650.

[75] S. A. Burke, J. A. Krzycki, Reconstitution of Monomethylamine:Coenzyme M Methyl Transfer with a Corrinoid Protein and Two Methyltransferases Purified fromMethanosarcina barkeri. *J. Biol. Chem.*, (1997), 272, 16570-16577.

[76] D. J. Ferguson, J. A. Krzycki, Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from Methanosarcina barkeri. *J. Bacteriol.*, (1997), 179, 846-852.

[77] D. J. Ferguson, N. Gorlatova, D. A. Grahame, J. A. Krzycki, Reconstitution of Dimethylamine:Coenzyme M Methyl Transfer with a Discrete Corrinoid Protein and Two Methyltransferases Purified fromMethanosarcina barkeri. *J. Biol. Chem.*, (2000), 275, 29053-29060.

[78] S. Herring, A. Ambrogelly, C. R. Polycarpo, D. Söll, Recognition of pyrrolysine tRNA by the Desulfitobacterium hafniense pyrrolysyl-tRNA synthetase. *Nucleic Acids Res.*, (2007), 35, 1270-1278.

[79] Y. Zhang, V. N. Gladyshev, High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Res.*, (2007), 35, 4952-4963.

[80] H. Neumann, S. Y. Peak-Chew, J. W. Chin, Genetically encoding N[epsi]-acetyllysine in recombinant proteins. *Nat Chem Biol*, (2008), 4, 232-234.

[81] T. Fekner, M. K. Chan, The pyrrolysine translational machinery as a genetic-code expansion tool. *Curr. Opin. Chem. Biol.*, (2011), 15, 387-391.

[82] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, T. A. Steitz, Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U. S. A.*, (2007), 104, 11268-11273.

[83] M. M. Lee, R. Jiang, R. Jain, R. C. Larue, J. Krzycki, M. K. Chan, Structure of Desulfitobacterium hafniense PylSc, a pyrrolysyl tRNA synthetase. *Biochem. Biophys. Res. Commun.*, (2008), 374, 470-474.

[84] D. Moras, Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci.*, (1992), 17, 159-164.

[85] T. Yanagisawa, R. Ishii, R. Fukunaga, O. Nureki, S. Yokoyama, Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina

mazei. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, (2006), 62, 1031-1033.

[86] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.*, (2008), 378, 634-652.

[87] T. Fekner, X. Li, M. K. Chan, Pyrrolysine Analogs for Translational Incorporation into Proteins. *Eur. J. Org. Chem.*, (2010), 2010, 4171-4179.

[88] A. Crnković, T. Suzuki, D. Söll, N. Reynolds, Pyrrolysyl-tRNA Synthetase, an Aminoacyl-tRNA Synthetase for Genetic Code Expansion. *Croat. Chem. Acta*, (2016), 89, 163-174.

[89] J. Krzycki, Translation of UAG as Pyrrolysine

Recoding: Expansion of Decoding Rules Enriches Gene Expression. *Nucleic Acids Mol. Biol.*, (2010), 24, 53-77.

[90] Y. S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.*, (2012), 134, 2950-2953.

[91] C. Liu, P. G. Schultz, Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.*, (2010), 79, 413-444.

[92] R. Jiang, J. A. Krzycki, PylSn and the Homologous N-terminal Domain of Pyrrolysyl-tRNA Synthetase Bind the tRNA That Is Essential for the Genetic Encoding of Pyrrolysine. *J. Biol. Chem.* , (2012), 287, 32738-32746.

[93] M. J. Schmidt, D. Summerer, Genetic code expansion as a tool to study regulatory processes of transcription. *Front. Chem.*, (2014), 2, 7.

[94] L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of Escherichia coli. *Science (New York, N.Y.)*, (2001), 292, 498-500.

[95] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.*, (2008), 15, 1187-1197.

[96] W. H. Zhang, G. Otting, C. J. Jackson, Protein engineering with unnatural amino acids. *Curr. Opin. Struc. Biol.*, (2013), 23, 581-587.

[97] S. H. W. Beiboer, B. Vandenberg, N. Dekker, R. C. Cox, H. M. Verheij, Incorporation of an unnatural amino acid in the active site of porcine pancreatic phospholipase A2. Substitution of histidine by 1,2,4-triazole-3-alanine yields an enzyme with high activity at acidic pH. *Protein Eng.*, (1996), 9, 345-352.

[98] R. Gan, J. G. Perez, E. D. Carlson, I. Ntai, F. J. Isaacs, N. L. Kelleher, M. C. Jewett, Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins. *Biotechnol. Bioeng.*, (2017), 114, 1074-1086.

[99] C. Fan, H. Xiong, N. M. Reynolds, D. Söll, Rationally evolving tRNA(Pyl) for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.*, (2015), 43, e156-e156.

[100] A. Tuley, Y.-S. Wang, X. Fang, Y. Kurra, Y. H. Rezenom, W. R. Liu, The genetic incorporation of thirteen novel non-canonical amino acids. *Chem. Commun.* , (2014), 50, 2673-2675.

[101] J. M. Tharp, Y. S. Wang, Y. J. Lee, Y. Yang, W. R. Liu, Genetic incorporation of seven ortho-substituted phenylalanine derivatives. *ACS Chem. Biol.*, (2014), 9, 884-890.

[102] V. Sharma, Y.-S. Wang, W. R. Liu, Probing the Catalytic Charge-Relay System in Alanine Racemase with Genetically Encoded Histidine Mimetics. *ACS Chemical Biology*, (2016), 11, 3305-3309.

[103] S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Green-Church, M. K. Chan, J. A. Krzycki, Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature*, (2004), 431, 333-335.

[104] W. Wan, J. M. Tharp, W. R. Liu, Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta*, (2014), 1844, 1059-1070.

[105] W. Stünkel, R. Campbell, Sirtuin 1 (SIRT1). *J. Biomol. Screen.*, (2011), 16, 1153-1169.

[106] A. Vaquero, M. Scher, D. Lee, H. Erdjument-Bromage, P. Tempst, D. Reinberg, Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell*, (2004), 16, 93-105.

[107] S. Kong, S. Kim, B. Sandal, S. Lee, B. Gao, D. D. Zhang, D. Fang, The Type III Histone Deacetylase Sirt1 Protein Suppresses p300-mediated Histone H3 Lysine

56 Acetylation at Bclaf1 Promoter to Inhibit T Cell Activation. *J. Biol. Chem.*, (2011), 286, 16967-16975.

[108] D. S. Wilson, A. D. Keefe, *Curr. Protoc. Mol. Biol.*, (2001), John Wiley & Sons, Inc.

[109] J. L. Feldman, K. E. Dittenhafer-Reed, N. Kudo, J. N. Thelen, A. Ito, M. Yoshida, J. M. Denu, Kinetic and structural basis for acyl-group selectivity and NAD(+)-dependence in Sirtuin-catalyzed deacylation. *Biochemistry*, (2015), 54, 3037-3050.

[110] Z. A. Wang, Y. Zeng, Y. Kurra, X. Wang, J. M. Tharp, E. C. Vatansever, W. W. Hsu, S. Dai, X. Fang, W. R. Liu, A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. *Angew. Chem. Int. Ed.*, (2017), 56, 212-216.

[111] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, An enhanced system for unnatural amino acid mutagenesis in E. coli. *J. Mol. Biol.*, (2010), 395, 361-374.

[112] Y. Lee, S. Mootien, C. Shoen, M. Destefano, P. Cirillo, O. A. Asojo, K. R. Yeung, M. Ledizet, M. H. Cynamon, P. A. Aristoff, R. A. Koski, P. A. Kaplan, K. G. Anthony, Inhibition of Mycobacterial Alanine Racemase Activity and Growth by Thiadiazolidinones. *Biochem. Pharmacol.*, (2013), 86, 222-230.

[113] T. Umehara, J. Kim, S. Lee, L. Guo, D. Söll, H. Park, N-Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett.*, (2012), 586, 729-733.

[114] H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, J. W. Chin, A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell*, (2009), 36, 153-163.

[115] J. Graff, L.-H. Tsai, Histone acetylation: molecular mnemonics on the chromatin. *Nat. Rev. Neurosci.*, (2013), 14, 97-111.

[116] A. J. M. Ruijter, A. H. Gennip, H. N. Caron, S. Kemp, A. B. P. Kuilenburg, Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.*, (2003), 370, 737.

[117] T. Finkel, C. X. Deng, R. Mostoslavsky, Recent progress in the biology and physiology of sirtuins. *Nature*, (2009), 460, 587-591.

[118] J. Rebek, On the structure of histidine and its role in enzyme active sites. *Struct. Chem.*, (1990), 1, 129-131.

[119] F. Schneider, Histidine in enzyme active centers. *Angew. Chem. Int. Ed. Engl.*, (1978), 17, 583-592.

[120] K. Hsieh, E. C. Jorgensen, Angiotensin II Analogues. 14. Roles of the Imidazole Nitrogens of Position-6 Histidine in Pressor Activity. *J. Med. Chem.*, (1979), 22, 1199-1206.

[121] Y. Ikeda, S. Kawahara, M. Taki, A. Kuno, T. Hasegawa, K. Taira, Synthesis of a novel histidine analogue and its efficient incorporation into a protein in vivo. *Protein Eng.*, (2003), 16, 699-706.

[122] D. C. Klein, J. L. Weller, K. L. Kirk, R. W. Hartley, Incorporation of 2-Fluoro-L-Histidine into Cellular Protein. *Mol. Pharmacol.*, (1977), 13, 1105-1110.

[123] J. T. Hammill, S. Miyake-Stoner, J. L. Hazen, J. C. Jackson, R. A. Mehl, Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat. Protoc.*, (2007), 2, 2601-2607.

[124] S. Sun, M. D. Toney, Evidence for a Two-Base Mechanism Involving Tyrosine-265 from Arginine-219 Mutants of Alanine Racemase. *Biochemistry*, (1999), 38, 4058-4065.

[125] C. A. Rohl, R. L. Baldwin, Comparison of NH Exchange and Circular Dichroism as Techniques for Measuring the Parameters of the Helix-Coil Transition in Peptides. *Biochemistry*, (1997), 36, 8435-8442.

[126] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Soll, O. Nureki, Pyrrolysyl-tRNA synthetase:tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature*, (2009), 457, 1163-1167.

[127] H. Xiao, F. B. Peters, P. Yang, S. Reed, J. R. Chittuluru, P. G. Schultz, Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.*, (2014), 9, 1092-1096.

[128] T. Yoshimura, K. Soda, *Alanine racemase: Structure and Function.* (2007), Wiley-VCH Verlag GmbH, Weinheim, Germany.

[129] K. Yokoigawa, Y. Okubo, H. Kawai, N. Esaki, K. Soda, Structure and function of psychrophilic alanine racemase. *J. Mol. Catal. B: Enzym.*, (2001), 12, 27-35.

[130] M. D. Toney, Reaction specificity in pyridoxal phosphate enzymes. *Arch. Biochem. Biophys.*, (2005), 433, 279-287.

[131] Y. Lin, J. Gao, A. Rubinstein, D. T. Major, Molecular dynamics simulations of the intramolecular proton transfer and carbanion stabilization in the pyridoxal 5′-phosphate dependent enzymes L-dopa decarboxylase and alanine racemase. *Biochim. Biophys. Acta.*, (2011), 1814, 1438-1446.

[132] R. B. Johnston, E. C. Schreiber, M. P. Davis, L. Jillson, W. T. Sorrell, M. E. Kirker, Catalytic properties of the active site of alanine racemase from B. subtilis. *Prog. Clin. Biol. Res.*, (1984), 144A, 339-350.

[133] J. L. Lynch, F. C. Neuhaus, On the Mechanism of Action of the Antibiotic O-Carbamyl-d-Serine in Streptococcus faecalis. *J. Bacteriol.*, (1966), 91, 449-460.

[134] F. R. Atherton, M. J. Hali, C. H. Hassall, R. W. Lambert, P. S. Ringrose, Phosphonopeptides as Antibacterial Agents: Rationale, Chemistry, and Structure-Activity Relationships. *Antimicrob. Agents Ch.*, (1979), 15, 677-683.

[135] M. D. Erion, C. T. Walsh, 1-Aminocyclopropanephosphonate: time-dependent inactivation of 1-aminocyclopropanecarboxylate deaminase and Bacillus stearothermophilus alanine racemase by slow dissociation behavior. *Biochemistry*, (1987), 26, 3417-3425.

[136] S. Mobashery, M. Johnston, Inactivation of alanine racemase by beta-chloro-L-alanine released enzymatically from amino acid and peptide C10-esters of deacetylcephalothin. *Biochemistry*, (1987), 26, 5878-5884.

[137] K. G. Anthony, U. Strych, K. R. Yeung, C. S. Shoen, O. Perez, K. L. Krause, M. H. Cynamon, P. A. Aristoff, R. A. Koski, New Classes of Alanine Racemase Inhibitors Identified by High-Throughput Screening Show Antimicrobial Activity against Mycobacterium tuberculosis. *PLoS One*, (2011), 6, e20374.

[138] M. Ciustea, S. Mootien, A. E. Rosato, O. Perez, P. Cirillo, K. R. Yeung, M. Ledizet, M. H. Cynamon, P. A. Aristoff, R. A. Koski, P. A. Kaplan, K. G. Anthony, Thiadiazolidinones: A New Class of Alanine Racemase Inhibitors with Antimicrobial Activity against Methicillin- Resistant S. aureus. *Biochem. Pharmacol.*, (2012), 83, 368-377.

[139] D. T. Major, J. Gao, A Combined Quantum Mechanical and Molecular Mechanical Study of the Reaction Mechanism and r-Amino Acidity in Alanine Racemase. *J. Am. Chem. Soc.*, (2006), 128, 16345-16357.

[140] A. Amadasi, M. Bertoldi, R. Contestabile, S. Bettati, B. Cellini, M. L. Salvo, C. Borri-Voltattorni, F. Bossa, A. Mozzarelli, Pyridoxal 5'-Phosphate Enzymes as Targets for Therapeutic Agents. *Curr. Med. Chem.*, (2007), 14, 1291-1324.

[141] T. Yoshimura, M. Goto, D-Amino acids in the brain: structure and function of pyridoxal phosphate-dependent amino acid racemases. *FEBS J.*, (2008), 275, 3527-3537.

[142] A. Watanabe, T. Yoshimura, B. Mikami, H. Hayashi, H. Kagamiyama, N. Esaki, Reaction Mechanism of Alanine Racemase from Bacillus stearothermophilus: X-Ray Crystillographic Studies of the Enzyme Bound With N-(5′-Phosphopyridoxyl) Alanine. *J. Biol. Chem.*, (2002), 277, 19166-19172.

[143] J. P. Shaw, G. A. Petsko, D. Ringe, Determination of the structure of alanine racemase from Bacillus stearothermophilus at 1.9-A resolution. *Biochemistry*, (1997), 36, 1329-1342.