SMALL MOLECULES FOR RIBOSWITCH DETECTION AND N-TERMINAL PROTEIN MODIFICATION

by

Pradeep Budhathoki

Bachelor of Science, 2009
Texas Christian University
Fort Worth, Texas

Submitted to the Graduate Faculty of the
College of Science and Engineering
Texas Christian University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

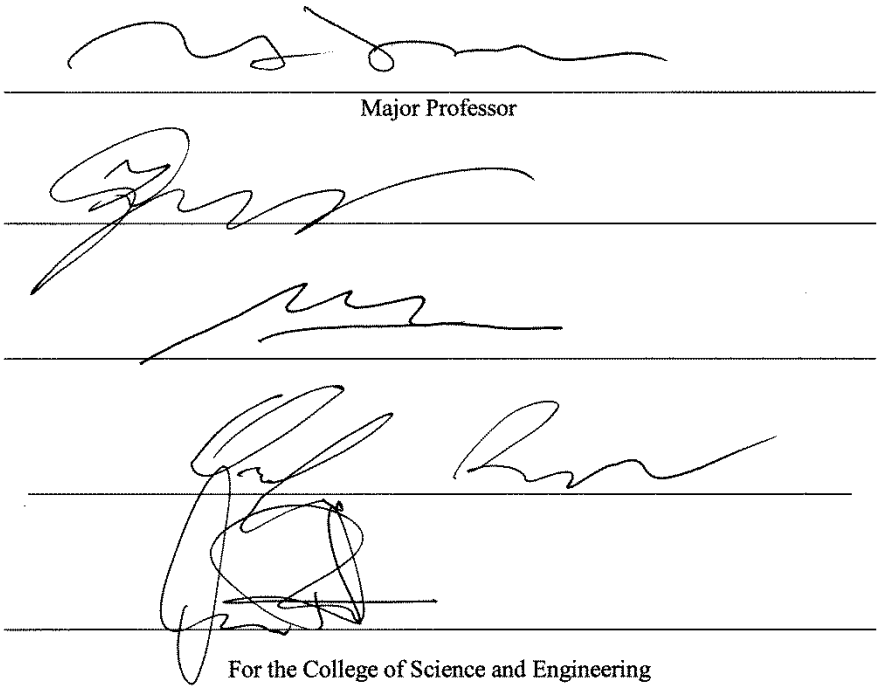
2015

SMALL MOLECULES FOR RIBOSWITCH DETECTION AND N-TERMINAL PROTEIN MODIFICATION

by

Pradeep Budhathoki

Dissertation approved



Major Professor

For the College of Science and Engineering

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Youngha Ryu for giving me the opportunity to join his research group. I am very grateful for his continued guidance and advice during my PhD program.

I also would like to thank each member of my committee, Dr. Onofrio Annunziata, Dr. Jean-Luc Montchamp and Dr. Zygmut (Karol) Gryczynski for their assistance, guidance, and valuable time overseeing my graduate studies. I am thankful to Dr. Annunziata for his help with the biophysical aspect of the riboswitch project and to Dr. Gryczynski for his advices on issues relating to fluorescence. I also deeply appreciate Dr. Montchamp for all his advice and suggestions especially pertaining to synthesis.

I want to offer special thanks to my group member, Dr. Lina Bernal-Perez for her support, assistance and friendship. I would also like to thank Dr. Eric Simanek and his group for letting me use their mass spectrometer and Dr. David Minter for his assistance in obtaining NMR spectra. I am also grateful to all my friends as well as all the members in the department. I am greatly indebted to the chemistry department and Texas Christian University for giving me the opportunity to pursue my PhD and for providing funds for our research.

Lastly, I would like to sincerely acknowledge my family and loved ones for being very patient and showering me with constant love and support during my doctoral studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ABBREVIATIONS

AEC	L-aminoethylcysteine
Ala	Alanine
Arg	Arginine
BLAST	Basic Local Alignment Search Tool
Boc-X	N-*tert*-Butoxycarbonyl-X
CP	2-Methyl-2,4-thiazolidinedicarboxylic acid
Cys	Cysteine
Dansyl	5-(Dimethylamino)naphthalene-1-sulfonyl
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	2,2-Dimethylthiazolidine
ESI	Electrospray ionization
GlcN6P	Glucosamine-6-phosphate
Gly	Glycine
HEPS	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HMP	4-Amino-5-hydroxylmethyl-2-methylpyrimidine
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
MAP	Methionine aminopeptidase
MES	2-(N-Morpholino)ethanesulfonic acid
Met	Methionine

Met_1	N-Terminal Methionine
mRNA	Messenger RNA
NBD	Nitrobenzoxadiazole
NCL	Native Chemical ligation
NHS	N-Hydroxysuccinimide
Ni-NTA	Nickel Nitrilotriacetic acid
nts	Nucleotides
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate buffer saline
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEG	Polyethylene glycol
PEP	Phosphoenolpyruvate
Pro	Proline
PTM	Posttranslational modification
pTsOH	*p*-Toluenesulfonic
RBS	Ribosome binding site
SAM	S-Adenosylmethionine
SD	Shine-Dalgarno sequence
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
Ser	Serine
TCEP	Tris-(2-carboxyethyl)phophine
Thr	Threonine
TLC	Thin layer chromatography

TMP	Thiamine monophosphate
TPP	Thimaine pyrophosphate
Tris	Tris(hydroxyemthyl)aminomethane
Trp	Tryptophan
TSTU	N,N,N′,N′-Tetramethyl-*O*-(*N*-succinimidyl)uronium tetrafluoroborate
UTR	Untranslated region
Val	Valine

Chapter 1. Riboswitches and the methods used for their analysis

The central dogma of molecular biology postulates the linear directional flow of genetic information from DNA to protein with RNA serving as central messenger (mRNA). Along with its role as the template for translation process, RNA is also known for its unique structures and catalytic functions, for instance in ribonucleoprotein complexes such as the ribosome. In addition, recent genomic explorations of bacteria have discovered novel functions of RNA in the realm of gene regulation, demonstrating that the roles of RNA are not limited to just as information carrier and catalytic components. This newly established RNA-based gene regulation, which hints at the RNA world hypothesis (pre-protein world) is principally based on the regulation of gene expressions mediated through transcription attenuation as in tryptophan operon,[1] thermosensing[2] and small regulatory RNA[3] in response to environmental and metabolic changes. The newest members of the gene regulating RNA are a specific region of mRNA called riboswitches, which were discovered in 2002.[4]

1.1 Riboswitches

The term "riboswitch" reflects the ability of these noncoding RNAs (ribo) to turn on and off (switch) associated genes. Typically found in the 5' untranslated regions (5'-UTRs) of certain bacterial mRNAs, riboswitches are structured RNA molecules that can specifically bind to cellular metabolites and regulate the genes associated with the

biosynthesis, transport, or degradation of the corresponding cellular metabolites.[5] Such binding between nucleic acids and metabolites is relatively rare and has been postulated to be the artifact of the pre-protein world. So far, more than 20 different structurally distinct classes of naturally occurring riboswitches have been discovered for many essential cellular metabolites including coenzymes, amino acids, nucleotide bases, sugars and metal ions (table 1.1).[6]

Riboswitch	Group	Ligand	K_d*
RFN	Coenzymes	Flavin mononucleotide (FMN)	~10 nM[7]
THI box		Thiamine pyrophosphate (TPP)	210-850 pM[8]
B_{12}		Adenosylcobalamin	~300 nM[9]
SAM-I,II, III		S-adenosylmethionine (SAM)	~150 nM[10]
THF		Tetrahydrofolate	~70 nM[11]
L Box	Amino acids	Lysine	~360 nM[12]
Glycine		Glycine	~ 20 µM[13]
Guanine	Nucleobase	Gauanine, hypoxanthine	~5nM[14]
Adenine		Adenine	~100 nM[14]
PreQ1		7-Aminoethyl 7-deazaguanine	~2 nM[15]
GlmS	Sugar	Glucosamine-6-phosphate	~200µM[16]
Mg	Metal	Magnesium	~400 µM[17]

Table 1.1 Major classes of riboswitches with their cognate ligands. *k_d varies based on the organism that the riboswitch is obtained from.

Usually a few hundred nucleotides (nts) in length, riboswitches consist of an evolutionarily conserved ligand binding aptamer domain along with a variable sequence, called the expression platform. Riboswitches generally utilize feedback repression to regulate the downstream coding sequences. In presence of abundant cellular metabolite,

the selective binding of the metabolite to the aptamer domain induces conformational change in the expression platform leading to the modulation of the downstream genes. Attenuation uses a similar mechanism of conformational change for gene regulation.[18] However, riboswitches are unique in their ability to directly sense and bind various metabolites without the need for intermediate molecules. In contrast to other *trans*-acting small RNAs that control gene expression, riboswitches are primarily *cis*-acting regulatory elements; i.e., they are found in the vicinity of the coding sequence that they modulate.

1.2 Mechanism of riboswitch-mediated gene regulation

All the riboswitch classes characterized to date have been found in eubacteria with the only exception being thiamine pyrophosphate (TPP) riboswitches, which also have been found in archaea[19] and eukaryotes.[20] The regulatory activity of bacterial riboswitches is mainly directed at either transcription or translation of genes involved in the metabolite transport and biosynthesis. All riboswitch mediated gene regulation is based on the ligand-dependent changes in RNA conformations. There are several different mechanisms that can either repress or, more seldom, induce gene activity.

1.2.1 *Transcription attenuation*

Transcription attenuation is the most prevalent mechanism of riboswitch-mediated gene regulation. In the presence of a low concentration of metabolite, a ligand-free riboswitch

adopts a conformation containing an antiterminator structure that prevents the formation of a transcription termination loop and permits the continuation of transcription. Upon binding the ligand to the aptamer domain, the sequence contributing to the formation of antiterminator becomes sequestered, alternatively leading to the formation of intrinsic or Rho-dependent transcription terminator that consequently halts transcription prematurely.

Based on computational prediction of putative intrinsic terminator sequences, the transcription attenuation mechanism is prevalent in Gram-positive bacteria.[21] For example, the *ilv-leu* operon in *B. subtilius*, which is involved in the biosynthesis of branched- chain amino acids (valine, isoluceine, and leucine), is regulated via transcription attenuation mechanism by the T-box riboswitch.[22]

1.2.2 *Sequestration of the ribosome binding site (RBS)*

Along with transcription termination, translation termination is a common riboswitch-mediated gene regulation mechanism in prokaryotes. For gene regulation at the translation level, the ligand bound riboswitch forms an anti-SD stem-loop conformation that blocks the accessibility of the Shine-Dalgarno sequence, which recruits the small ribosomal subunit to initiate translation. According to putative-sequence specific computational studies, this mechanism is mostly found in bacterial groups including archeae other than Gram-positive bacteria.[21] Some classes of riboswitch, such as the SAM aptamer, contain both intrinsic transcription terminator stem and sequence that

interferes with the accessibility of Shine-Dalgarno sequences, utilizing dual regulation at both transcription and translation levels.[23]

1.2.3 *Riboswitch as ribozyme*

In another mechanism of gene regulation, the riboswitch acts as a metabolite-responsive ribozyme. For example, in the *B. subtilis glmS* riboswitch, the binding of glucosamine-6-phosphate (GlcN6P) induces a conformational change in the leader region of the riboswitch and consequently activates the allosteric ribozyme that influences the fate of downstream *glmS* genes.[24] The activation of the ribozyme leads to self-cleavage at a specific site that promotes transcript destabilization, as the newly generated 5' hydroxyl group targets the mRNA for degradation by RNase J.[25]

1.2.4 *Splicing mRNA*

The mechanism of splicing mRNA is used by TPP riboswitches in eukaryotes. In contrast to almost all bacterial riboswitches found in the 5' UTR, the TPP riboswitch is found in the intron of fungal[26-27] and plant pre-mRNA.[20] When bound to the ligand, the riboswitch appears to regulate alternative splicing of introns that are expressed at different levels, thereby regulating the thiamine metabolism genes.[26]

1.3 Significance and importance of riboswitches

1.3.1 Riboswitches as antibacterial agent

Riboswitches are of profound interest for reasons that go beyond their status as unique and widespread gene regulatory elements. Given that riboswitches play significant roles in the genetic control almost exclusively in bacteria, they have been proposed as targets for discovery of potential antibiotics.[28] The premise of this approach is that the analogs of cellular metabolites maintain the riboswitch binding property, but are not metabolically useful and thus could cause misregulation of critical genes and pathways crucial for bacterial survival. Riboswitches are fundamentally different drug targets than the conventional RNA drug targets, in which the design of new ligands directed at RNA targets could be problematic due to the issue of selectivity. Fortunately, riboswitches have evolved as structured receptors for the purpose of binding small molecule ligands with precise selectivity.

Several known antibacterial compounds have been confirmed to function by targeting riboswitches. Two analogs of lysine, L-aminoethylcysteine[29] (AEC) and DL-4-oxalysine[30] (Fig. 1.1) were originally reported to inhibit certain Gram-positive bacteria. But only recently, it was revealed that both compounds bind to lysine riboswitch and repress lysine regulated genes in *B. subtilis*.[12] Inhibition of lysyl-tRNA synthetase (LysRS) by AEC has also been reported to be the primary antimicrobial action of AEC.[31] However in this case, AEC resistance achieved by mutations to the *lysC* promotes more effective competition of lysine with AEC for binding to LysRS, indicating an indirect

effect of the lysine riboswitch-mediated increase in aspartokinase levels and cellular concentration of lysine.

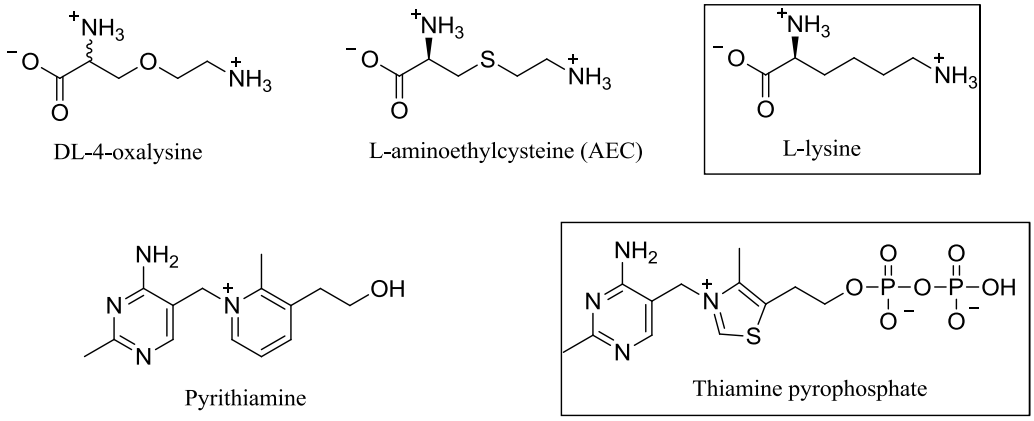


Figure 1.1 Structures of known antibacterial agents believed to target riboswitch RNAs compared to cellular metabolites (boxed)

Similarly, pyrithiamine (Fig. 1.1), in its phosphorylated form, possibly inhibits bacterial and fungal growth by targeting one or more TPP riboswitch-mediated gene control in these organisms.[32-33] Hence, studying novel ligand and riboswitch binding interactions especially with the help of three dimensional (3D) structures of different riboswitches, would be helpful for the discovery of new antibiotics.

1.3.2 Riboswitches as biosensors

In addition to drug targets, riboswitches are also of huge interest due to their great potential as a tool in synthetic biology and chemical biology. The ability of the aptamer domain of riboswitch to selectively bind small molecules to regulate genes has attracted

the attention of researchers. Although the riboswitch-mediated gene regulation was conclusively revealed in the natural system for the first time in 2002, [4] the idea that RNA could bind a wide range of proteins and small molecules ligands was well established in early 1900's. Several groups independently reported methods for *in vitro* selection and evolution of RNA sequences that could bind wide range of molecules.[34-36] These aptamers showed great binding selectivity and affinity similar to proteins, while working with only four different nucleotides compared to 20 different amino acids. These aptamers were created by the process known as SELEX (Systematic Evolution of Ligand by Exponential Enrichment), by which a large pool of usually 10^{14} randomized RNA sequences is subjected to a selection process.[35] First, the starting pool of RNA sequences is passed through a column containing covalently immobilized target ligand. Then the column is washed to remove the unbound RNA. Finally, washing the column with free ligand elutes out potential aptamers. This initial pool of aptamers is enriched by the process of reverse transcription and polymerase chain reaction (PCR), and the entire procedure is repeated to ultimately obtain aptamers that have increased affinity and selectivity to the desired target ligand.

SELEX can be used to develop engineered riboswitches. When coupled to certain reporter genes, the synthetic riboswitches can act as biosensors. These synthetic riboswitches, when incorporated in the 5' UTR region of certain reporter genes, have shown to control translation initiation in mammalian cell, yeast, and bacteria in response to unconventional ligands such as Hoechst dye,[37] tetracycline,[38] and theophylline[39] respectively. Similarly, engineered riboswitches have been successfully used to control cellular transcription. For example, one aptamer, selected *in vitro* against the bacterial

repressor protein, TetR, regulated TetR controlled genes at the transcription level when screened *in vivo*.[40] The theophylline aptamer linked to ribozyme, induced self-cleavage and inhibition without the involvement of specific transcriptional or translational elements.[41] Hence it is feasible that riboswitch can be engineered to respond to almost any ligand of choice and are, therefore, of great interest for applications in synthetic biology, especially for developing new genetic tools to control gene expression in response to exogenous ligands. Synthetic riboswitches can also be used as sensors that could effectively detect any target ligands *in vitro*.

1.4 Current methods to study the riboswitch metabolite binding interaction

Originally, a number of riboswitch classes were identified by delving into the literature reporting gene regulation in the absence of protein factor. Currently, however, comparative genomic approaches have become a significant tool in the discovery of new riboswitches. Comparative genomic approaches compare the genomic features of different organisms to make some alignments of sequences and look for orthologous sequences. For example, the use of bioinformatics tool, BLAST (Basic Local Alignment Search Tool) enabled in the identification of UTR homologous sequences,[24] which were inspected for conserved structure and later experimentally confirmed as glycine riboswitch.[42] Once the potential riboswitch candidate is identified, it is necessary to demonstrate that a genetic control element is induced by the selective binding of a specific metabolite in absence of any regulatory proteins. Metabolite binding in absence of proteins can be studied by several methods including in-line probing, equilibrium

dialysis, *in vitro* transcription termination assay, and isothermal titration calorimetry (ITC).

1.4.1 In-line probing of aptamer structures

The in-line probing assay has been frequently used for detecting RNA-metabolite binding interactions and determining apparent K_d (dissociation constant) values for these interactions.[43] This technique also provides hints about the bound and unbound conformational states of riboswitches.

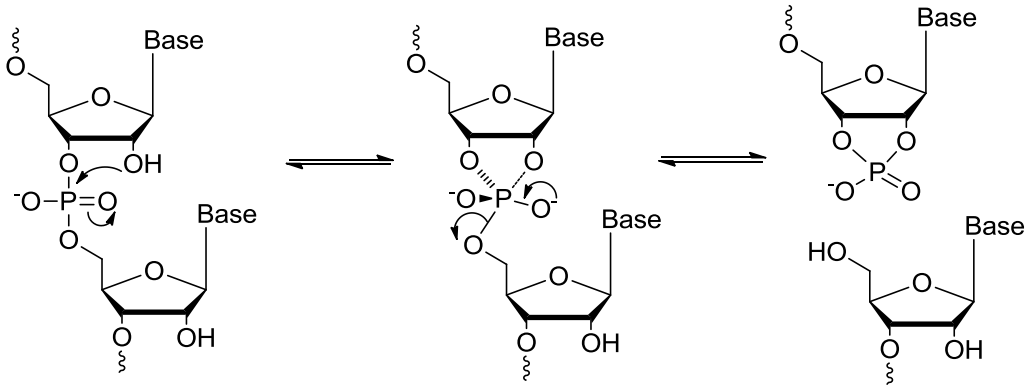


Figure 1.2 RNA degradation due to internal transesterification and phosphodiester cleavage

This technique exploits the RNA degradation due to internal transesterification. The slow and non-enzymatic cleavage of phosphodiester initiates with the "in-line" nucleophilic attack of the 2' oxygen on the adjacent phosphorus center, followed by intramolecular displacement of the 5′ oxygen in the adjacent phosphorus center to efficiently cleave the RNA linkage. This cleavage reaction occurs more frequently in flexible single-stranded RNA and less commonly in strands involved in the secondary

structure. When an aptamer binds to a metabolite, change can be observed in the pattern and the intensity of cleavage, which are typically at different concentration of metabolites distinct to those observed for ligand unbound aptamer. The cleavage patterns at different metabolite concentrations are monitored by PAGE (Polyacrylamide Gel Electrophoresis) and the intensity of each fragment is used to find the binding constant, also providing clues about the conformations and the mechanism of the riboswitch action.

1.4.2. *Aptamer-ligand binding analysis by equilibrium dialysis*

Equilibrium dialysis is a simple biophysical method for riboswitch-ligand binding analysis.[44] This technique uses a two-reservoir system in which one chamber contains an excess of the riboswitch aptamer while the other contains radiolabeled ligand substrate or other analogs to test for interactions. The two chambers are separated by a permeable dialysis membrane that permits the diffusion of the small ligands between the chambers while the aptamer is restricted. The shift in distribution of the radiolabeled ligands towards the chamber containing the aptamer indicates affinity and specificity (in presence of competitor) while equal distribution of the ligand within the two chambers indicates no interaction between the aptamer and the target ligand. The gathered data can be used to determine the binding constant and stoichiometry of the riboswitch-ligand interaction.

1.4.3. *In vitro transcription termination assay for aptamer-ligand binding*

In vitro transcription assay has also been used to analyze the aptamer-metabolite binding interaction.[45] This assay employs the riboswitch-mediated transcription termination. It uses a template containing a potential riboswitch (with an intrinsic terminator sequence as part of its expression platform) and T7 RNA polymerase for *in vitro* transcription, which generates transcripts of varying lengths depending on the concentration of the ligand. The presence or the absence of the ligand determines the formation of the transcription terminator and the read-through of the polymerase. With the assistance of PAGE, the ratio between the amount of truncated transcript (where termination occurs) and that of full transcript (when read-through occurs can be monitored to determine the affinity between the aptamer and the ligand.

1.4.4. *Isothermal titration calorimetry (ITC) analysis of riboswitch-ligand interaction*

Recently, isothermal titration calorimetry has been used to analyze the interaction of both natural ligands and their analogs with their cognate riboswitch.[46] This technique measures the heat evolved or absorbed during a reaction or binding. It is used to determine the enthalpy, entropy, stoichiometry, and binding constants of the reaction. The interaction of the aptamer domain with a specific ligand typically involves heat. Therefore, the associated thermodynamic parameters and the binding affinity can be efficiently measured by ITC. A significant advantage of ITC over other methods is that the heat

associated with the binding without the specific labeling molecules. In addition, ITC can be performed over a broad range of pH and temperature.

1.5. Specific research objective

The main objective of the first part of this dissertation is to report a fluorescence-based method to analyze riboswitch-ligand binding. Although the aforementioned techniques have been successfully used to analyze the riboswitch-ligand interactions, they do have their own disadvantages. The In-line probing assay requires a multi-step procedure and a long incubation period for RNA cleavage. In addition, radiolabelled molecules are hazardous and require extra precaution.

The *in vitro* transcription assay works only for the riboswitch with the transcription termination mechanism. There are other riboswitch-mediated mechanisms of gene regulation. Hence, negative results from this assay do not explicitly demonstrate a lack of interaction between aptamer and ligand, since some riboswitches have been shown to regulate gene expression by translational control[47] and also by RNA processing or splicing.[20]

Although very affordable, equilibrium dialysis is slow and requires a large volume. New ITC instruments can analyze a small volume of analytes but riboswitch analysis by ITC is still not economical.

Over these aforementioned methods, fluorescence-based methods offer several distinct advantages. Unlike in-line probing and equilibrium dialysis using radioactive

labels, the fluorescence-based assay is less hazardous. High sensitivity is very useful, especially when working with a small amount of samples. In addition, the fluorescence measurements are fast and multiple fluorescence properties (quenching, anisotropy, and resonance energy transfer) can be utilized to determine the binding constant, the conformational dynamics and the mechanism of action. Fluorescence-based assays can be performed in parallel formats for high throughput screening.[48] Even though modifying the ligand can potentially influence binding, it can be prevented by careful design of the analogs.

Fluorescence-based assays, in the literature, were used to monitor riboswitch-binding interactions specifically for intrinsically fluorescent ligands such as riboflavin[49] and 2-aminopurine.[50] The fluorescence quenching that occurs when a riboswitch binds to riboflavin and 2-aminopurin was monitored to obtain the binding affinity of these ligands to their riboswitches. However, these ligands often do not have the desired fluorescence properties and most riboswitch ligands do not have intrinsic fluorescence. As such, we aimed to develop a fluorescence-based method to analyze the binding interaction of several riboswitches.

Chapter 2. Rationally designed fluorescent probes for various riboswitches

2.1 Rational design of fluorescent riboswitch-binding probes on the basis of crystal structures

We aimed to rationally design and synthesize novel probes that could effectively and selectively bind to various classes of riboswitches. The structure-based designs of these probes rely on the X-ray crystal structure of the ligand-bound riboswitches. These 3D crystal structures not only provide information about the ligand recognition but also provide clues for developing potential probes with higher affinity and selectivity. The investigation of the binding pocket of several riboswitches allowed us to identify the accessible positions of several native riboswitch ligands, to which fluorescent dyes or generic RNA binding molecules could be attached without significantly affecting the riboswitch-ligand interactions. The fluorescence properties of the labels were used to analyze the binding interaction.

2.1.1 *Crystal structure of the lysine riboswitch and possible accessible points in lysine*

The lysine riboswitch uses a transcriptional attenuation mechanism to repress the production of an enzyme, aspartate-semialdehyde dehydrogenase,[47] which is involved in

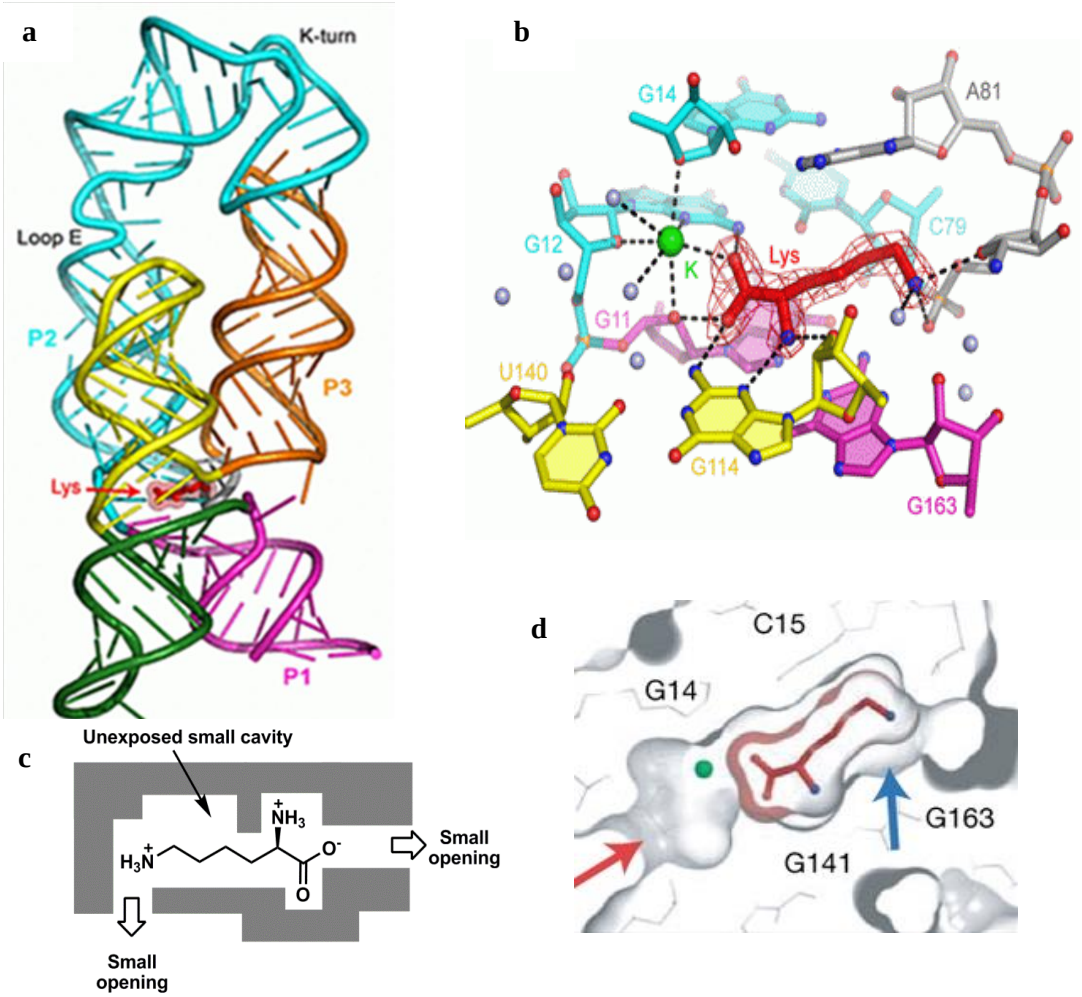


Figure 2.1. Structure of lysine-bound *lysC* riboswitch from *Thermotoga maritima*; **a**) Overall lysine riboswitch structure in a ribbon representation; **b**) Binding interaction between the residues in the binding pocket and lysine at the core junction (red); **c**) Schematic representation of lysine-bound binding pocket showing the two opening as in figure **d**; **d**) Cross-section through the surface view of the lysine binding pocket displaying accessible positions, as indicated by red arrow (carboxylate end of lysine) and blue arrow (amino end of lysine); (Adapted from Serganov et al. 2008)[51]

the biosynthesis of a precursor for amino acids of aspartate family that includes methionine, threonine, and lysine. The X-ray crystal structures of the 174-nucleotide long aptamer domain of the *Thermotoga maritima* lysine riboswitch in the lysine-bound and

free states revealed the molecular basis of amino acid recognition.[51] The structure of the lysine-bound riboswitch, also known as an "L-box", features three-helical and two-helical bundles radiating from a compact five-helical junction. Lysine interacts with this junctional core of the riboswitch through shape-complementarity within the elongated binding pocket and through several direct and K^+-mediated hydrogen bonds and coordination bonds to its charged ends (Fig. 2.1a). Interestingly, the crystal structure of the ligand-bound *Thermotoga maritima* lysine riboswitch revealed two small openings: one at the carboxylate end and the other at the ε-amino group of lysine in the binding pocket (Fig. 2.1 b and c).[51]

These two charged ends of lysine were suggested as potential sites for modification to develop potential lysine-like antibacterial analogs. L-Homoarginine and N^6-1-iminoethyl-L-lysine, which have the lysine ε-amino group substituted with guanidinium and acetamidine group, respectively, showed selective binding to the lysine riboswitch with k_d values of 7 ± 4 μM and 11 ± 3 μM, respectively.[12] The lysine riboswitch also has shown tolerance to lysine analogs with modifications at C4 positions of lysine with an isosteric group such as O, S and SO_2 because of the presence of small cavity between the C4 and N7 (Fig. 2.1 b and d).[51] Compounds S-(2-aminoethyl)-L-cysteine (AEC) and L-4-oxalysine, which contain sulfur and oxygen at position C4, respectively have shown to be placed within the binding pocket in similar manner to lysine. Despite similar placements, AEC (k_d = 30 μM) and L-4-oxalysine (k_d = 13 ± 2 μM) showed weaker binding affinities to the aptamer possibly due the electronegativity of O and S substitution at the C4 position. Lysine analogs with modification at α- amino group have shown poor binding affinity, such as, N^2-acetyl-L-lysine and N^2-methyl-L-

lysine both have k_d >1000.[12] On the other hand, the carboxylate end was yet to be explored; therefore was the targeted site for our structural modifications.

2.1.2 *Crystal structure of the TPP riboswitch and possible accessible points in TPP*

A conserved sequence named THI-box, found in the 5' UTR region, represses thiamine biosynthetic or transport genes in response to thiamine pyrophosphate (TPP), a crucial coenzyme for many enzymes involved in carbohydrate metabolism.[19] The TPP riboswitch regulates several genes involved in the biosynthesis of thiamine precursors, 4-amino-5-hydroxymethyl-2-methyl pyrimidine pyrophosphate and 5-(2-hydroxyethyl)-4-methyl thiazole monophosphate.[52] The gene regulation mechanism of the THI-box varies with organisms. It regulates transcription termination and translational inhibition in prokaryotes while it controls intron splicing in fungi such as in *Aspergillus*.[26]

The X-ray crystal structure of the 80-nucleotide long sensing domain of the TPP riboswitch from *E. coli* provides structural insights into TPP recognition and gene regulation.[53] The TPP aptamer consisted of conserved nucleotides and a secondary structure that formed two parallel helices connected to a third helix by means of a three way junction. TPP is positioned perpendicularly to these two main helices. Two separate pockets accommodate the ends of the ligand by hydrogen bonds (Fig. 2.2a and c), whereas the central thiazole ring forms an inter-domain bridge between the two helices. Similarly, the binding pocket of *Arabidopsis thaliana* TPP riboswitches shows the 4-amino-5-hydroxylmethyl-2-methylpyrimidine (HMP) moiety of TPP is enclosed in one domain via H-bond interactions and intercalation, and the pyrophosphate (PP) moiety

coordinates to the Mg^{2+} ions in the another domain.[54] The central thiazole ring is exposed to solvent through large openings (Fig. 2.2 b and d). In addition, the structure of the TPP riboswitch bound to benfotiamine (Fig. 2.2e), a synthetic acyclic S-benzoyl derivative of thiamine monophosphate (TMP), indicated that the TPP riboswitch could accommodate a substantial modification of the thiazole ring.[55] Therefore, the central thiazole ring was the target for our structural modification.

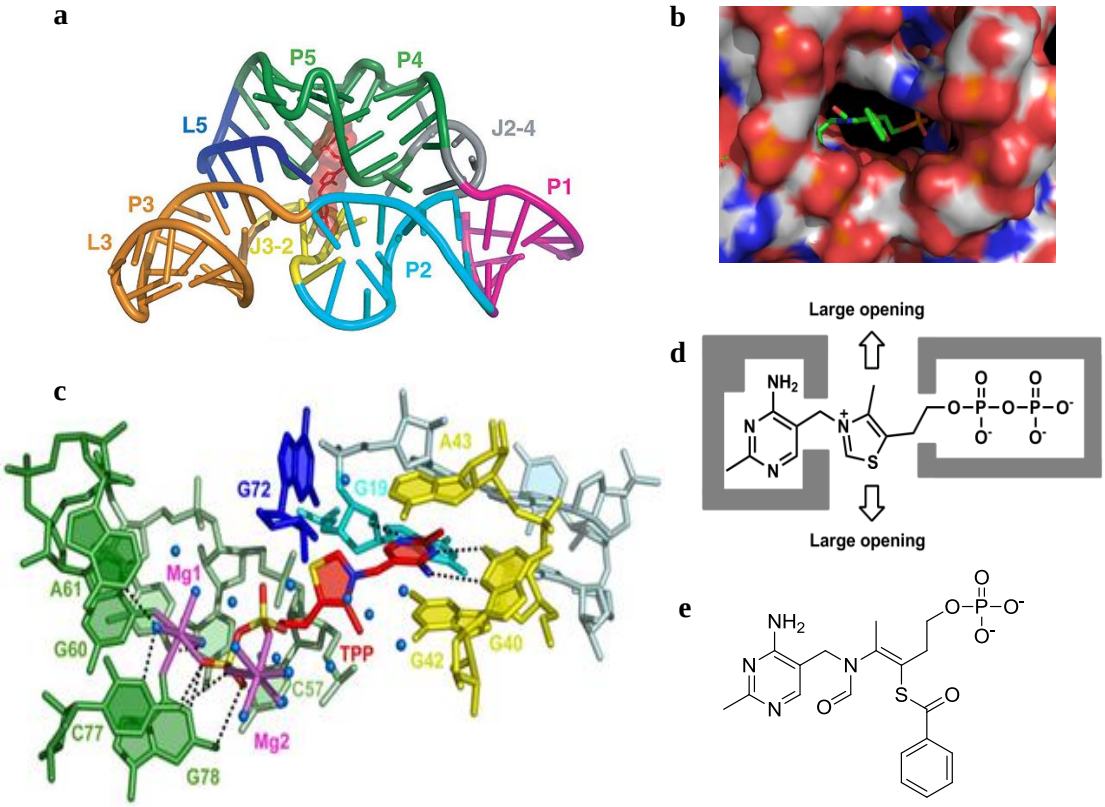
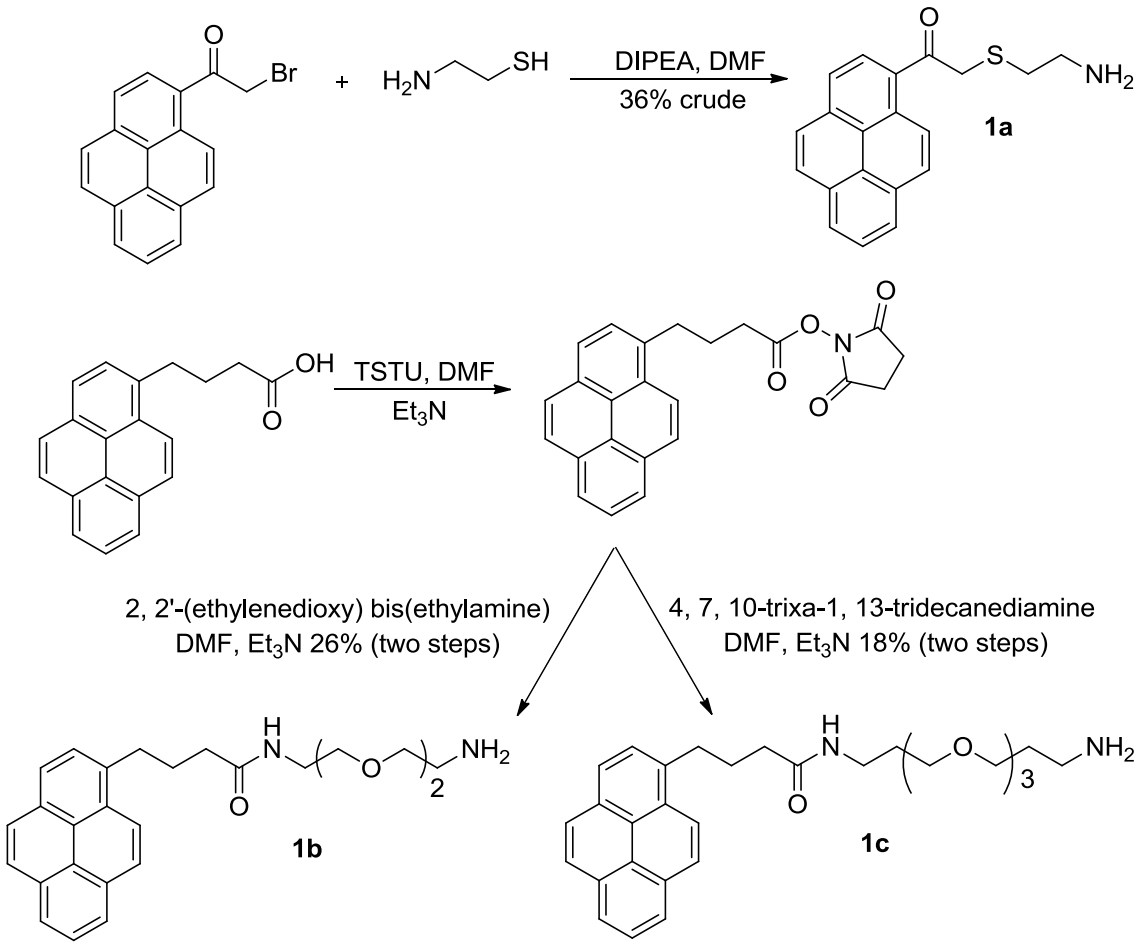


Figure 2.2. Structure of TPP-bound TPP riboswitch from *E. coli* and *Arabidopsis thaliana*; **a**) Crystal structure of the TPP-bound (red) sensing domain of TPP riboswitch from *E. coli*; **b**) Space filling representation of the residues and the stick representation of TPP clearly depicting the exposed central thiazole ring; **c**) Binding interaction between TPP (red) and the residues in the binding pocket of TPP riboswitch from *E. coli*; **d**) Schematic representation of the binding pocket displaying the accessible positions of TPP; **e**) Structure of benfotimaine. (Adapted from Serganov et al. 2006)[53]

2.2. Results and discussion

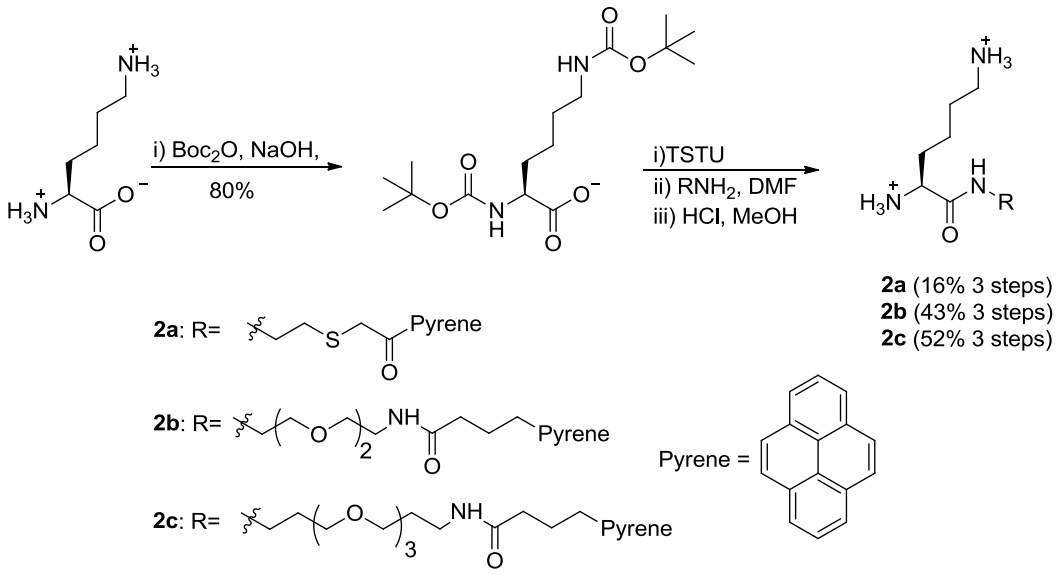
2.2.1 *Synthesis of lysine riboswitch binding fluorescent probes*

For the design of our fluorophore-labeled lysine riboswitch probes, we extended the carboxylate group of lysine via an amide bond and labeled it with a pyrene fluorophore. Pyrene, a polycyclic hydrocarbon consisting of four fused benzene rings, has been a useful



Scheme 2.1 Synthesis of pyrene-labeled amide linkers

fluorophore for numerous biophysical studies. We synthesized three pyrene analogs, **1a**, **1b** and **1c,** in which pyrene is attached to linkers of varying length. The pyrene labeled amine linkers were synthesized as shown in Scheme 2.1. S-alkylation of cysteamine with 1-(bromoacetyl) pyerene (PA-Br) produced **1a** (36% crude), which was used without purification for subsequent reaction. Mono N-acylation of two diamines reagents, 2,2'-(ethylenedioxy)bis(ethylamine) and 4,7,10-trioxa-1,13-tridecanediamine with the NHS-ester of pyrenebutyric acid (PBA-NHS) gave **1b** (26%) and **1c** (18%) respectively, after purification using preparative TLC developed with 2% methanol in chloroform. The polyethylene glycol linkers (PEG) were used to enhance water solubility and also to mimic several water molecules present in the channel at the carboxylate end.

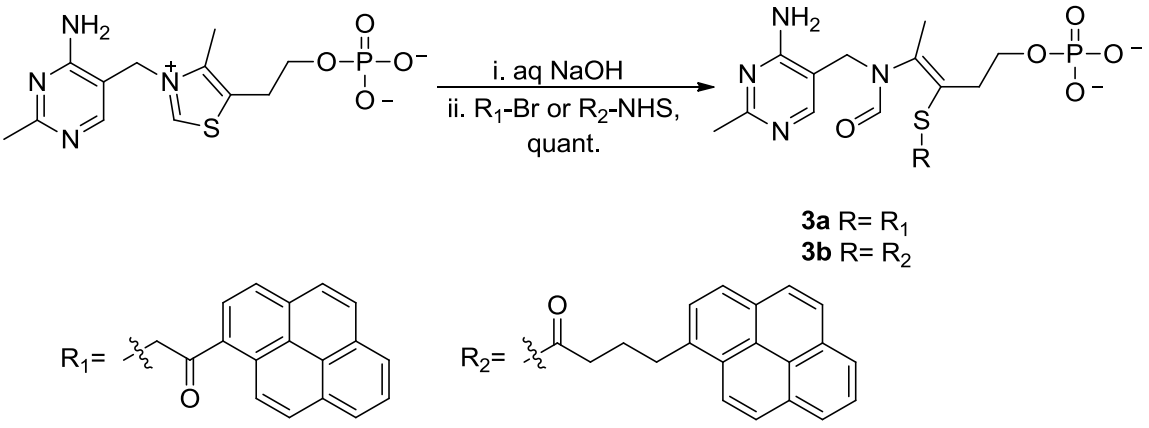


Scheme 2.2. Synthesis of pyrene-conjugated lysine probes

Boc-protected lysine was prepared by protecting two amino groups of lysine using Di-*tert*-butyl dicarbonate, as described in the literature.[56] The NHS ester of Boc-protected

lysine was prepared by the reaction of Boc-protected lysine with the coupling agent N,N,N′,N′-tetramethyl-*O*-(*N*-succinimidyl)uronium tetrafluoroborate (TSTU). The reaction of these fluorophore-labeled linkers with N-hydroxylsuccinimidyl (NHS) ester of N-*tert*-butoxycarbonyl (Boc)-protected lysine and subsequent purification using preparative TLC developed with 2% methanol in chloroform and Boc deprotection in methanolic HCl gave the compounds **2a** (16% in 3 steps), **2b** (43% in 3 steps) and **2c** (52% in 3 steps) (Scheme 2.2). Prior to the deprotection step, **2a** was purified using silica gel chromatography (1:1 hexane/ethyl acetate) while **2b** and **2c** were purified using preparative TLC (1:10 methanol/chloroform).

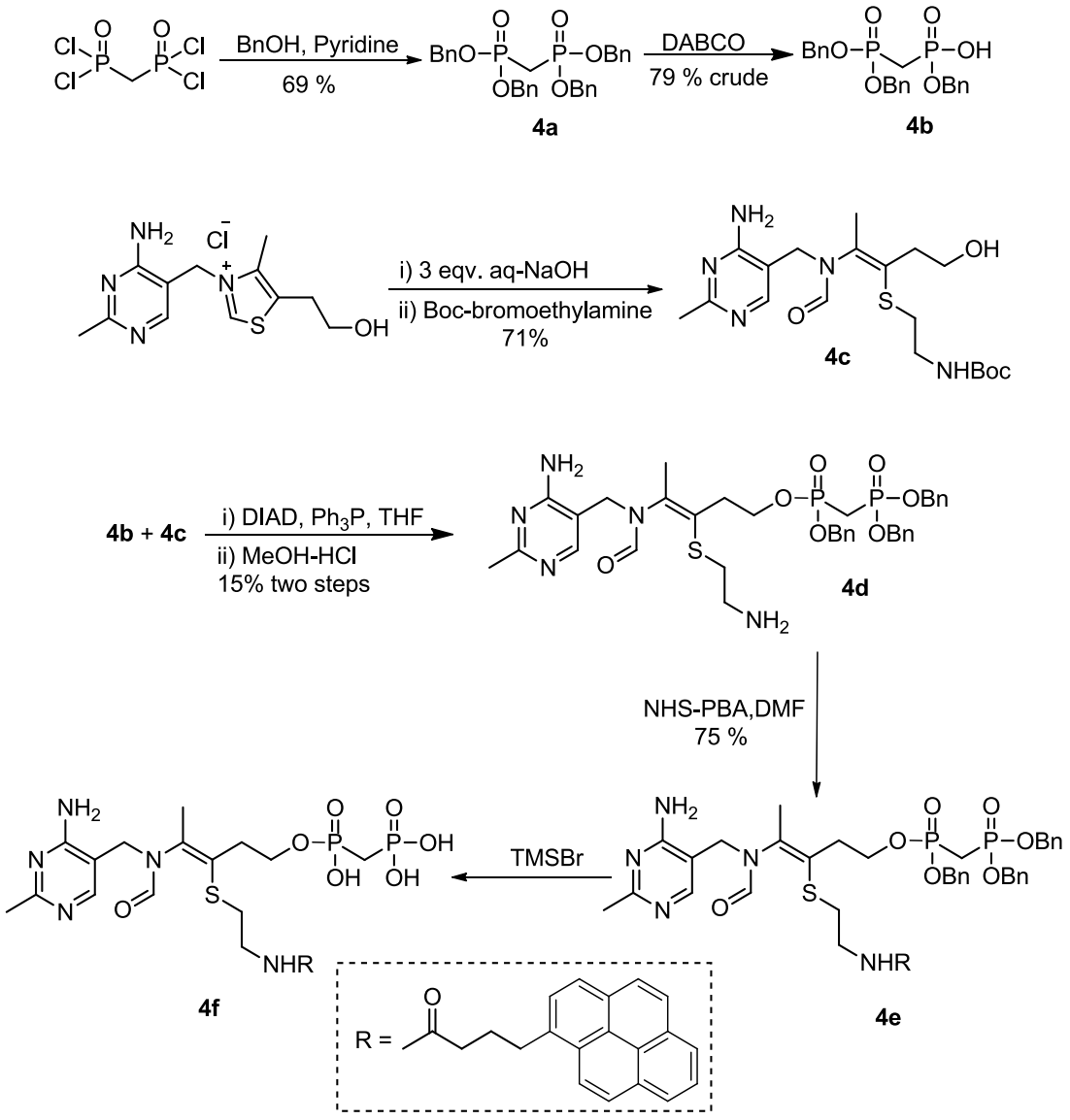
2.2.2 *Synthesis of TMP and TPP riboswitch fluorescent probe*



Scheme 2.3. Synthesis of pyrene-conjugated acyclic TMP probes

TPP and TMP fluorescent probes were prepared by modifying the central thiazole moiety. The thiazole ring of thiamine or TMP undergoes a reversible ring opening reaction in aqueous NaOH solution (pH 10) to form sodium a thiolate intermediate[57] to

which the pyrene fluorophore was attached. Studies have shown that TMP and its acyclic synthetic analog benfotiamine, bind to the *E. coli* thiM TPP riboswitch with increased affinity despite their lack of pyrophosphate (PP) group. Therefore, pyrene-derivatized acyclic TMP analogs **3a** and **3b** were synthesized by reacting the thiolate intermediate of TMP with PA-Br and PB-NHS, respectively (Scheme 2.3).



Scheme 2.4: Synthesis of pyrene-conjugated bisphosphonate analog of acyclic TPP

TPP could not be used as the starting material for TPP analogs as it readily decomposed to TMP in alkaline condition. Therefore, more stable methylene bisphosphonate analogs of TPP from thiamine were synthesized (Scheme 2.4). The reaction of methylenebis(phosphonic dichloride) with benzyl alcohol in the presence of pyridine gave **4a** (69%) which was purified by silica gel chromatography using 5% hexane in ethyl acetate.[58] Mono-deprotection of the benzyl group using 1,4-diazabicyclo[2.2.2]octane (DABCO) produced **4b** (79% crude).[59] The thiolate intermediate of thiamine formed under basic condition was alkylated with Boc-protected bromoethylamine to afford **4c** (71%) after silica gel purification using 1-4% methanol in chloroform. Mitsunobu reaction of **4b** and crude **4c** using diisopropyl azodicarboxylate (DIAD) with triphenylphosphine and Boc-deprotection in methnolic HCl gave **4d** (15% in two steps) after silica gel purification using 1-4% methanol in chloroform. The coupling of **4d** with NHS-PBA and silica gel purification using 1-3 % methanol in chloroform gave **4e** (75%). The debenzylaton of **4e** using bromotrimethylsilane (TMSBr) gave the TPP analog **4f** as only characterized by ESI-MS (Scheme 2.4).

2.2.3 *Construction of plasmid containing lysine and TPP riboswitches*

The genes *thiC* and *thiM* are among the twelve genes responsible for the biosynthesis of thiamine in prokaryotes; *thiC* is involved in the pyrimidine biosynthesis while *thiM* is a kinase gene.[52] In response to TPP, the *thiC* riboswitch represses the *thiC* gene by premature transcription termination. The *thiM* riboswitch controls the *thiM* gene at the translational level by sequestering SD sequence.[60] These two riboswitches were PCR

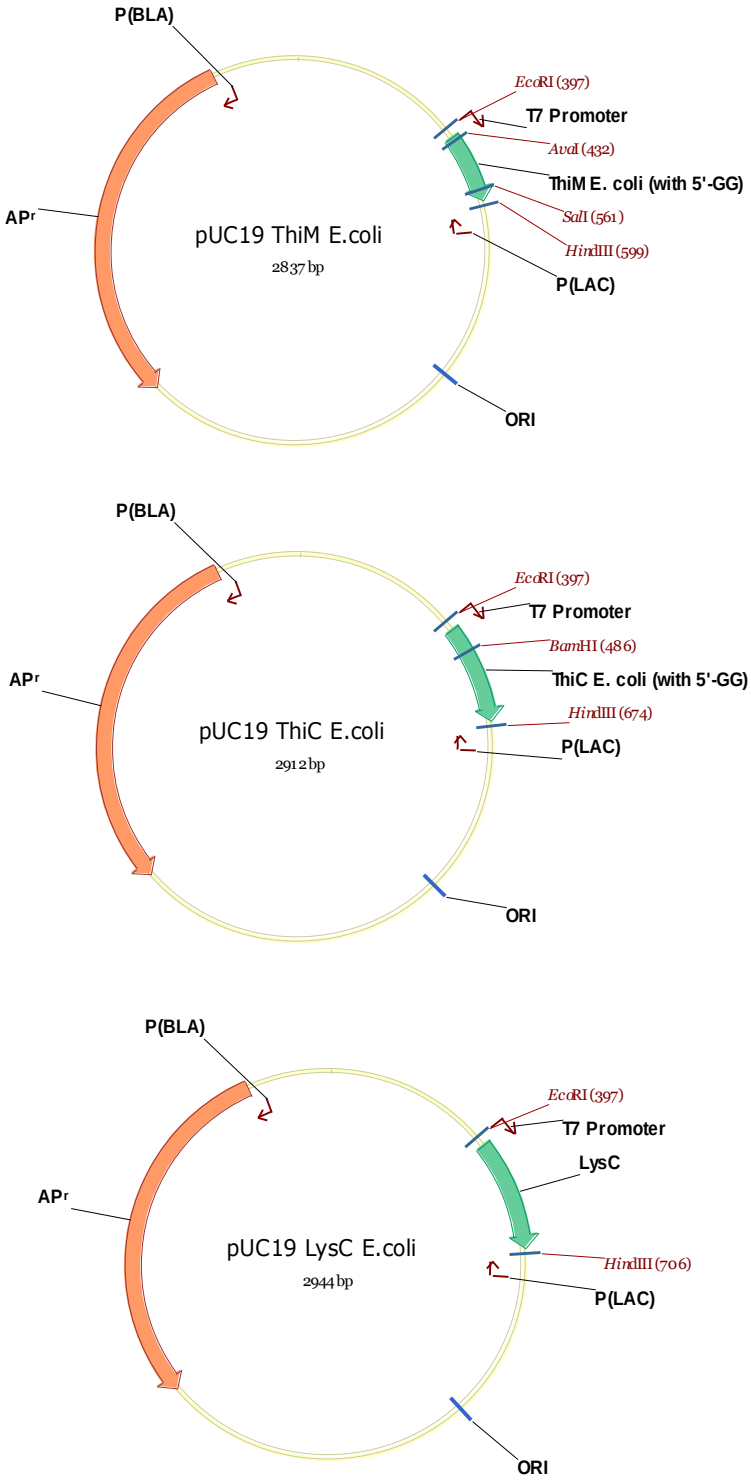


Figure 2.3 Diagrams of recombinant vectors **a**) pUC19-thiM, **b**) pUC19-thiC and **c**) pUC19-lysC showing the engineered restriction sites used for cloning the riboswitch genes into the pUC19 vector system

amplified from *E. coli* genomic DNA (Fig. 2.4). The PCR products were inserted into the pUC19 plasmid to obtain pUC19-thiC and pUC19-thiM, which were confirmed by DNA sequencing analysis (Fig. 2.3a and b). The *in vitro* transcription of these plasmids provided 163-base *thiM* and 238-base *thiC* TPP riboswitches (Fig 2.5).

Similarly, the aptamer domain of *lysC* was PCR amplified from *E. coli* genomic DNA (Fig 2.4) and was used to construct pUC19-lysC vector (Fig. 2.3c). Using this plasmid as the template, 271-base *lysC* lysine riboswitch was prepared by *in vitro* transcription (Fig 2.5).

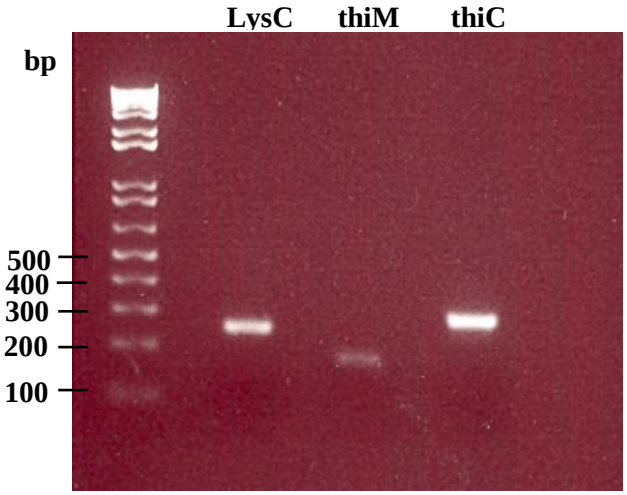


Figure 2.4 Gel electrophoresis analysis: 1 % agarose gel of the PCR products obtained after amplification using the genomic DNA as the template along with 1 kb Plus DNA marker as the reference.

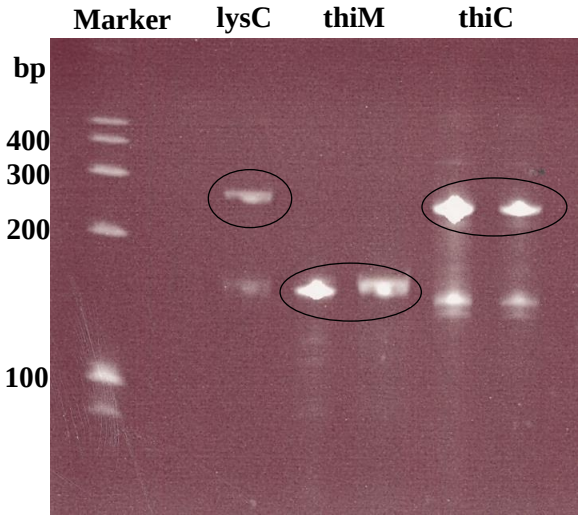


Figure 2.5 Gel electrophoresis analysis following *in vitro* transcription: 271-base *lysC* lysine riboswitch,163-base *thiM* and 238-base *thiC* TPP riboswitch transcripts visualized using 8% denaturing PAGE stained with ethidium bromide. RNA Century™ was used as the sized Marker.

2.2.4 *Fluorescence binding assays and initial problems*

To examine the binding affinities of synthesized probes, **2a**, **2b** and **2c** were incubated with the lysine riboswitch and the changes in the fluorescence intensity were monitored using a spectrofluorometer. Initial measurements showed decrease in the intensities of the fluorescence emission of the probes, indicating the binding with the *lysC* riboswitch. However, the decrease in intensity at different concentrations of the riboswitch was found to be unreliable. No correlation could be established between the emission intensity of the dye and the riboswitch concentration.

All the pyrene labeled probes exhibited the same issue. The false positive fluorescence quenching observed resulted from factors other than the presence of the RNA. It was observed that even the fluorescence emission intensity of the probes by

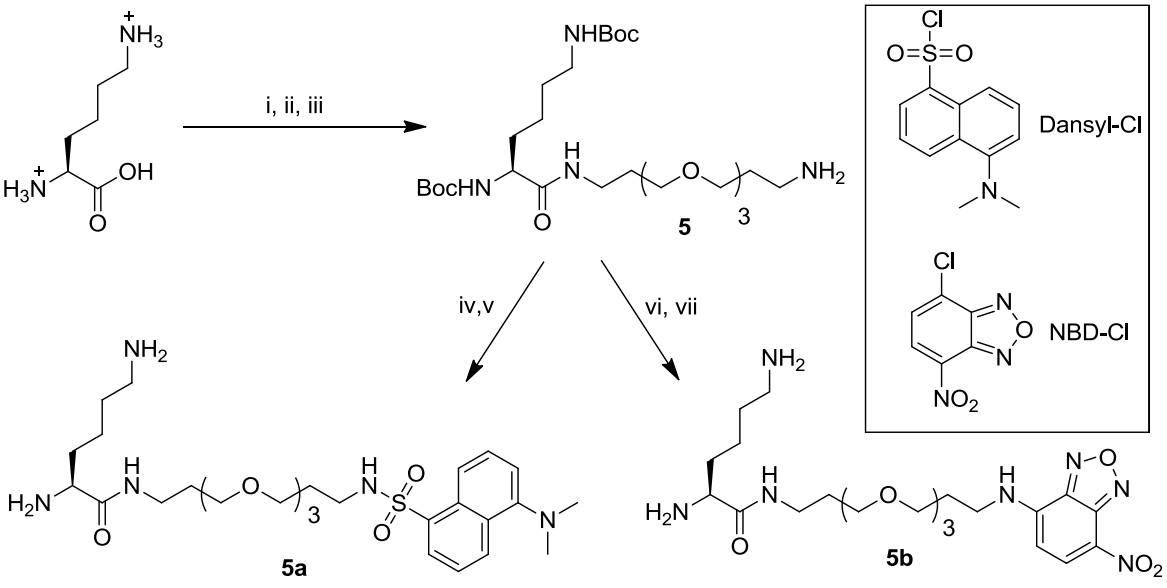
themselves in just buffer showed the quenching effect. Molecular oxygen is known to be an efficient fluorescence quencher of aromatic hydrocarbons.[61] Therefore, pyrene and its derivatives have been effectively used as a probe for measuring oxygen concentration and diffusion in living cells and artificial membranes.[62-63] The ability to quench a significant fraction of fluorescence using dissolved oxygen is due to long fluorescent life-time of pyrene and its derivatives. Their fluorescence lifetimes are reported to be in the range of 100-200 nsec.[64]

To eliminate the effect of dissolved oxygen in the solution, the samples were vacuum-degassed and also bubbled with nitrogen. Reagents such as pyrogallol, which is known to absorb oxygen in aqueous solution, were also used. The experiments were also performed in different types of buffer like Tris (tris(hydroxyemthyl)aminomethane), HEPS (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), MES (2-(N-morpholino)ethanesulfonic acid) to prevent the diminishing of the emission intensity. Despite, all these efforts, the unusual fluorescence quenching of the pyrene labeled probes could not be avoided.

2.2.5 *Synthesis of probes using dansyl and NBD groups*

To circumvent the quenching issues of pyrene-labeled probes, other environmentally sensitive fluorophores such as 5-(dimethylamino)naphthalene-1-sulfonyl (dansyl) and nitrobenzoxadiazole (NBD) were used to synthesize the riboswitch analogs. These relatively small dyes have effectively been used to label various biomolecules for a wide range of applications [65,66] without interfering with bio-activity.[67]

Lysine riboswitch probes were synthesized to examine the viability of using dansyl and NBD groups for riboswitch-ligand interactions (Scheme 2.5). Boc-protected NHS ester of lysine was prepared by first reacting lysine with di-*tert*-butyl-dicarbonate in presence of NaOH[56] and then TSTU. The Boc-protected NHS ester was then reacted with 4,7,10-trioxa-1,13-tridecane-diamine yielded **5** (69%) after silica gel purification using 1-4% methanol in chloroform. The subsequent N-alkylation of **5** using dansyl chloride and NBD-Cl followed by deprotection of the Boc groups using methanolic HCl provided lysine riboswitch probes **5a** and **5b,** respectively after purification using preparative TLC developed with 10% methanol in chloroform. The molar extinction coefficients were determined to be 200 M^{-1} cm^{-1} at 328 nm for **5a** and 2000 M^{-1} cm^{-1} at 480 nm for **5b**.



Scheme 2.5 Synthesis of lysine probes with NBD and dansyl groups: (i) Di-*tert*-butyl dicarbonate, NaOH 80%; (ii) TSTU, iPr_2NEt, DMF; (iii) 4,7-10-trioxa-1,13-tridecanediamine, Et_3N, DMF, 69% in two steps; (iv) Dansyl chloride, Et_3N, CH_2Cl_2, 50%; (v) HCl, MeOH, 82%; (vi) NBD-Cl, Et_3N, CH_2Cl_2, 30%; (vii) HCl, MeOH, 90%; inset) Structures of Dansyl chloride and NBD chloride.

2.2.6 *Fluorescence-based binding assay of **5a** and **5b** to the lysC riboswitch*

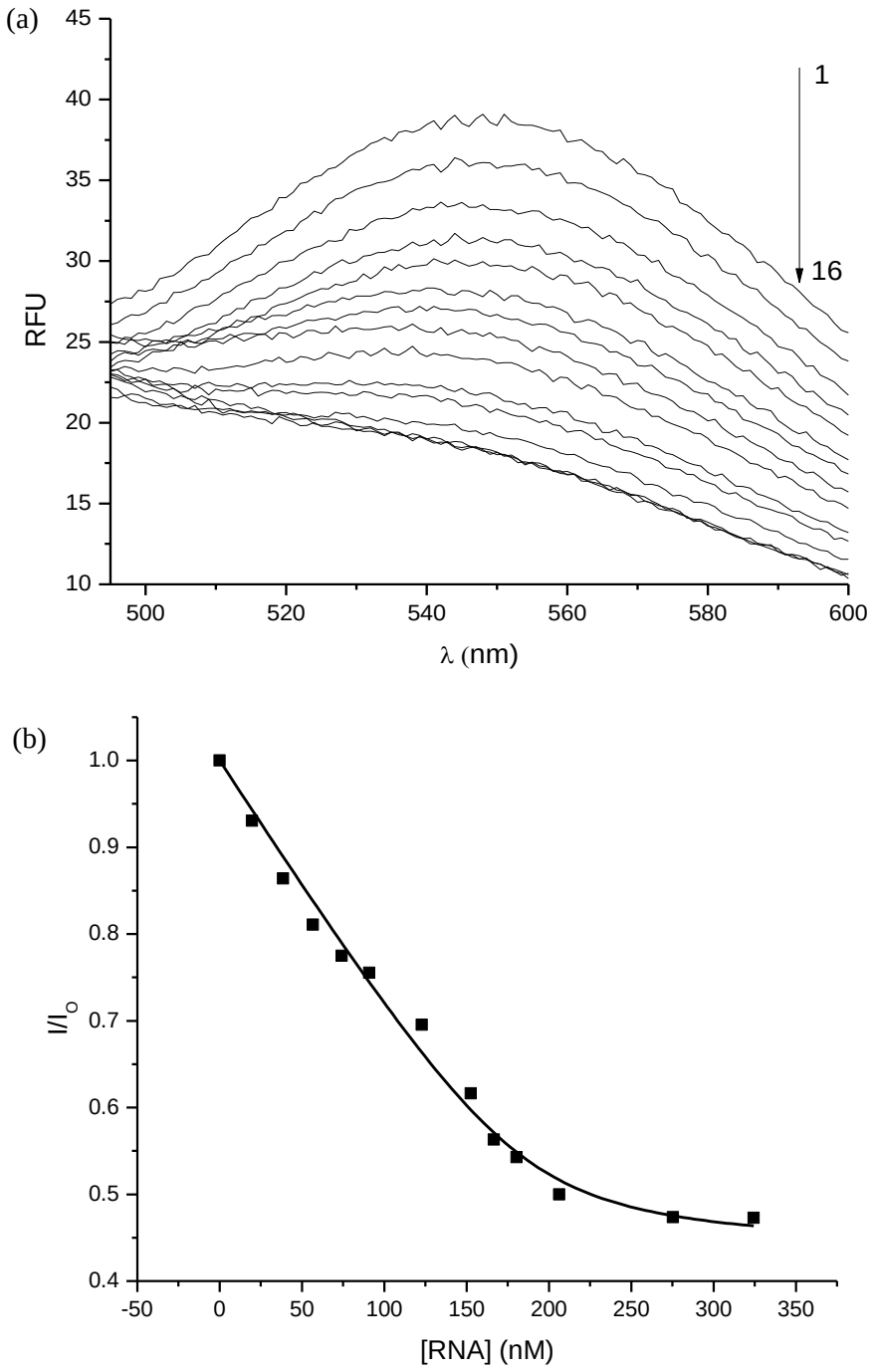


Figure 2.6 (a) Fluorescence intensity of 200 nM **5a** alone (1) and with 20 nM (2), 38 nM (3), 57 nM (4), 74 nM (5), 107 nM (6), 123 nM (7), 138 nM (8), 153 nM (9), 167 nM (10), 180 nM (11), 194 nM (12), 206 nM (13), 219 nM (14), 275 nM lysC (15) 324 nM (16) of the *lysC* riboswitch RNA; (b) Relative fluorescence of 200 nM **5a** as a function of increasing concentrations of the *lysC* riboswitch RNA.

The fluorescence intensity change of **5a** at 550 nm (excitation at 328 nm) was monitored at room temperature in the presence of varying concentrations of the *lysC* riboswitch to examine whether **5a** binds to *E. coli lysC* riboswitch *in vitro*. As shown in Figure 2.6(a), the emission intensities of 200 nM of **5a** consistently diminished with the increase in concentration of *lysC* riboswitch This fluorescence quenching effect indicates binding between the probe and the riboswitch with the formation of a complex between them.

A well-fitted curve for an equation derived for 1:1 complex formation was obtained on plotting relative fluorescence intensities against the concentration of the riboswitch (Fig. 2.6 b) with a dissociation constant K_d = 7.8± 6 nM, a value at least 25-fold lower than the reported dissociation constant for lysine (360 nM).[12] This result indicates an enhanced affinity of **5a** to the lysine riboswitch as compared to lysine. The enhancement in the binding affinity for this probe can be attributed to the diethylene glycol bis(3-aminopropyl) ether linker in between dansyl and lysine moieties of **5a**, which probably binds to the solvent-exposed hydrophilic pocket at the carboxyl end of lysine (Fig. 2.1) and enhances the overall binding affinity.

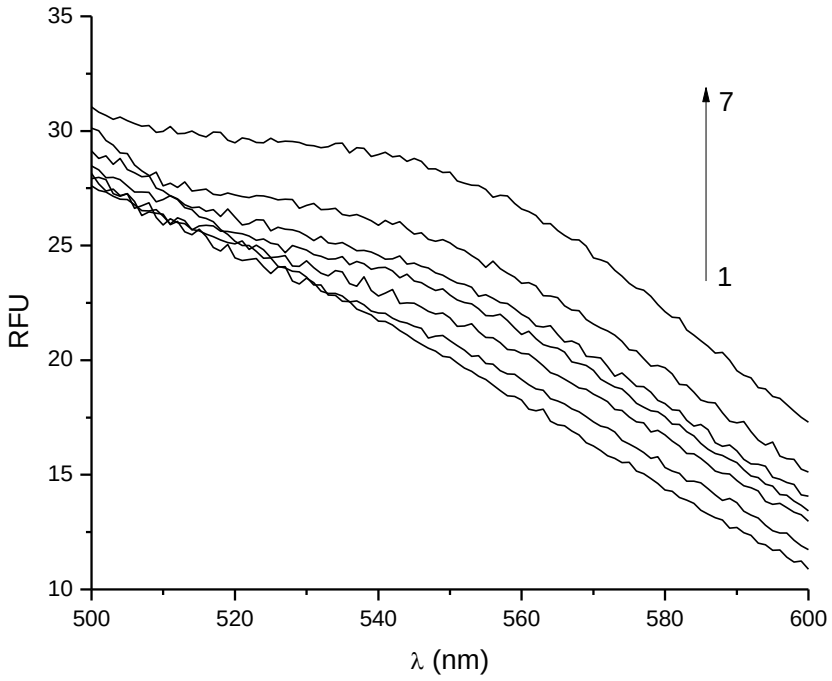


Figure 2.7 Fluorescence intensity of 200 nM **5a** in the presence of 500 nM the lysC riboswitch RNA (1) and with 571 nM (2), 721 nM (3), 888 nM (4), 1.0 μM (5), 1.1 μM (6), and 1.2 μM (7) of lysine.

Competition assay between lysine and **5a** was performed in presence of the riboswitch to assess if the binding of **5a** to the *lysC* riboswitch is selective and whether it shares the same binding site as lysine. As shown in Figure 2.7, the fluorescence intensity of 200 nM **5a** quenched by the lysine riboswitch was recovered by the addition of increasing concentration of lysine. This result indicates the existence of competitive affinity towards that lysine riboswitch between **5a** and lysine molecule.

Similarly, the selectivity of **5a** was also investigated by observing the fluorescence intensity change in the presence of different riboswitch other than *lysC*, in this instance the *E. coli thiM* riboswitch, which specifically recognizes thiamine pyrophosphate. As shown in Figure 2.8, relatively very little fluorescence intensity changes were observed for **5a** even in the presence of 5 molar excess of the *thiM*

riboswitch compared to *lysC* riboswitch. These results clearly suggest that **5a** does not non-specifically bind to *thiM* riboswitch and in doing so reveal the selectivity of **5a** towards lysine riboswitch.

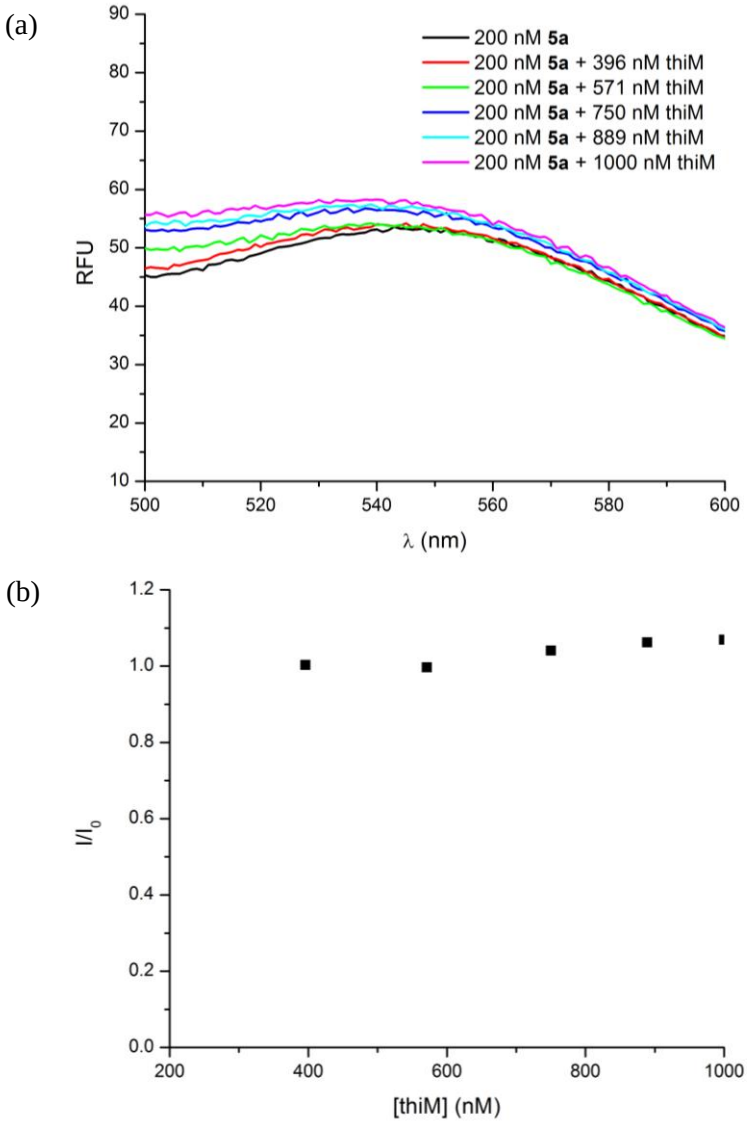


Figure 2.8 (a) Fluorescent intensity of 200 nM **5a** in the presence of increasing concentration of the *thiM* riboswitch RNA. (b) Relative fluorescent intensity of 200 nM **5a** as a function of increasing concentration of the *thiM* riboswitch RNA.

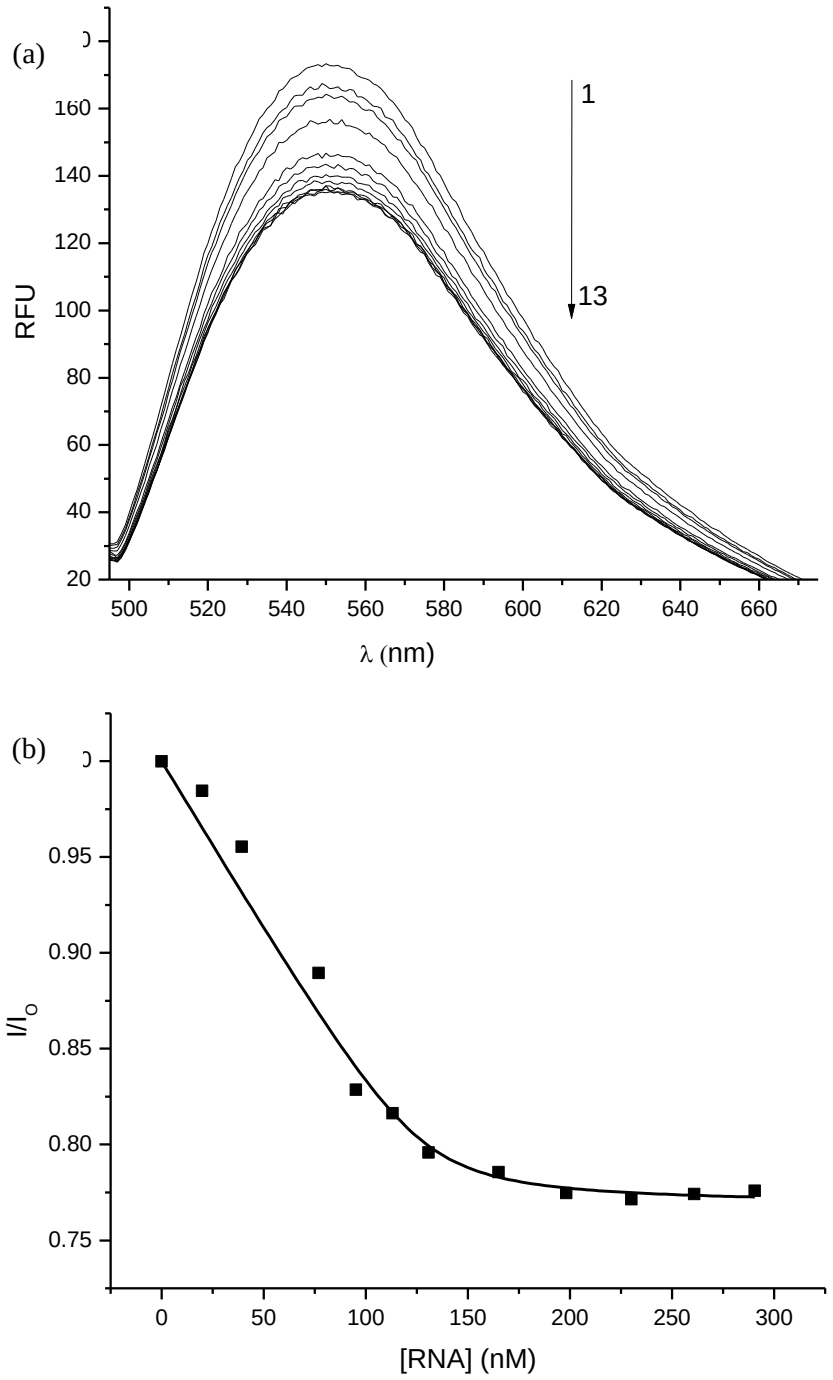


Figure 2.9 (a) Fluorescence intensity of 200 nM **5b** alone (1) and with 20 nM (2), 39 nM (3), 58 nM (4), 77 nM (5), 95 nM (6), 113 nM (7), 131 nM (8), 165 nM (9), 198 nM (10), 230 nM (11), 261 nM (12), and 291 nM (13) of the lysC riboswitch RNA. (b) Relative fluorescence intensity of 200 nM **5b** as a function of increasing concentrations of the lysC ribowitch RNA.

Due to relatively low sensitivity of **5a** as evident in Figure 2.6a, whose molar extinction coefficient at 328 nm was 200 M^{-1} cm^{-1}, NBD-conjugated lysine probe **5b** (molar extinction coefficient 2000 M^{-1} cm^{-1} at 480 nm) was used to re-evaluate the binding analysis. On comparing the relative fluorescent intensities, we observed that **5b** exhibited relatively better sensitivity than **5a** (Figure 2.6a and 2.9a).

The change in fluorescence emission intensities in presence of the riboswitch was monitored to examine if **5b** also binds to the *lysC* riboswitch RNA. This change in fluorescent intensities of **5b** in the presence of riboswitch RNA was assessed using the excitation wavelength of 480 nm and emission at 550 nm. Similar to **5a**, the emission intensity of **5b** also decreased with the addition of increasing amount of the *lysC* (Figure 2.9a), indicating the binding of **5b** to *lysC* riboswitch RNA.

A well-fitted curve for an equation derived for 1:1 binding stoichiometry was obtained on plotting relative fluorescence intensity of **5b** against the concentration of the riboswitch with a dissociation constant K_d = 3.0± 2 nM, a value at relatively similar to that of **5a** (Figure 2.9b). This result indicates an enhanced affinity of **5b** to the lysine riboswitch compared to lysine (K_d = 360 nM). As explained before, the enhancement in the binding affinity for this probe can also be attributed to the diethyleneglycol bis(3-aminopropyl)ether linker in between NBD and lysine moieties of **5b**.

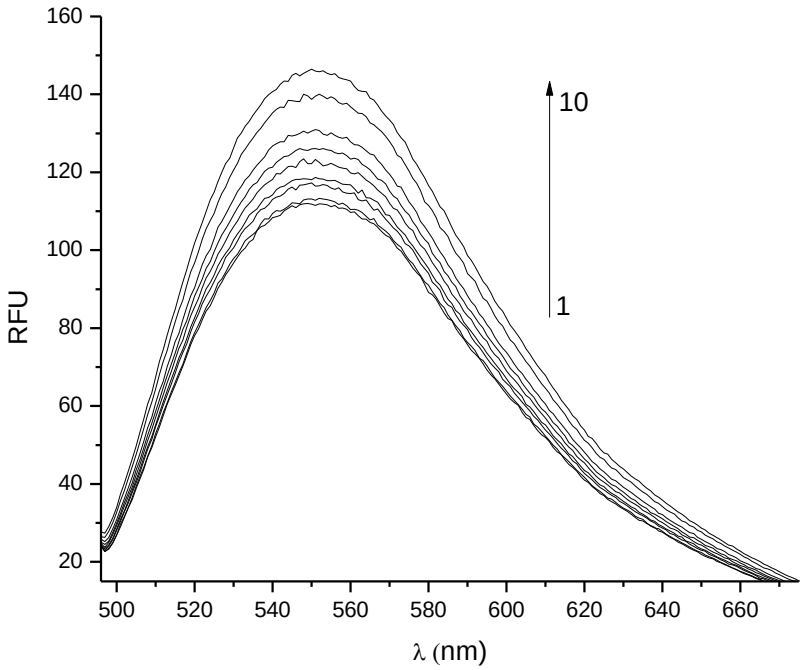


Figure 2.10 Fluorescence intensity of 200 nM **5b** in the presence of 500 nM lysC riboswitch RNA (1) and with 95 nM (2), 182 nM (3), 333 nM (4), 571 nM (5), 750 nM (6), 889 nM (7), 1.0 μM lysine (8), 1.1 μM (9), and 1.2 μM (10) of lysine.

The quenched fluorescence of **5b** in the presence of 500 nM *lysC* riboswitch RNA was recovered by the addition of increasing concentration of lysine which clearly indicates that **5b** also competes with lysine for binding to the *lysC* riboswitch RNA (Figure 2.10). In the presence of 1.2 μM lysine, the fluorescence intensity of **5b** is almost completely recovered to the level of the probe itself. This result confirms that **5b** binds to the same lysine binding pocket within the lysine riboswitch RNA.

The fluorescence quenching effect of **5b** was monitored in the presence of *E. coli thiM* riboswitch RNA to examine the selectivity of **5b** towards the *lysC* riboswitch RNA,. As shown in Figure 2.11a and b, the fluorescence intensities of **5b** changed very little with the increase in the concentration of *thiM* riboswitch even up to 1 μM. This result

clearly shows that **5b** does not bind non-specifically to any RNA other than the *lysC* riboswitch RNA.

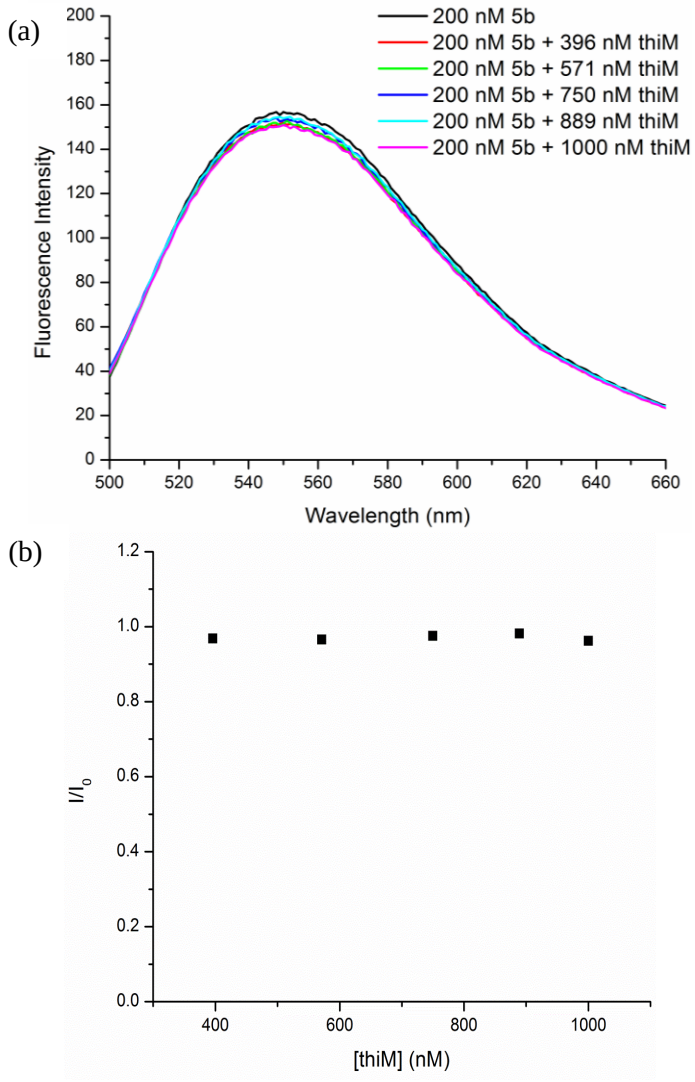


Figure 2.11 (a) Fluorescence intensity of 200 nM **5b** in the presence of increasing concentration of the *thiM* riboswitch RNA. (b) Relative fluorescence intenisty of 200 nM **5b** as a function of increasing concentraions of the *thiM* riboswitch RNA.

2.2.7 *Investigation of bacterial growth inhibition using the lysC riboswitch probes **5a** and **5b***

Indicated by their low K_d values, **5a** and **5b** could act as potential bacterial growth inhibitors. In comparison, two lysine analogs, L-aminoethylcysteine (AEC) and DL-4-oxalysine, which are known antibacterial agents have a K_d values of ~30 µM and 13±2 µM, respectively.[12] This indicates that both **5a** and **5b** should bind the lysine riboswitch aptamer with significantly greater affinity than both AEC and DL-4-Oxalysine and can inhibit bacterial growth by riboswitch mediated gene expression mechanism.

MG1655 *E. coli* cells were incubated on M9 minimal medial plates containing 1 mM **5a** or **5b** to investigate whether **5a** and **5b** could in fact inhibit bacterial growth. On incubating the cells at 37 °C for 24 hours, no inhibition of cell growth was observed, probably because the probes were not transported into the cell. Evidently, *E. coli* colonies on the media plates were not fluorescent while the media plate displayed fluorescence from **5a** and **5b**. Therefore, *in vivo* studies of these molecules were not further pursued.

2.3 Conclusion

In summary, various fluorescent probes for both lysine and TPP riboswitches were rationally designed and synthesized. For some unknown reason, the pyrene-based probes displayed unusual quenching most probably due to molecular oxygen quenching. Therefore, they could not be effectively used for riboswitch binding analysis. However, both dansyl and NBD derivatives of lysine were successfully used to monitor the binding

interaction and measure the binding affinity between the probes and the riboswitch RNA. These probes will be useful for high-throughput identification of new riboswitch modulators from small molecule libraries and for the analysis of cellular metabolites in complex chemical and biological samples *in vitro*.

2.4 Experimental methods

2.4.1 Media and reagents

LB (BD Difco™) and M9 minimal media (Sigma) were prepared according to the manufacturer's standard procedures. M9 minimal medium was supplemented with 0.8 % glycerol (Fisher Scientific) as the carbon source and 0.1 % casamino acids (BD Difco™). LB Agar (BD Difco™) was added at 3.7 % to prepare agar plates. XL1-Blue *E. coli* cells (Stratagene) were used for cloning and maintaining plasmids. Agar plates and LB culture were supplemented with carbenicillin (100 µg/ml Apollo Scientific Ltd.). Plasmids were isolated and purified using QIAprep Spin Miniprep Kit (Qiagen). DNA sequencing analysis was performed using Applied Biosystems 3130xl Genetic Analyzer (Hitachi). All enzymes and biological reagents were purchased from New England Biolabs unless indicated otherwise. *Taq* DNA polymerase was used for polymerase chain reaction (PCR) while T7 RNA polymerase was used for the *in vitro* transcription. Synthetic oligonucleotides were obtained from IDT (Integrated DNA Technologies). DNA and RNA concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Fish Scientific).

2.4.2 Preparation of the E. coli lysC riboswitch DNA

The PCR of the *E. coli* genomic DNA with The primer pair:

> LysC-F (5'-AGC TGA ATT C*TA ATA CGA CTC ACT ATA* GGT ACT ACC TGC GCT AGC GC-3') and

LysC-R (5’-CAT GAA GCT TGG ATG GAT CAC CTG GGC ACA AGG GAA GAG CGG-3’)

generated the 271-base *E. coli lysC* riboswitch gene under control of T7 promoter (italic in the primer LysC-F) along with the EcoRI and HindIII restriction sites (underlined in the primers LysC-F and LysC-R, respectively) at the 5’ and 3’-ends, respectively. The PCR product was inserted between the EcoRI and HindIII restriction sites of pUC19 to prepare pUC19-lysC. The plasmid was verified by sequencing. The linear DNA template for *in vitro* transcription was prepared by PCR using pUC19-lysC as the template and two primers:

T7-F (5’-TAA TAC GAC TCA CTA TAG GG-3’) and

LR (5’-GGC ACA AGG GAA GAG CGG-3’)

2.4.3 Preparation of the E. coli thiC and thiM riboswitch DNA

Similar strategy was employed to generate the *thiC* and *thiM* riboswitch DNA. The PCR of the genomic *E. coli* was performed using the following primers:

thiC-F (5’-AGC TGA ATT C*TA ATA CGA CTC ACT ATA* GGA ATG CCC CAT TTG CGG GGC-3’) and

thiC-R (5’-CAT GAA GCT TGG ATG GAT CAC CTG CGC GTT GTT CGC GGC GGG-3’) to generate the *E coli. thiC* riboswitch gene. The primer pair:

thiM-F (5’-AGC TGA ATT CTA ATA CGA CTC ACT ATA GGA ACC AAA CGA CTC GGG G-3’) and

thiM-R (5’-CAT GAA GCT TGG ATG GAT CAC CTG TTG CGC TGA ACC CAG CAG G-3’) to generate the *E. coli thiM* riboswitch gene.

EcoRI and HindIII restriction sites are underlined and the T7 promoter sequence italicized as in primers for *lysC* riboswitch. The PCR products obtained were inserted between the two restriction sites of pUC19 to generate the pUC19-thiM and pUC19-thiC. These plasmids were amplified, isolated and verified in the same manner as the pUC19-lysC. These plasmids were used as the template to obtain the linear DNAs for *in vitro* transcription by PCR with primers:

T7-F (5'-TAA TAC GAC TCA CTA TAG GG-3') and

TC-R (5'-CGC GTT GTT CGC GGC GGG -3') for *thiC* riboswitch.

T7-F (5'-TAA TAC GAC TCA CTA TAG GG-3') and

TM-R (5'-TTG CGC TGA ACC CAG CAG G -3') for *thiM* riboswithc.

2.4.4 Preparation and purification of the lysC riboswitch

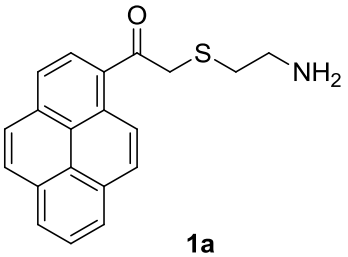
The transcription reaction (200 µL) was carried out in a buffer containing 40 mM Tris-HCl (pH 7.9), 2 mM spermidine, 14 mM $MgCl_2$, 10 mM dithiothreitol, 4 mM each of ATP, UTP, GTP and CTP, 160 U RNase inhibitor, 2 U yeast inorganic pyrophosphatase (Sigma), 3.6 µg DNA template and 200 U T7 RNA polymerase. The reaction mixture was incubated at 37 °C for 7 h and quenched by the addition of 16 U RNAse-free DNAse I and 80 mM $CaCl_2$. After the incubation at 37 °C for 1h, the reaction was stopped with the addition of 100 mM EDTA. The crude transcript was obtained by ethanol precipitation and purified on 8% preparative denaturing polyacrylamide gel. The RNA band was visualized by UV-shadowing and cut from the gel. The finely minced gel slices were extracted with 6 mL 100 mM sodium acetate (pH 4.5) containing 1 mM EDTA and

0.1% SDS overnight and filtered. The filtrate was concentrated by successive extraction with n-butanol. Ethanol precipitation provided the pure *lysC* riboswitch (82 µg) which was dissolved in DEPC-treated water and stored at -20 °C.

2.4.5 General synthetic methods

All commercial chemicals were used without further purification. Nuclear magnetic resonance (NMR) spectra were recorded on a Varian INOVA 400 spectrometer. Chemical shifts are reported in δ ppm and coupling constants (*J*) are reported in Hz. Assignments of ^{1}H resonances were assisted by the COSY spectral data. High resolution mass spectrometry (ESI-HRMS) data were obtained by using the Fourier-transform ion cyclotron resonance (FTICR) operating in tandem with an LTQ linear ion trap mass spectrometer equipped with an electrospray ionization (ESI) source (LTQ-FT, Thermo).

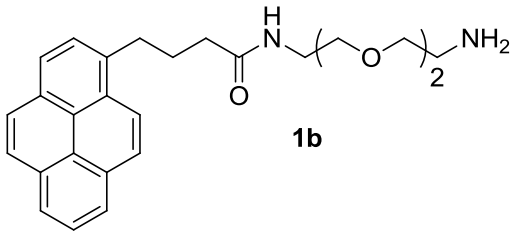
*Synthesis of pyrene-labeled liners **1a***



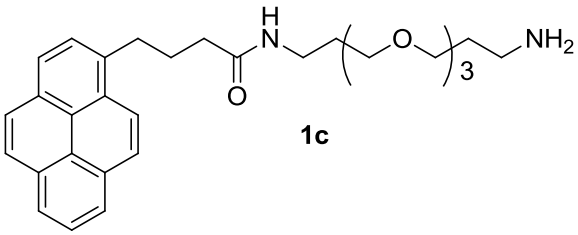
1a

Aminothanethiol·HCl (0.113 g, 1 mmol) dissolved in DMF (3 ml) in the presence of DIPEA (0.522 ml, 3 mmol) was treated with 1-(bromoacetyl)pyrene (0.323 g, 1mmol) The reaction was left overnight. TLC in methanol/chloroform (10% v/v) showed the completion of the reaction. The reaction was washed with saturated $NaHCO_3$ and water. The organic layer was dried using Na_2SO_4. The solvent was removed *in vacuo* to give a yellow crude product (36%) which was subsequently used for the next reaction without purification.

*Synthesis of pyrene-labeled liners **1b*** and ***1c***



1b



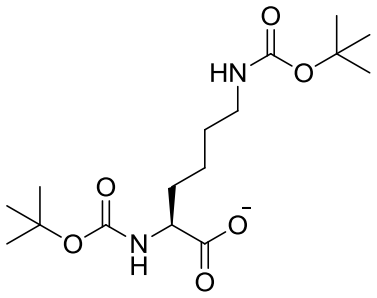
1c

To a solution of pyrenebutyric acid (200 mg, 0.69 mmol) and triethylamine (586 µl, 4.2. mmol) in DMF (2 ml), (O-(N-Succinimidyl)-1,1,3,3-tetramethyl uronium tetrafluoroborate) (TSTU) (340 mg, 1.1 mmol) was added. The reaction completion was monitored using TLC in 1:1 hexane/ethyl acetate. The reaction mixture was split into two halves. To the first half, 2,2'-(ethylenedioxy)bis(ethylamine) (761 µl, 5.19 mmol) was added and to the second half 4,7,10-trioxa-1,13-tridecanediamine (1.14 ml, 5.2 mmol) was added. Both the reactions were left stirring overnight at room temperature. Purification using preparative TLC with 10:1 $CHCl_3/CH_3OH$ as the elution solvent provided **1b** and **1c** (42% and 52% respectively). NMR spectroscopy and ESI-MS data were consistent with literature.[68]

Synthesis of Boc-protection of Lysine

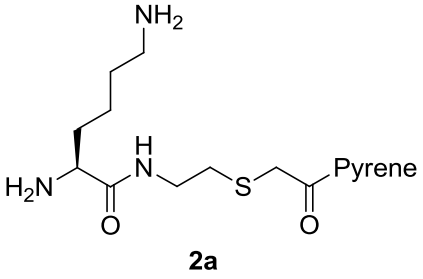
Lysine·HCl (3.65 g, 20 mmol) was dissolved in 2M NaOH (40 ml) to which dioxane (60 ml) was added. To this stirring mixture of Boc-anhydride (8.73 g, 40 mmol) was added. The reaction was stirred at room temperature for 5 hrs. On completion of the reaction as indicated by the absence of lysine on TLC stained with phosphomolybdic acid (PMA) water (50 ml) was added to the reaction mixture and washed with dichloromethane (2X

50 ml). The aqueous layer was cooled in ice bath and the pH was brought to 2 by adding 10% aqueous citric acid. The solution was extracted with dichloromethane (2X 50 ml) and the combined organic layer was dried using $MgSO_4$. Removal of the solvent under reduced pressure gave oily product (5.5 g, 80%). NMR spectra were consistent with literature.[56]



*Synthesis of **2a***

To a stirred solution of Boc-protected lysine (0.346 g, 1 mmol) in DMF (3 ml) and DIPEA (0.522 ml, 3 mmol), TSTU (0.451 g, 1.5 mmol) was added. On completion of reaction molar equivalent amount of **1a** was added to the reaction mixture and reaction was left overnight at room temperature. Solvent was removed under reduced pressure and silica gel chromatography (1:1 hexane/ethyl acetate) was used for purification. Boc-deprotection in methonolic-HCl (2M) gave yellowish oily **2a** (73 mg. 16 % in 3 steps). **2a** was characterized with ESI-MS but was not used for the fluorescent-based binding assay due to the fluorescence quenching of pyrene.

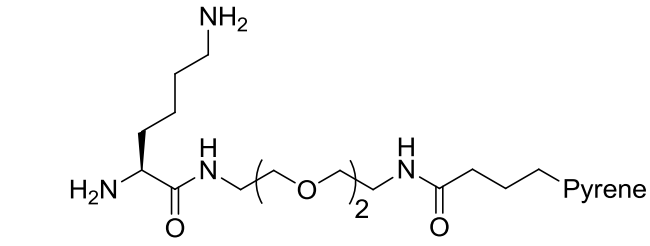


2a

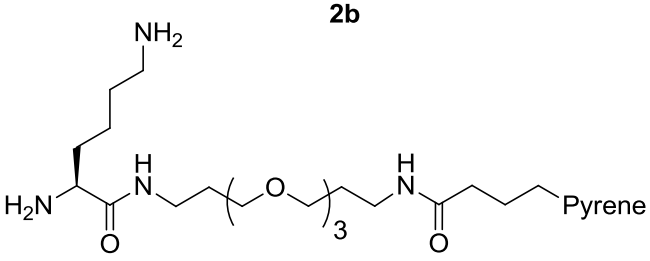
*Synthesis of **2b** and **2c***

Boc-protected lysine-NHS-ester was prepared by reacting Boc-protected lysine (346 mg, 1 mmol) in DMF (3 ml) and DIPEA (0.522 ml, 3 mmol) with TSTU (0.361 g, 1.2 mmol). 168 µl of NHS ester of Boc-protected lysine reaction mixture were added to a solution of

1b (20 mg, 0.047 mmol) in DMF (1 ml) and DIPEA (26 µl, 0.14 mmol). 143 µl of NHS ester of Boc-protected lysine reaction mixture were added to a solution of **1c** (20 mg, 0.041 mmol) in DMF (1 ml) and DIPEA (26 µl, 0.14 mmol). The products were purified using preparative TLC with 10:1 $CHCl_3$-CH_3OH as the elution solvent. On subsequent overnight deprotection of boc group using Methanolic HCl (1 ml) gave yellowish oily product **2b (**11 mg, 42.5 %) and **2c** (13 mg, 51.2%). **2b** and **2c** were characterized with ESI-MS but were not used for the fluorescent-based binding assay due to the fluorescence quenching of pyrene.



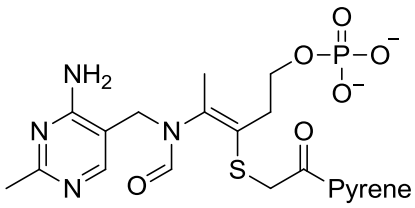
2b



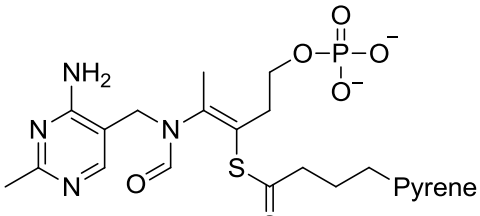
2c

*Synthesis of **3a** and **3b***

To a stirring solution of TMP (34.5 mg, 0.1 mmol) dissolved in 1 M aq. NaOH (156 µl) a solution of 1- (bromoacetyl)pyrene (32.3 mg of 0.1 mmol) in DMF (400 µl) was added. **3a** with formed with quantitative yield under less the 10 minutes as monitored by ESI-MS and TLC. Similarly, to a stirring solution of TMP (16.2 mg, 0.038 mmol) in of 1 M NaOH (156 µl) a solution of PB-NHS ester (15 mg) in DMSO (200 µl) was added. **3b** was obtained with



3a



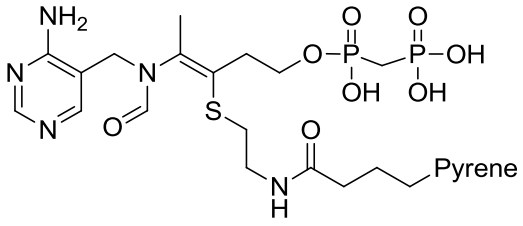
3b

quantitative yield in less than 10 minutes as monitored by TLC and ESI-MS. **3a** and **3b** was characterized with ESI-MS but were not used for the fluorescent-based binding assay due to the fluorescence quenching of pyrene.

*Synthesis of **4f***

To a stirring solution of thiamine HCl (1.26 g, 3.74 mmol) in 1 M NaOH (11.2 ml), a solution of 0.838 g (3.74 mmol) Boc-protected aminoethane bromide[69] (0.83 g, 3.74 mmol) in acetonitrile (7 ml) was added stirred for 6 hrs. On completion of the reaction, the solvent was removed under reduced pressure and the reaction mixture was extracted with dichloromethane. The organic layer was dried using $MgSO_4$. On purification using silica gel chromatography (1-4 % MeOH in $CHCl_3$) white crystal of **4c** (1.32 g, 71%) was obtained.

To a stirring solution of **4b** (0.2 g, 0.44 mmol), Diisopropyl azodicarboxylate (0.2 g, 0.47 mmol) and triphenylphosphine (and 0.122 g, 0.47 mmol) in 3 ml of THF at 0 °C, a solution of **4c** (0.092 ml, 0.47 mmol) in 1.5 ml of toluene was added over 40 min. The reaction was left stirring at room temperature for 17 hours. Silica gel chromatography using 1-4% of MeOH in $CHCl_3$ was used to purify the product **4d** (60 mg, 15% in two steps). **4d** on de-protection using methanolic-HLC and reaction with N-hydroxylsuccimide pyrene butyric acid (1.2 molar equivalent) in DMF (1ml) and DIPEA (3 molar equivalent) yielded **4e** (54 mg, 74%) after silica gel purification using 1-3% of MeOH in $CHCl_3$. The treatment of **4e** with

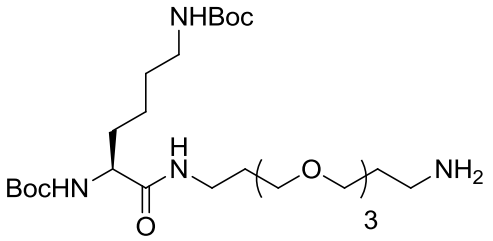


4f

bromotrimethylsilane (TMSBr) gave **4f**. Product confirmed by ESI-MS. **4f** was characterized with ESI-MS but was not used for the fluorescent-based binding assay due to the fluorescence quenching of pyrene.

*Synthesis of **5***

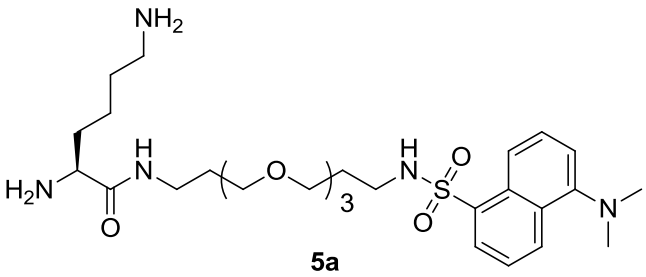
A solution of Boc-Lys(Boc)-OH (0.524 g, 1.51 mmol), N,N-diisopropylethylamine (1 mL, 5.7 mmol) and 2-succinimido-1,1,3,3-tetramethyluronium tetrafluoroborate (0.455 g, 1.51 mmol) in DMF (5 mL) was stirred at room temperature for 4 h. To this solution at 0 °C 4,7,10-trioxa-1,13-tridecanediamine (0.331 µL, 1.51 mmol) in DMF (3 mL) was added dropwise. The reaction mixture was stirred at room temperature overnight and evaporated to give an oily residue, which was purified by silica-gel column chromatography with 1-4% methanol in chloroform to give **5a** (0.554 g, 69 %). 1H NMR (300 MHz, $CDCl_3$) δ 8.17 (bs, 1H), 7.11 (bs, 2H), 5.59 (bd, J=7.2, 1H), 4.88 (bs, 1H), 4.02 (1H, bs), 3.75 (m, 2H), 3.65 (m, 12H), 3.51 (m, 2H), 3.30 (m, 2H), 3.10 (m, 2H), 2.65 (m, 2H), 1.98 (m, 2H), 1.76 (m, 4H), 1.52 (m, 2H), and 1.58-1.30 (m, 22H). MS (ESI) m/z= 549.3 $[M+H]^+$.



5

*Synthesis of **5a***

A mixture of **5** (50 mg, 0.09 mmol), triethylamine (53 µl, 38 mmol) and dansyl chloride (24.3 mg, 0.09 mmol) in CH_2Cl_2 (0.5 ml) was stirred



5a

overnight at room temperature and evaporated. The residue was purified using preparative TLC with 10% methanol in chloroform to give the fully-protected intermediate (35 mg, 50%). A mixture of the intermediate (17 mg, 0.022 mmol) in 1.25 M HCl in methanol (1 mL) was stirred and evaporated to give **5b** (10.6 mg, 82%). ^{1}H NMR (400 MHz, D_2O) δ 8.73 (d, J = 8.9, 1H, $C^{Ar}H$), 8.44 (d, J= 8.8, 1H, $C^{Ar}H$), 8.36 (d, J= 7.4, 1H, $C^{Ar}H$), 8.05 (d, J= 8.1, 1H, $C^{Ar}H$), 7.89 (dd, J=7.4, 8.9, 1H, $C^{Ar}H$), 7.89 (dd, J= 8.1, 8.8, 1H, $C^{Ar}H$), 3.92 (t, J= 6.7, 1H, C□*H*), 3.58 (s, 6H, N*Me*$_2$), 3.51 (t, J= 6.5, 2H, C*H*$_2$O), 3.30 (m, 12H, 5 C*H*$_2$O and NC*H*$_2$), 2.99 (m, 4H, NC*H*$_2$ and $C^{\varepsilon}H_2$), 1.88 (m, 2H, $C^{\beta}H_2$), 1.76 (m, 2H, C*H*$_2$), 1.66 (m, 2H, $C^{\delta}H_2$), 1.57 (m, 2H, C*H*$_2$), and 1.41 (m, 2H, $C^{\gamma}H_2$). ^{13}C NMR (100 MHz, D_2O) δ 169.3 (C=O), 139.0, 134.9, 130.5, 128.6, 128.0, 126.7, 126.0, 125.8, 125.7, 119.2 (10 naphtyl C), 69.4, 69.3, 69.2, 68.9, 68.1, 67.4 (6 *C*H$_2$O), 53.0 (*C*H), 46.5 (2 *C*H$_3$), 39.4, 38.8, 36.4 (2 NCH$_2$ and C^{ε}), 30.2 (C^{β}), 28.2, 28.0 (2 *C*H$_2$), 26.2 (C^{δ})and 21.2 (C^{γ}). The other four aromatic carbons of the NBD group were not shown in the ^{13}C NMR spectrum even with a good signal to noise ratio after 25,000 scans probably due to very fast quadrupole relaxation caused by neighboring nitrogen atoms. HRMS (ESI) calculated for $C_{28}H_{48}N_5O_6S^+$ $[M+H]^+$: 582.3320; found 582.3318.

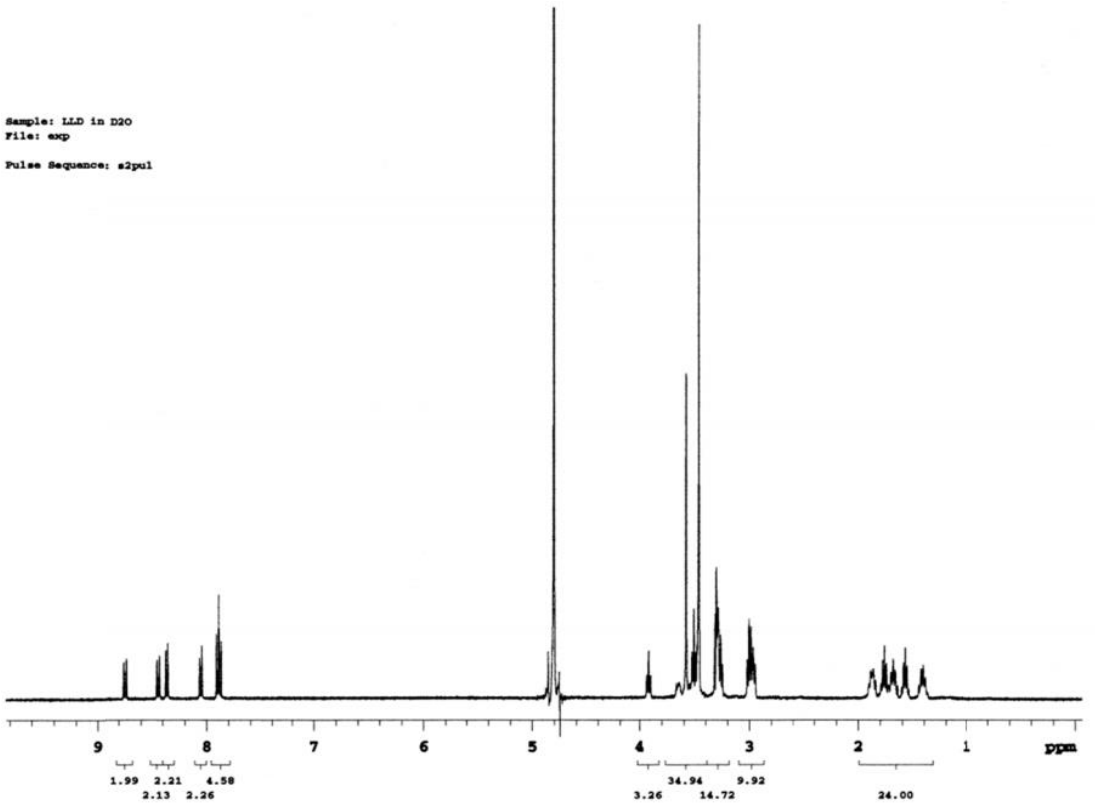


Figure 2.12 ^{1}H NMR of **5a**

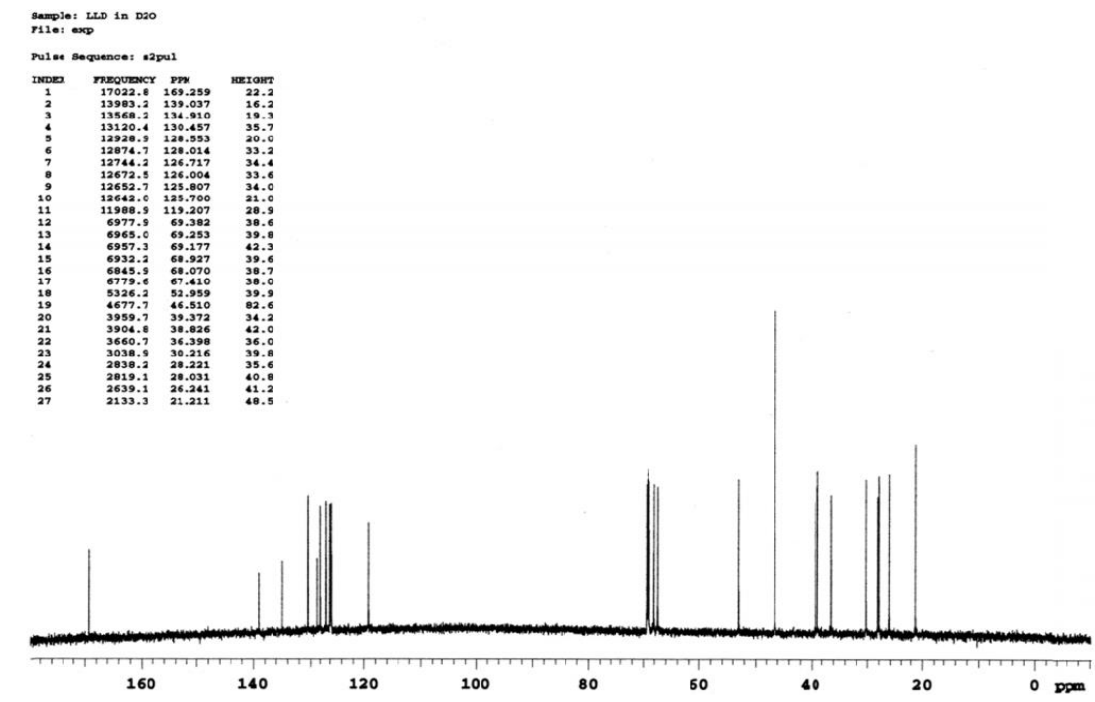
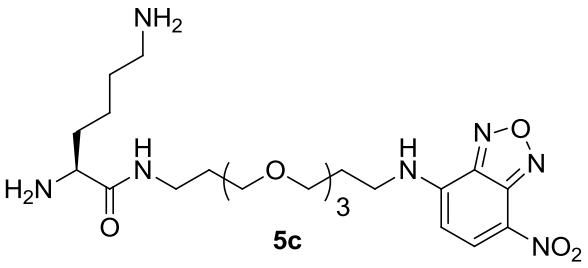


Figure 2.13 ^{13}C NMR of **5a**

*Synthesis of **5c***



5c

A mixture of **5a** (44 mg, 0.08 mmol), triethylamine (48 µl, 34 mmol) and 4-chloro-7- nitrobenzofurazan (15.9 mg, 0.08 mmol) in CH_2Cl_2 (0.5 ml) was stirred overnight at room temperature and evaporated. The residue was purified using preparative TLC with 10% methanol in chloroform to give the fully-protected intermediate (17 mg, 30%). A mixture of the intermediate (8 mg, 0.011 mmol) in 1.25 M HCl in methanol (1 mL) was stirred and evaporated to give **5c** (5.3 mg, 91%). ^{1}H NMR (400 MHz, D_2O) δ 8.42 (d, J= 9.2, 1H, C^6H), 6.33 (d, J= 9.2, 1H, C^5H), 3.91 (t, J= 6.8, 1H, $C^\alpha H$), 3.68-3.58 (m, 12H, 6 CH_2O), 3.49 (t, J= 6.4, 2H, NCH_2), 3.24 (t, J= 7.0, 2H, NCH_2), 2.97 (t, J= 7.6, 2H, $C^\varepsilon H_2$), 2.03 (m, 2H, CH_2), 1.86 (m, 2H, $C^\beta H_2$), 1.76 (m, 2H, CH_2), 1.66 (m, 2H, $C^\delta H_2$), and 1.41 (m, 2H, $C^\gamma H_2$). ^{13}C NMR (100 MHz, D_2O) 169.7 (C=O), 139.2 (C^6), 99.8 (C^5), 69.5, 69.5, 69.3, 69.2, 68.1, 68.1 (6 CH_2O), 53.0 (CH), 40.7, 38.8, 36.3 (2 NCH_2 and C^ε), 30.4 (C^β), 28.0, 27.3 (2 CH_2), 26.2 (C^δ) and 21.2 (C^γ). HRMS (ESI) calculated for $C_{22}H_{38}N_7O_7^+$ $[M+H]^+$: 512.2827; found 512.2820.

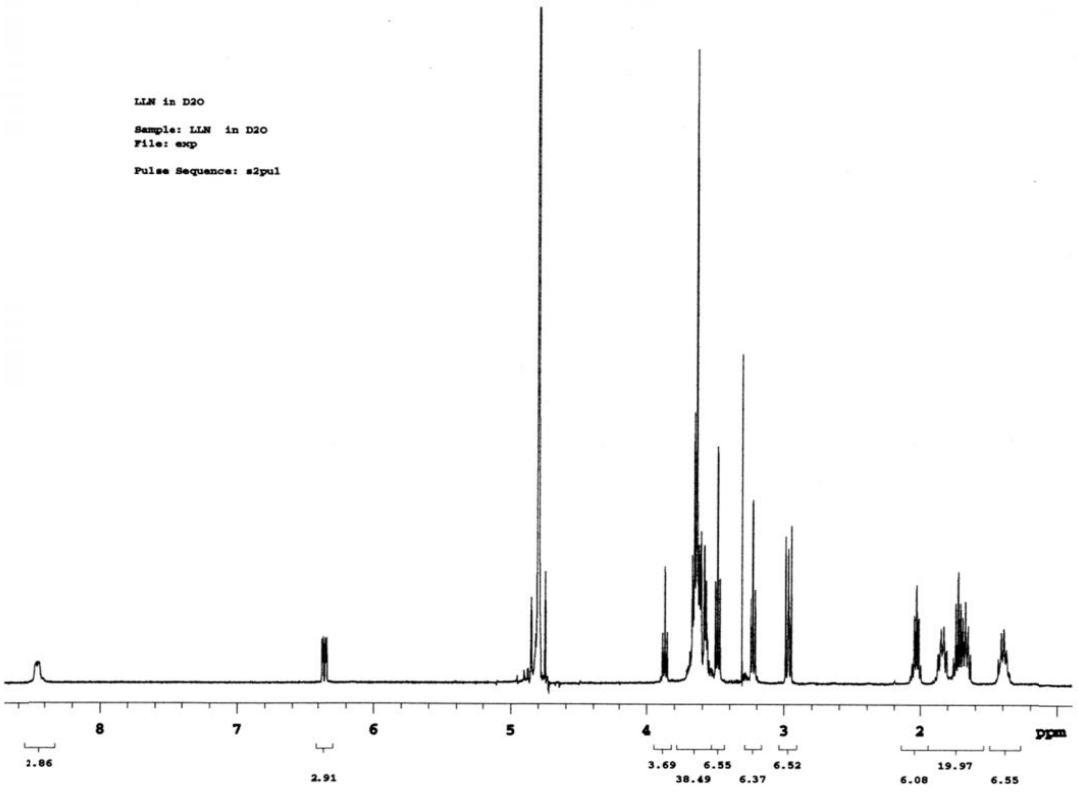


Figure 2.14 ^{1}H NMR of **5b**

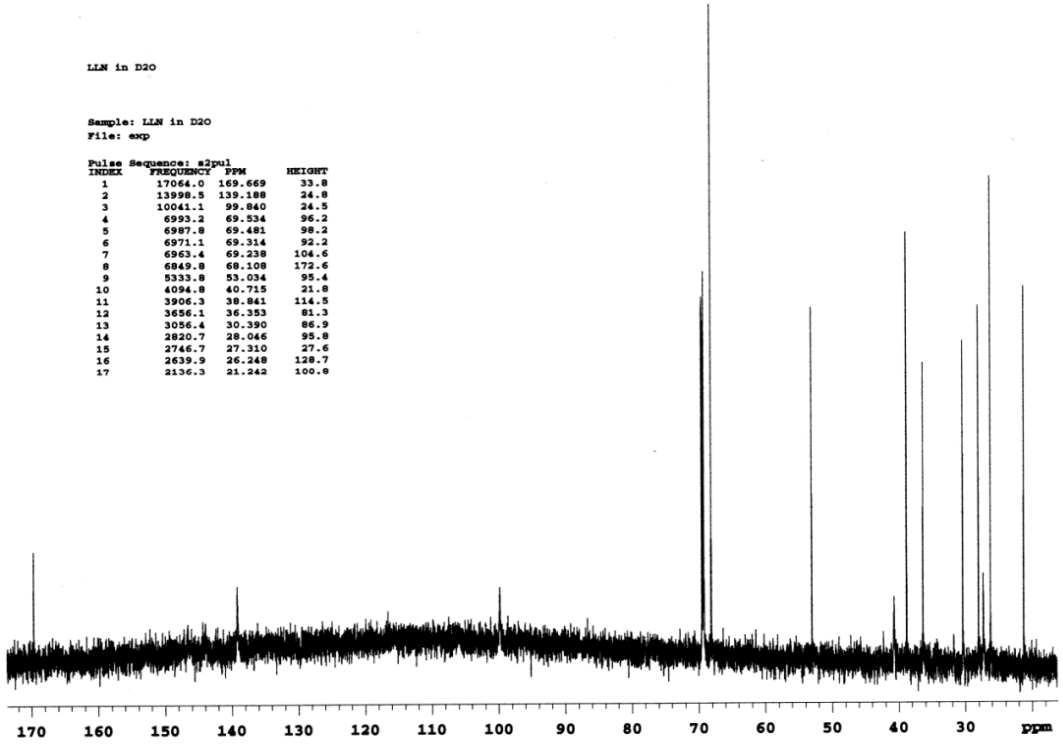


Figure 2.15 ^{13}C NMR of **5b**

2.4.6 Fluorescence measurements

The *lysC* riboswitch RNA was folded in the binding buffer (50 mM HEPES pH 7.5, 20 mM $MgCl_2$, and 100 mM NaCl and 100 mM KCl) by incubating at 90 °C for 1 min, 4 °C for 5 min and 37 °C for 30 min. The molar extinction coefficients (ε) of **5b** and **5c** were determined by measuring absorbance at different known concentrations of the dye solutions. They were determined to be 200 $M^{-1}cm^{-1}$ at 328 nm and 2000 $M^{-1}cm^{-1}$ at 478 nm, respectively. For titration, a stock solution of the folded RNA containing 200 nM dye in the binding buffer were added to a 200 nM dye solution in the binding buffer. Fluorescent intensity measurements were performed on a Shimadzu RF-5301PC fluorescence spectrometer using quartz cuvettes. The emission intensities were recorded in 5 min after each addition. The **5b** and **5c** samples were excited at 328 nm and 480 nm, respectively, and the emission was monitored at 550 nm for both samples.

Determination of dissociation constants

The binding constant between RNA (R) and Dye (D) is defined as

$$R + D \leftrightharpoons R{\cdot}D$$

$$K = \frac{[R{\cdot}D]}{[R][D]} \quad (1)$$

$$[R{\cdot}D] = K[R][D] \quad (2)$$

The fluorescence emission intensity of the sample solution is defined by

$$i = \alpha_U i_o + \alpha_B i_\infty = \alpha_U i_o + (1-\alpha_U) i_\infty \quad (3)$$

where, α_U and α_B are the fraction of the unbound and bound dye, respectively, and i_o and i_∞ are the fluorescence emission intensity of the unbound and bound dye, respectively.

The fraction of the unbound dye is defined by

$$\alpha_U = \frac{[D]}{[R \cdot D]+[D]} \quad (4)$$

Equations (2) and (4) give

$$\alpha_U = \frac{1}{1+K[R]} \quad (5)$$

Equations (3) and (5) give

$$\frac{i}{i_o} = \frac{1+\theta K[R]}{1+K[R]} \quad (6)$$

Where, $\theta = \frac{i_\infty}{i_o}$, the ratio of the fluorescence intensity in the presence of an infinite concentration of RNA and in the absence of RNA.

The initial concentrations $[R]_0$ and $[D]_0$ are defined by

$$[R]_0 = [R]+[R \cdot D] \quad (7)$$

$$[L]_0 = [L]+[R \cdot D] \quad (8)$$

Equations (2), (7), and (8) give the free RNA concentrations as

$$[R] = \frac{-(K[D]_0 - K[R]_0 + 1) + \sqrt{(K[D]_0 - K[R]_0 + 1)^2 + 4K[R]_0}}{2K} \quad (9)$$

Then the following equation is obtained from Equations (6) and (9):

$$\frac{i}{i_o}=\frac{1+0.5\theta\left\{-(K[D]_0-K[R]_0+1)+\sqrt{(K[D]_0-K[R]_0+1)^2+4K[R]_0}\right\}}{1+0.5\left\{-(K[D]_0-K[R]_0+1)+\sqrt{(K[D]_0-K[R]_0+1)^2+4K[R]_0}\right\}} \quad (10)$$

The nonlinear curve fitting by Equation (10) was performed using Origin Pro 8.0 and provided the binding constant K. The dissociation constant K_D is the reciprocal of K.

Chapter 3. Selective modification of cysteine in proteins

3.1 Protein modification

Proteins are primarily responsible for diverse cellular functions. Proteins play roles as structural components, transporters, signal transducers, defense mechanisms, and biological catalysts. The ability of proteins to perform a wide variety of functions is due to their molecular diversity. In fact, proteins are the most diverse of all the macromolecules: each cell accommodates thousands of different proteins. Despite being the product of the genome, the proteome – a cell's repertoire of proteins, is larger than the genome itself, especially in eukaryotes. The human genome project found only about 21,000 genes that account for approximately two to three order of magnitude more proteins in humans.[70]

The proteome diversification is achieved mainly by two distinct methods. The first method is mRNA splicing, including alternative splicing that occurs in a tissue specific manner and/or under specific cellular conditions.[71] For example, two completely different proteins, calcitonin, a thyroid hormone, and calcitonin gene related peptide (CGRP), a neurotransmitter produced by the nervous system, are products of the same gene generated by alternative splicing.[72-73] The second method for expanding the proteome diversity is the post-translational modification (PTM) of proteins at one or more sites.[74] For these posttranslational modifications, specific enzymes add some chemical moiety to proteins or cleave the peptide backbone. Found in all three domains

of life, the PTM includes SAM-dependent methylation, ATP-phosphorylation, acetyl-CoA-dependent acetylation, glycosylation, biotinylation, proteolysis, and lipidation.

To emulate the way nature diversifies the proteome to perform a myriad of cellular functions, researchers have developed strategies to introduce unnatural modification to a protein of interest for desired applications. The unnatural protein modification has provided valuable tool to study and characterize proteins. Such modifications are significant not only for fundamental research but also for biotechnology and biomedical applications. Proteins have been modified by genetic or chemical tools.

3.1.1 Protein modification by genetic methods

Genetic methods of protein modification use recombinant DNA technology. The gene of interest is cloned in-frame (maintaining the open reading frame) with the sequence that encodes a molecular tag which enables rapid and effective characterization, purification, and *in vivo* localization of the gene product. The green fluorescent protein (GFP) and its variants have, for example, been of one of the most important tool due to their stability and sensitivity[75] for monitoring gene expression levels in living cells or tissue,[75] cellular localization of the fusion proteins and the cellular dynamics. Typically added to the N- or C-terminus, the polyhistidine tag (His-tag) containing at least six histidine residues, can bind to transition metal ions such as Ni^{2+}, Co^{2+} or Zn^{2+}.[76] The His-tag has been widely used in protein purification using immobilized affinity chromatography (IMAC). It has

also been used for immobilization of a protein on a surface.[77] In addition, the genetic code has been expanded for the incorporation of unnatural amino acids (other than the standard 20 proteinogenic amino acids) to any specific site in the target protein. So far, about 71 different unnatural amino acids with a wide range of reactivity and functionality have been successfully added to proteins of *E. coli*, yeast, and mammalian cells.[78]

Despite its wide range of application, genetic tools for protein modification have several limitations. The diversity of genetic modification is normally limited to the genetic code. Fused peptides could potentially perturb the function, localization, and interaction of the target protein. Compared to organic fluorescent dyes, fluorescent proteins are relatively less sensitive to the environment and display lower photostability and brightness. Most fusion proteins or labels are primarily added to the N- or C-terminus, which could be involved in protein interaction or could be subject to terminal proteolysis.

3.1.2 Chemical modification of proteins

Proteins can be chemically modified with small molecules with intrinsic biophysical properties such as fluorescence, affinity, isotopic activity and unique reactivity. In this method, the fully synthesized proteins are conjugated to chemical labels *in vivo* or *in vitro* by chemo-selective reactions of specific amino acid residues. Chemical modification is widely useful in the identification, detection, functional characterization, purification and localization of proteins depending on the properties of molecular tags.

Proteins chemical modification has several advantages over the genetic modification. Relatively smaller molecular tags are less likely to interfere with inherent structure and function of the proteins. Moreover, chemical reagents can flexibly modify either terminal or internal amino acids. Chemical labeling of proteins depends on the intrinsic reactivity of the amino acid residues. Many chemoselective reagents haven been developed for labeling of proteins both *in vivo* and *in vitro*.[79]

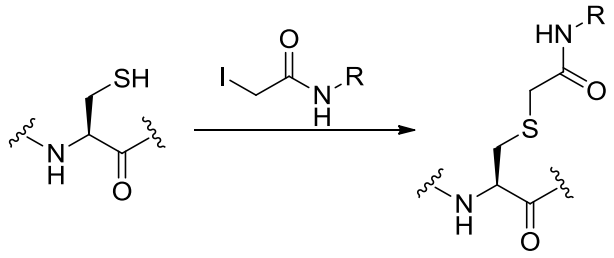
3.3 Methods for chemical protein labeling

Chemical protein labeling has been achieved by the careful selection of reagents with different reactivity and selectivity towards the nucleophilic amino acid residues such as cysteine and lysine. Due to its lower pKa to that of ε-amine of lysine, the protein N-terminal amino group can also be selectively modified.[80] Traditionally, the remaining 18 amino acids were rarely targeted for the chemical modification. However, some of these amino acids have recently been modified with new selective reagents. In addition to the side chains, the N-terminal amino group and the C-terminal carboxylate are important targets for the selective protein modification. In this dissertation, the methods used to modify the Cys residue in protein will be discussed in more detail.

3.4 Specific labeling of proteins via cysteine

Among the proteogenic amino acids, cysteine is an excellent target for the selective bioconjugation due to its strong nucleophilicity. The cysteine residues are naturally less abundant[81] and can be easily introduced to desired location of a protein by site-directed mutagenesis. With a pKa value of about 8.5, the thiol group (SH) of cysteine easily undergoes deprotonation to form the strongly nucleophilic thiolate anion, which is susceptible to disulfide formation, nucleophilic displacement or conjugate addition. Some conventional and recent methods for the chemical modification of the cysteine residue are discussed in the following paragraphs.

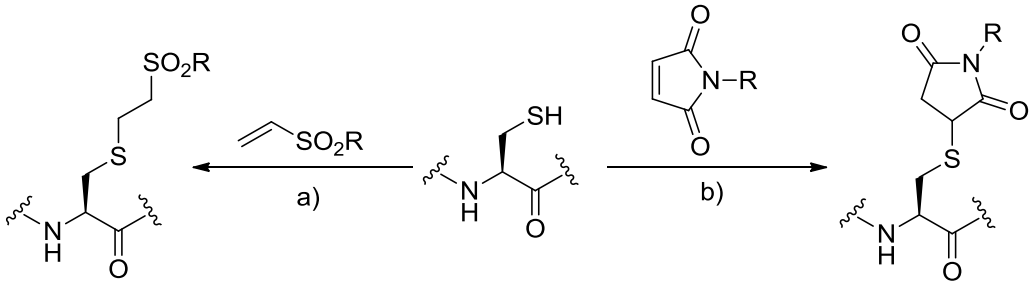
3.4.1 Alkylation of cysteine



Scheme 3.1 Modification of cysteine using iodoacetamides

The most common way to modify cysteine is the direct alkylation of the thiolate ion. Iodoacetamides, α- halocarbonyl electrophiles were among the first reagents for the direct alkylation of cysteine to form a stable thioether linkage (Scheme 3.1). Iodoacetamides were used for modifying surface cysteine residue of keratin for its characterization.[82] Similarly, the cysteine residues in the enzyme active site have been modified with

iodoacetamides for its functional studies.[83] In some specific cases, iodoacetamides can also react with methionine[84] and lysine residues, compromising the cysteine-selective reactivity but this issue can be solved by the use of less electrophilic chloroacetamide.[85]

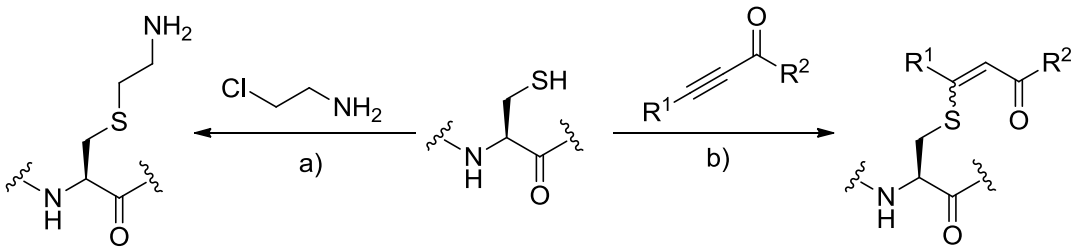


Scheme 3.2 Modification of cysteine with a) vinyl sulfones and b) maleimides

Cysteine can be alkylated by conjugate addition to Michael acceptors, due to the adequate nucleophilicity of the thiolate ion as Michael donor. The reaction rate of the conjugate addition increases with pH due to the increased concentration of thiolate anion.[86] Maleimides, vinyl sulfones and α,β-unsaturated ketones are widely used for the cysteine modification. The reaction of a maleimide reagent with cysteine is a very selective process.[87] Although most maleimide reactions are irreversible, bromomaleimides provides bioconjugation with additional site for further bioconjugation.[88] Vinyl sulfone and its derivatives have been used for cysteine conjugation because of good water solubility, mild reaction condition, and excellent yields.[89-90]

Although frequently observed in organic synthesis, conjugate addition of thiols to alkynes was rarely used for protein modification. Only recently, the cysteine thiols have been added to electron deficient alkynes such as alkynoic amides for peptide modification

(Scheme 3.3b).[91] The resulting thio-olefin can easily undergo a second addition for further modification.



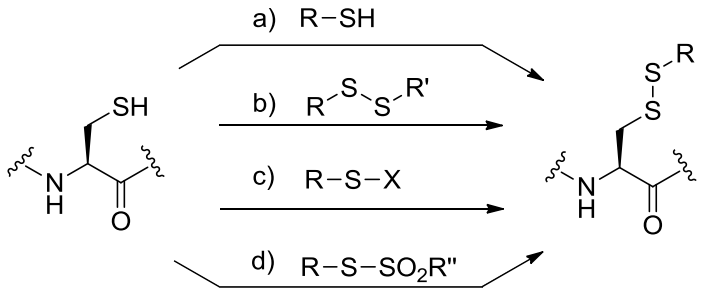
Scheme 3.3 Modification of cysteine: a) aminoethylation to generate lysine mimic b) Use of electron deficient alkynes where R^1= H, alkyl and R^2=alkyl, aryl, NHR or OR.

Aminoethylation of cysteine using chloroethylamine and its derivatives have been used to generate lysine analogs (Scheme 3.3a). This method has been especially used to determine the role of lysine in the active site or the functional domain by converting the inactive cysteine residue to lysine mimics.[92]

3.4.2 Oxidation of cysteine

The sulfur atom of cysteine can be converted into a wide range of oxidized forms that exhibit distinct reactivity. The thiolate ion is easily oxidized to form a disulfide bond with another thiol. This S-S bond formation is important for the stability and folding of many proteins. It has been used for the chemical modification of proteins. This reaction is simply performed by mixing a protein with thiol containing reagents. However, these reactions form mixed products and require long incubation periods. These problems have been addressed by developing activated reagents like sulfenyl halides,[93] alkylthiosulfonate[94] and disulfide exchangers.[95] The reaction of proteins with of 4-

nitrophenylsulfenyl chloride in aqueous acetic acid covalently modifies the cysteine residue to yield mixed disulfide.[96] Sulfenyl halides are mostly useful for analytical purpose, whereas alkylthiosulfonate and its derivatives are used for functional studies. Glycomethanethiosulfonates have been used for the synthesis of glycolyated proteins in high yield.[94]



Scheme 3.4 Methods of disulfide formation of cysteine: a) air oxidation, b) disulfide exchange, c) sulfenyl halides, and d) alkylthiosulfonate

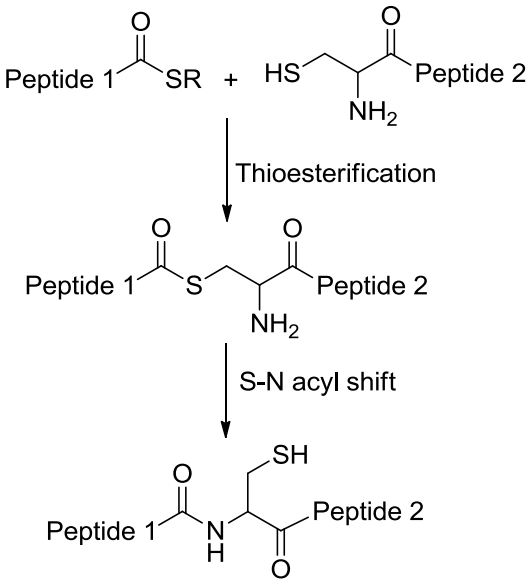
Although efficient to specifically modify biomolecules, the oxidation of cysteine to form disulfide is limited by the inherent instability of the disulfide bond especially in a reducing environment. However, this instability has been used as an advantage in drug delivery where the tethered modification is released once the system reaches the destination with a reducing environment.[97]

3.5 Ligation strategies at N-terminal cysteine

The reactivity of the N-terminal cysteine is comparable to internal cysteine residue for most of the aforementioned reactions. However, the aminothiol moiety of the N-terminal cysteine provides unique reactivity and has been used for selectively modifying protein *in*

vivo and *in vitro* by methods such as Native Chemical Ligation (NCL) and thiazolidine ligation.

3.5.1 Native chemical ligation at cysteine (NCL)

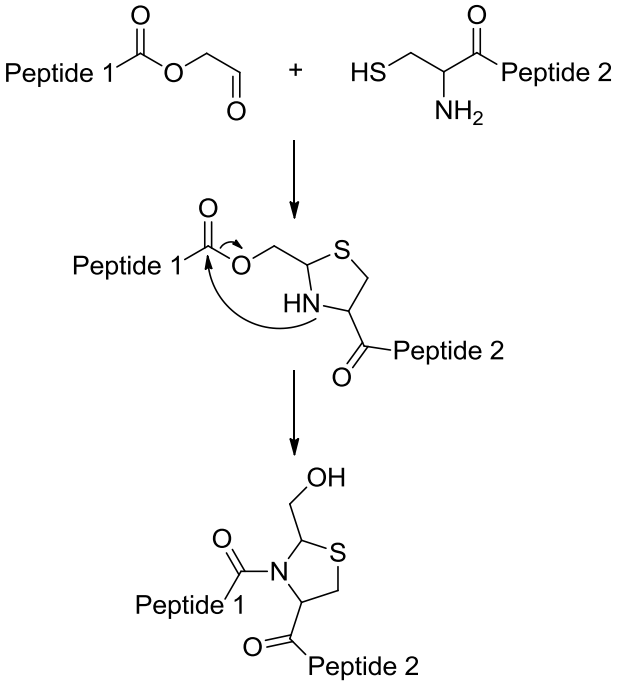


Scheme 3.5 Native chemical ligation of two peptides

Kent and co-workers introduced a thioether ligation technique termed as native chemical ligation.[98] This transformation involves the coupling of one peptide containing N-terminal cysteine residue and another peptide containing C-terminal thioester to form a single polypeptide with a "native" amide bond at the ligation site. Mechanistically, NCL involves two steps (Scheme 3.5): 1) the reversible trans-thioesterification by the nucleophilic attack of thiolate group of the N-terminal cysteine on the C-terminal thioester to form thioester intermidate, and 2) the rearrangement of the intermediate by an intramolecular S-N acyl transfer to form a native peptide bond.

The success of NCL can be attributed to its regioselective and chemoselective nature. The ligation occurs without reaction with internal cysteine residues and racemization at the ligation site.[99] Moreover, protecting groups are not required and the reaction can be performed in water at neutral pH even in the presence of denaturants. NCL is a powerful tool for the modification, synthesis and semi-synthesis of proteins. NCL has proved possible to make every native protein that has been attempted by the chemical ligation with size beyond the traditional limits of solid-phase synthesis.[100] In addition, the approach has been used for efficient addition of isotopic markers,[101] chemical probes and posttranslational modification to proteins and peptides of interest.[102]

3.5.2 Thiazolidine ligation



Scheme 3.6 Thiazolidine ligation of two peptides as described by Liu et.al

The thiazolidine ligation is also used to conjugate two large unprotected peptides. First

reported in 1994 by Liu *et al*., this method employs the reaction of alkyl aldehyde or glycolaldehyde with the 1-2, aminothiol component of N-terminal cysteine residue at acidic pH to form a thiazolidine ring.[103] This pseudoproline ligation is initiated by an imine formation between the C-terminal aldehyde with the N-terminal amino nucleophile which then imine then undergoes a rapid rearrangement. In general, the thiazolidine ligation process consists of three steps:

1. Introduction of glycolaldehyde ester into the carboxyl terminus of the first unprotected peptide,
2. Thiazolidine ring formation by the dehydration reaction between the aldehyde and the N-terminal mercaptamine group of the second unprotected peptide, and,
3. The *O,N*-acyl rearrangement in which O-acyl ester linkage is transferred to N-acyl amide linkage to form a peptide bond with a pseudoproline (2-hyroxylmethyl thiaproline) structure.

The final step, in which the secondary amine of the thiazolidine ring conducts a nucleophilic attack, is the rate determining step.[103] While the ring formation step is very fast, the rearrangement step is generally slow and depends on pH and on the steric and electronic environment.[104] The ligation is usually performed under acidic conditions at room temperature to avoid side reactions of aldehyde with other nucleophiles within the peptides.

Like NCL, this method allows the synthesis of the long peptides and proteins that cannot be normally synthesized by solid phase peptide synthesis. For instance, this

method was successfully used to synthesize a 50 residue-long analog of transforming growth factor-α[104] and a 99 residue HIV-1 protease analog with biological activities comparable to native protein.[105] The unnatural pseudoproline structure also did not affect the cellular uptake or function of a peptide fragment ligated to a membrane permeable sequence.[106]

3.6 Specific objective

The objective of the second part of this dissertation is to introduce a new and efficient method for the reversible chemical modification of protein N-terminal cysteine. This strategy involves chemoselective reaction between the N-terminal cysteine and rationally designed and synthesized pyruvate analogs to form a stable five-membered thiazolidine ring at the ligation site. In contrast to the method developed by Liu *et al*., which was exclusively used for the peptide ligation, the new pyruvate method can be generally used to label N-terminal cysteine with varieties of functional handles both *in vitro* and *in vivo*.

Although valuable for protein chemistry, Lui's thiazolidine ligation method is less desirable in certain circumstances. For example, the C-terminal aldehyde is very reactive, therefore its preparation and handling require extra precaution. For the same reason, pseudoproline ligation reactions are relatively not very chemo-specific. The imine capture can be achieved by the N-terminal amino group of non-cysteine residues such as Ser, Thr, Trp, His, and Asn that form a variety of heterocyclic rings at the ligation site

before the final O,N-acyl rearrangement step.[107] This could be especially problematic for the ligation of a mixture of proteins.

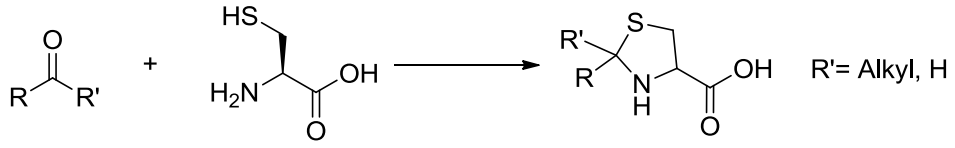
The thiazolidine ligation reaction sometimes requires very strong acid, such as trifluoroacetic acid, which is not very suitable for some proteins. Such acidic condition is not feasible for the protein modification *in vivo*. Furthermore, introducing the aldehyde moiety at the C-terminus requires a heavy metal reagent.[105]

Much of these issues will be addressed in the mild and efficient methodology presented henceforth.

Chapter 4. Reversible modification of the protein N-terminal cysteine residue *in vitro*

4.1 Introduction to the reaction of cysteine with pyruvic acid

In the 1930's literature, it has been reported that cysteine reacts with carbonyl compounds to form cyclic compounds. For example cysteine reacts with various aldehydes and ketones to form thiazolidine compounds (Scheme 4.1).[108-109] Independently, Ratner *et al.* reported that cysteine reacts with formaldehyde to form thiazolidinecarboxylic acid.[110]

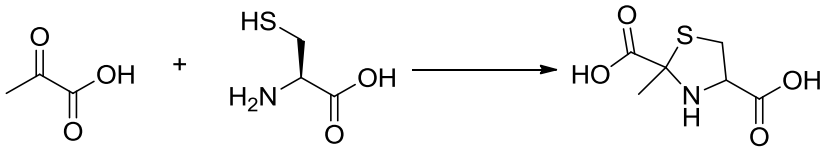


Scheme 4.1 The condensation reaction of cysteine with aldehyde and ketones to form proline like thiazolidines

This chemistry has been used in peptide chemistry and pharmacology. For example, 2,2-dimethylthiazolidine derived by the condensation of cysteine with acetone was used to protect cysteine from oxidation and β elimination for peptide synthesis.[111] The thiazolidine ligation strategy used for the synthesis of peptides and proteins discussed in chapter 3 was the reaction of N-terminal Cys with aldehyde compounds. Similarly, thiazolidines as pseudoproline derivatives improved the solvation and coupling kinetics for the peptides synthesis by disrupting intermolecular aggregation and providing access to 'inaccessible' peptides.[112-113] Various derivatives of thiazolidine carboxylic acids

synthesized from cysteine and aldehydes have also been used as cysteine prodrugs to avoid spontaneous oxidation and other reactions of cysteine.[114-115]

In his initial report, Schubert described that pyruvic acid reacted with cysteine to form a simple acyclic S-adduct. However, in his subsequent publication, he reported the complex formed between pyruvic acid and cysteine was 2-methyl-2,4-thiazolidinedicarboxylic acid (CP), as shown in scheme 4.2.[116] His conclusion was based on his unsuccessful attempts to substitute the amino and hydroxyl groups in his originally proposed acyclic structure.



Scheme 4.2 Reaction of cysteine and pyruvate to form 2-methyl-thiazolidine-2, 4-dicarboxylic acid (CP)

Schubert was able to synthesize CP under mild condition by simply mixing cysteine and pyruvic acid in the presence of pyridine in ethanol. This method was used repeatedly for the synthesis of CP by various groups.[117-118] CP can also be formed from cysteine and pyruvic acid at physiological pH.[119-120] Wlodek *et al.* identified the CP formation by electron impact mass spectroscopy and also reported the pH dependence of the CP synthesis.[121] The authors observed that the CP formation was optimal at pH 7.0-8.0.

The pursuit of developing the methodology of proteins modification using pyruvate analogs was motivated by the observation of CP formation *in vivo* as side

reaction. Rose and his group, for the first time, reported the pyruvate attachment to the N terminal cysteine of a recombinant DNA binding protein.[122] The recombinant protein expressed in *E. coli* was 70 mass units heavier than predicted for the determined sequence. The difference in mass was due to loss of water as pyruvic acid is 88 Da. The pyruvate adduct of the N-terminal cysteine was confirmed by NMR spectroscopy, which recorded no amide, ester, or thioester bond. The formation of CP is a spontaneous side reaction at the N-terminal cysteine with cellular pyruvate during the expression of recombinant proteins.[122-124] Therefore, it should be feasible to modify recombinant proteins using pyruvate analogs *in vivo*.

Unlike cysteine's reaction with aldehyde and ketones, its reactions with pyruvate have not been used for any biological application. Hence, this part of the dissertation aims to use various pyruvate analogs for the N-terminal specific labeling of proteins in vitro and in vivo.

4.1.1 Reversibility of thiazolidine formation

Ratner et al. reported that thiazolidinecarboxylic acids were hydrolyzed back to formaldehyde and cysteine by treating with boiling HCl, with ferric chloride at pH 10, or with sodium sulfite at pH 6.[110] King *et al.* reported that 2,2-dimethylthiazolidine (DMT) protecting group of Cys could be removed by using mercury derivatives, followed by the treatment of hydrogen sulfide to remove the mercury derivative.[125] As DMT gained popularity in peptide synthesis, mild methods were developed for thiazolidine hydrolysis.

The use of HCl in methanol was also de-protected DMT to generate free Cys.[111] A mixture of 1:1 water and alcohol was also successfully used for the hydrolysis of DMT attached to a Gly-Gly-Ala tripeptide.[126]

The thiazolidine ring can also be efficiently removed by the reaction with aminoxy compounds such as methoxyamine[127,128] Since methoxyamine-HCl is relatively inert towards the main protein composition, it was used for efficiently re-generate the intact original protein after the modification using pyruvate analogs.

4.1.2 Design of pyruvate analogs

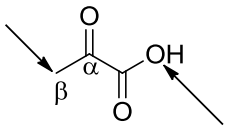


Figure 4.1 Structure of pyruvic acid with sites for modification (indicated by arrows)

Pyruvic acid is a simple alpha-keto acid with pKa 2.45 (Fig 4.1). Pyruvate is a major component in several cellular metabolic pathways. The formation of CP occurs at the α-carbonyl. Therefore, the methyl group and carboxylate of the pyruvate can be modified for the synthesis of analogs, which can react with the N-terminal cysteine of proteins. A variety of pyruvate analogs containing fluorescent dyes and biotin will be useful in studying proteins. Pyruvate can also be modified with bioorthogonal functional groups that do not interfere or interact with the biological system. These functional groups are useful for the reactions between the orthogonal pairs, such as aldehyde/ketone and

hydrazide/alkoxyamine,[129] azide and alkyne,[130] azide and phosphine,[131] tetrazole and alkene.[132]

4.1.3 Generation of recombinant proteins with N-terminal Cys residue

Recombinant proteins with N-terminal Cys can be obtained by several methods. One method involves the cleavage of a precursor fusion protein with a site-specific protease to expose an internal cysteine as the new N-terminal residue.[133] Site-directed mutagenesis can be used to introduce N-terminal Cys. As the protein synthesis initiates with N-formylmethionine in prokaryotes, it is imperative to introduce a Cys mutation at the second position of the recombinant protein. During the synthesis of this mutant protein in bacterial system, the N-formyl group is removed by peptide deformylase to leave the N-terminal methionine (Met_1).[134] Methionine aminopeptidase (MAP) subsequently removes the Met_1 residue.[135,136,137] Hence, cysteine will be introduced at the second position of the model protein and expressed in *E. coli* MAP to generate N-terminal cysteine residue.

4.2 Significance

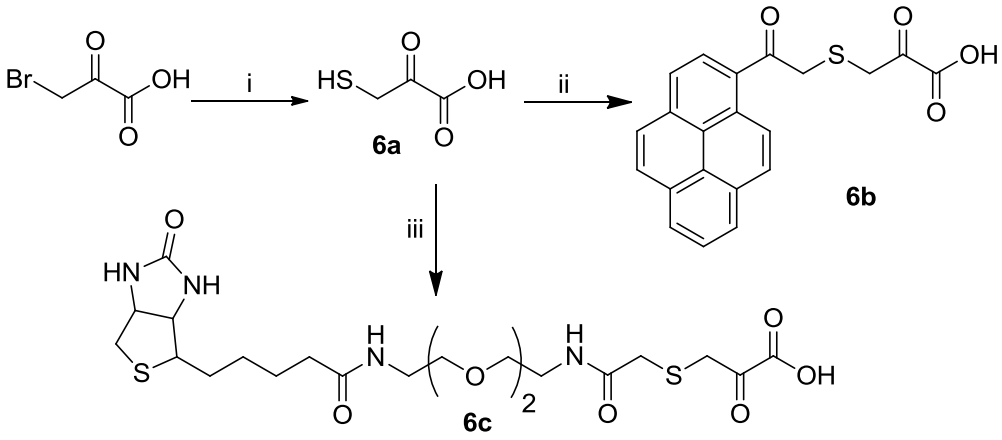
The use of pyruvate analogs to modify the N-terminal Cys residue of proteins will broaden the use of thiazolidine chemistry from peptide ligation to chemical tagging of proteins. The N-terminal specific incorporation of chemical reporters into proteins will

complement the modification methods described before. This method of labeling could be attractive due to following features:

1. *Reaction at physiological condition*-- which is crucial especially for the stability of biomolecules and reactions *in vivo*.
2. *Mild reactivity of pyruvate analogs*-- would ensure easy handling and preparations as well as little side reactions.
3. *Potential use in peptide conjugation*-- for the synthesis of peptides and proteins
4. *Reversibility of thiazolidine formation*—re-generation of the unmodified proteins under controlled condition after the intended application.

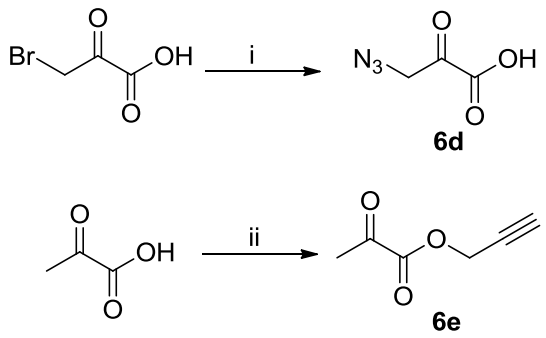
4.3 Results and discussion

4.3.1 Synthesis of pyruvate analogs



Scheme 4.3 Synthesis of pyruvate analogs: i) NaSH, CH_3OH, 48% ii) Bromoacetyl pyrene, K_2CO_3, DMF,57% crude iii) Biotin-PEG-iodoacetamide, K_2CO_3, DMF, quant.

We synthesized the pyruvate analogs with set of a commonly used tags with proteomical interest such as pyrene[138] (fluorescence) or biotin[139] (affinity). Commercially available 3-bromopyruvic acid was used to synthesize 3-mercaptopyurvate (**6a)** by reacting with sodium hydrosulfide as reported by Tanabe *et al.*[140] The S-alkylation of 3-mercaptopyruvate with 1-(bromoacetyl)pyrene and iodoacetyl-PEG_2-biotin produced the novel pyruvate analogs **6b** (57% crude) and **6c** (qant.)**,** respectively (Scheme 4.3).



Scheme 4.4 Synthesis of pyruvate analogs: i) NaN_3, DMSO, 76% ii) Propargyl alcohol, pTsOH, benzene, reflux, 15%

The azido and alkyne derivatives of pyruvate were also synthesized. 3-azido-2-oxo-propanioc acid (**6d)** was prepared by reacting 2-bromopyruvic acid with sodium azide as described in the literature (scheme 4.4).[141] All of these analogs were obtained by modifying C-3 of pyruvate. The carboxylate end of pyruvate was also derivatized. 2-Propynyl 2-oxopropanoate (**6e),** was prepared by refluxing pyruvic acid in presence of propargyl alcohol in the presence of a catalytic amount of p-toluenesulfonic acid, as reported in the literature (Scheme 4.4).[142]

4.3.2 Plasmid construction and protein expression

We used the Z domain as a model protein for the conjugation study. It is a small three helix-bundle immunoglobulin binding domain derived from *Staphylococcal* protein A. The Z domain was expressed with a C-terminal hexa-histidine tag from pETZ-21 plasmid. The amino acid sequence for the wild type Z domain expressed is as follow:

MTSVDNKINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAKKLN DAQAPKGSHHHHHH (theoretical mass of 7895.61 Da).

The codon for Thr-2 of the wild type was changed to Cys by site directed mutagenesis to produce pET21-Z (T2C) variant (Fig. 4.2). The introduction of cysteine at the second position ensures for the action of both endogenous defromylase and MAP to generate the Z domain variant with N-terminal cysteine when expressed in *E. coli*.

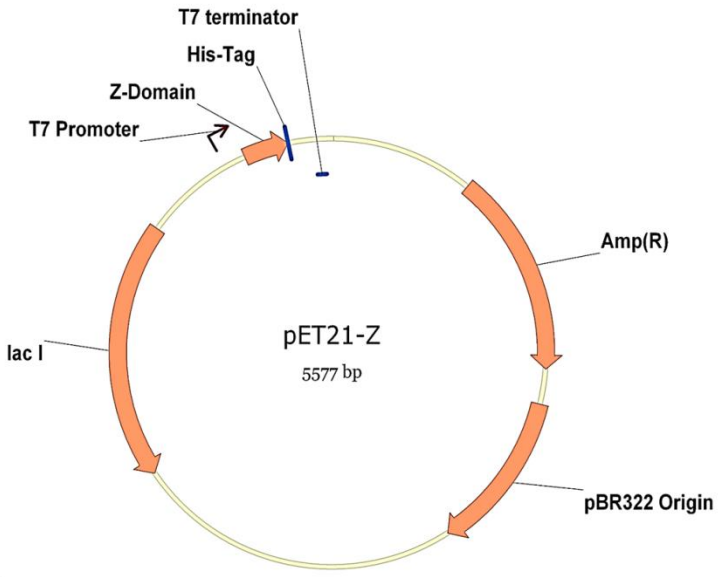


Figure 4.2 Diagram of constructed recombinant pET21-Z vector displaying the Z domain gene fused to his tag

The amino acid sequence for T2C Z domain variant following site-directed mutagenesis and the action of deformylase and MAP is as follow:

CSVDNKINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAKKLND AQAPKGSHHHHHH (theoretical mass is 7764.41 Da).

The Z domain (T2C) mutant was overexpressed in BL21(DE3) *E. coli* strain. The amino acid sequence of the C-terminal his-tag fusion was utilized to purify the expressed protein by IMAC and was visualized using SDS PAGE (Fig 4.3).

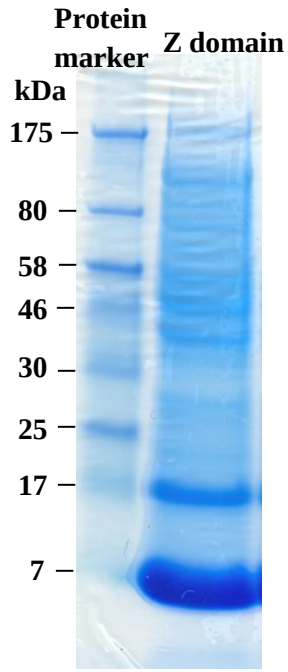


Figure 4.3 4-12 % SDS PAGE of purified T2C Z domain mutant. The first contains the Protein Marker, Broad Range (7-175 kDa) while the second land contains the Z domain protein near 7 kDa and its oxidized dimer at around 17 kDa

Mass spectrometry analysis of the purified protein confirmed the presence of N-terminal Cys (fig. 4.4) with no trace of 7895.61 Da (with Met_1) peak, which inherently confirms the activity of both deformylase and MAP. The observed mass of Z domain

mutant obtained after purification was 7835.66 Da, which corresponds to the pyruvate adduct that was removed with methylaminoxy treatment. The observed mass of Z domain variant with N-terminal cysteine after the removal of the pyruvate adduct provide was 7765.35 Da (Fig. 4.4) while the average calculated theoretical mass is 7764.41 Da. The mass spectrum also shows the presence of a minor peak 7863.60 Da (+ 98 Da) which is believed to be non-covalently attached adducts of either phosphoric or sulfuric acid from the mass spectrometer. Such adducts are observed during mass spectrometry analysis of proteins due to the presence of impurities and also can be easily removed.[143]

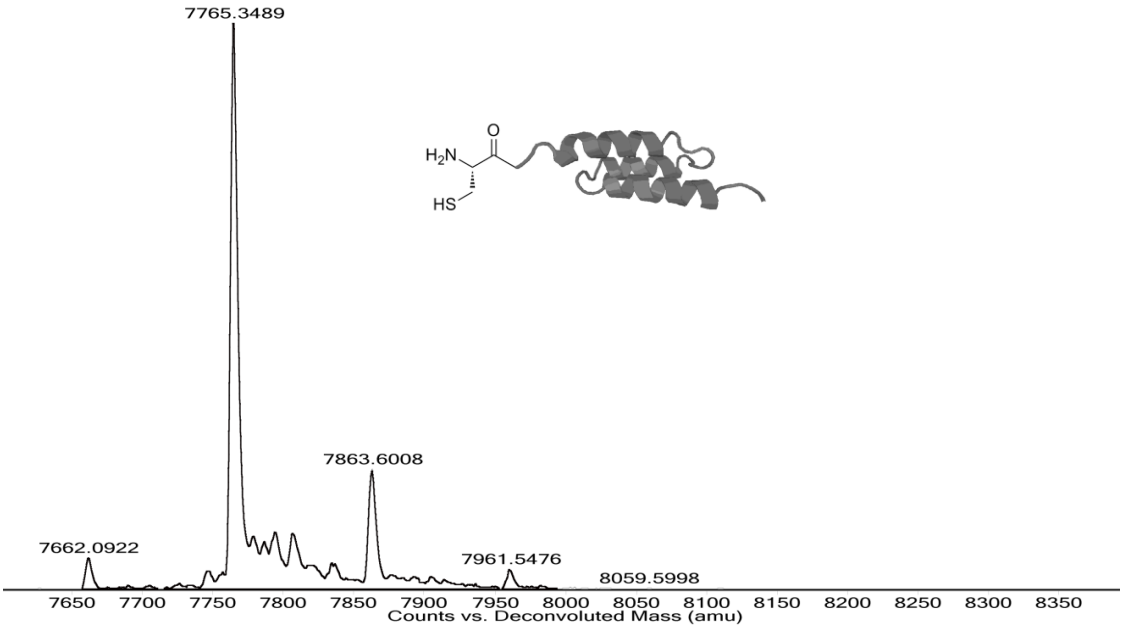
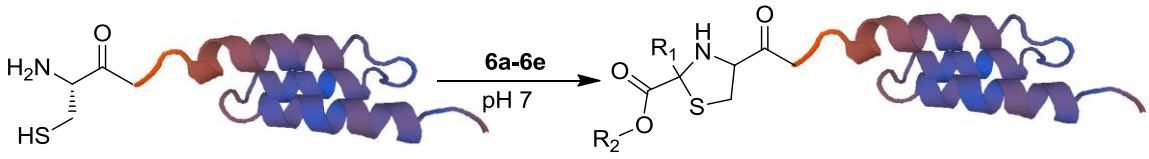


Figure 4.4 Deconvoluted ESI mass spectrum of T2C Z domain mutant and its cartoon representation.

4.2.3 Reactions of N-terminal cysteine of T2C Z domain mutant with pyruvate analogs



Scheme 4.5 The reaction between N-terminal Cys of T2C Z domain mutant with pyruvate analogs **6a-6e** performed in presence PBS buffer at pH 7 to produces thiazolidine moiety at the ligation site.

The overnight reaction of the T2C Z domain mutant with the synthesized pyruvate analogs in the presence of phosphate buffered saline (PBS), pH 7 at room temperature successfully generated the modified proteins. Prior to the addition of the pyruvate analogs into the reaction mixture, the T2C Z domain mutant was treated with *tris*(2-carboxyethyl)phosphine (TCEP) to ensure the absence of the oxidized disulfide dimer. The modified proteins were identified by electrospray ionization mass spectrometry (ESI-MS). Almost 100 % conversion was confirmed by the absence of T2C Z domain mutant (7765.35 Da) and the presence of single peak for the corresponding thiazolidine adduct in the mass spectrometry analysis.

For a test reaction, the T2C Z domain mutant was reacted with pyruvic acid. The mass spectrometry result displayed a formation of CP derivative form of T2C Z domain mutant with molecular mass of 7835.66 Da (theoretical mass 7834.46 Da), as shown in Figure 4.5. Similarly, the reaction with analogs **6a-6e** successfully provided the

corresponding modification of T2C Z domain mutant (Fig. 4.5, 4.6, 4.8, 4.10, 4.11, and 4.12). The theoretical and observed mass of the modified proteins are listed in Table 4.1.

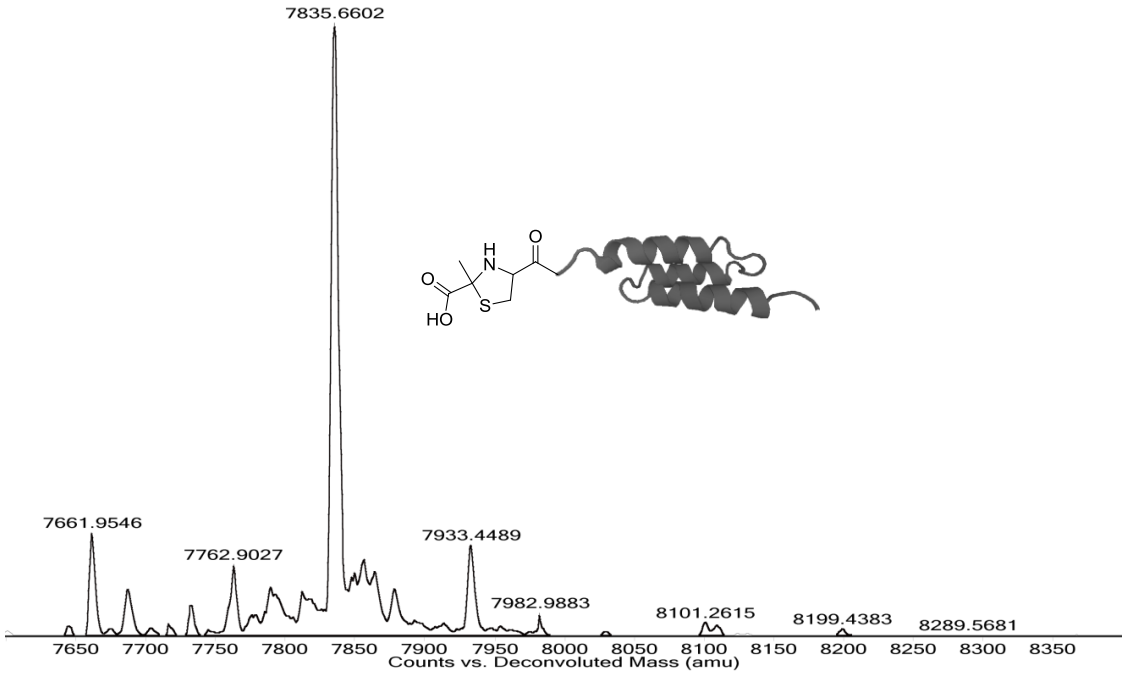


Figure 4.5 Deconvoluted ESI mass spectrum of CP derivatized T2C Z domain mutant obtained by reaction with pyruvate and its structure

	Theoretical mass (Da)	Observed mass (Da)
T2C Z domain mutant	7764.41	7765.34
T2C Z domain mutant + pyruvate	7834.45	7835.66
T2C Z domain mutant + **2a**	7866.52	7867.28
T2C Z domain mutant + **2b**	8108.79	8110.66
T2C Z domain mutant **+ 2c**	8281.04	8281.66
T2C Z domain mutant **+ 2d**	7873.49	7875.55
T2C Z domain mutant **+ 2e**	7872.50	7873.57

Table 4.1 The theoretical and observed mass of T2C Z domain mutant and its modified derivatives

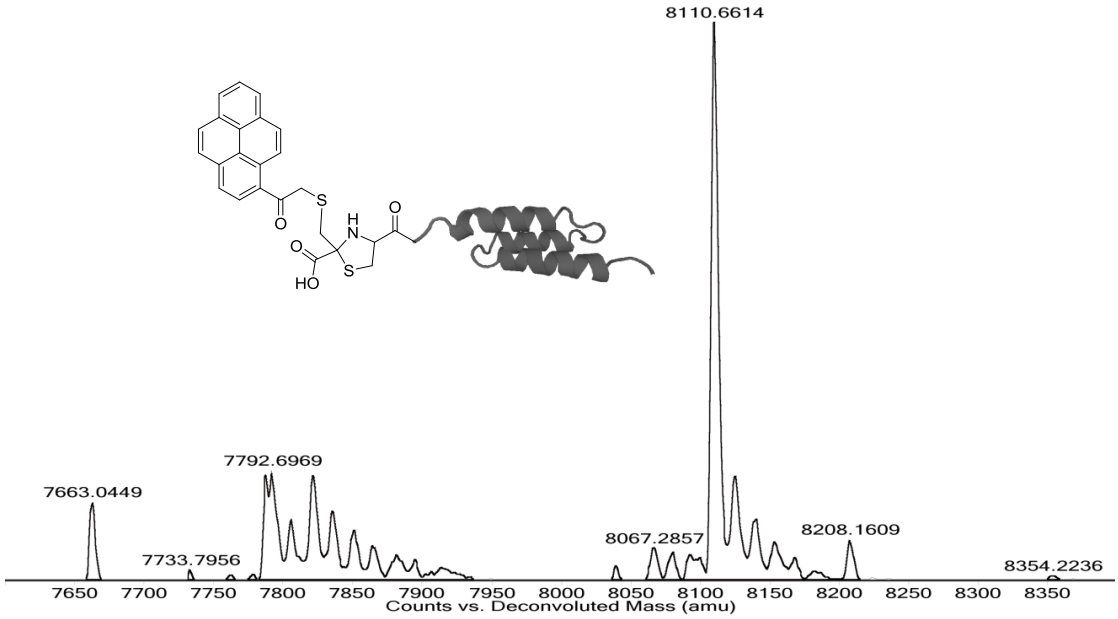


Figure 4.6 Deconvoluted ESI mass spectrum of pyrene labeled T2C Z domain mutant obtained by reaction with pyrene derivatized pyruvate analog **6b.**

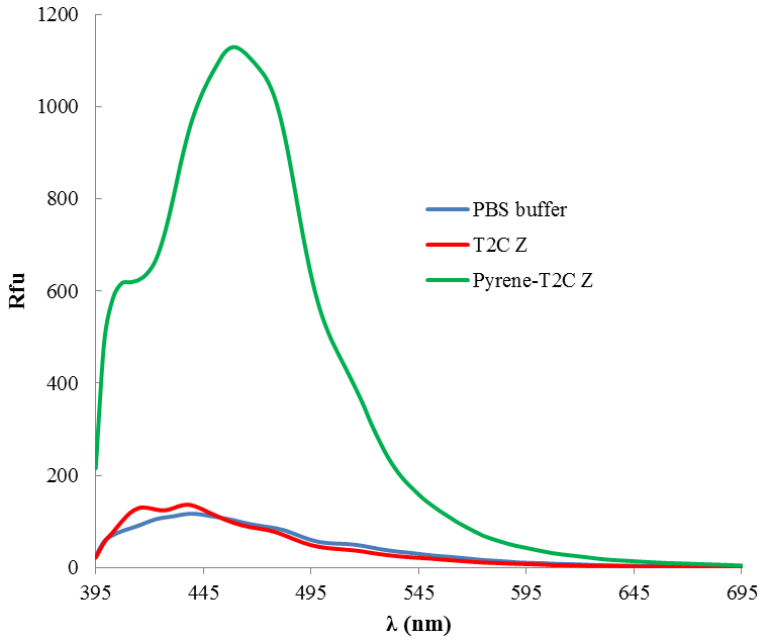


Figure 4.7 Fluorescence spectra for pyrene conjugated T2C Z domain mutant measured by NanoDrop 3000 fluorospectrometer (with wavelength range of 400-750 nm)

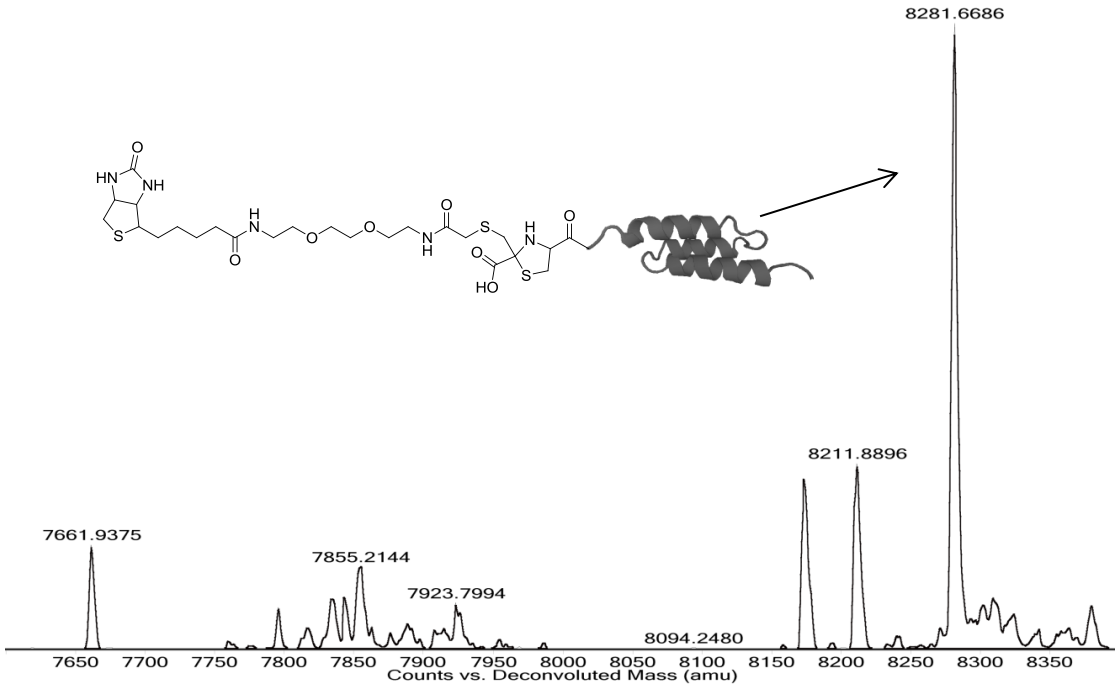


Figure 4.8 Deconvoluted ESI mass spectrum of biotinylated T2C Z domain mutant obtained by reaction with biotin derivatized pyruvate analog **6c**

In addition to the confirmation of biotinylation of T2C Z domain mutant by mass spectrometry (Fig 4.8), streptavidin gel-shift assay was also performed to inspect the proper functioning of the tagged biotin molecule. The binding of biotin to streptavidin is one of the strongest known non-covalent interactions with an unusually high affinity (dissociation constant $K_d = 10^{-15}$ M).[144] The absence of the biotinylated protein close to 7 Da marker in lane 3 confirms the binding of the biotinylated protein with streptavidin (Fig 4.9).

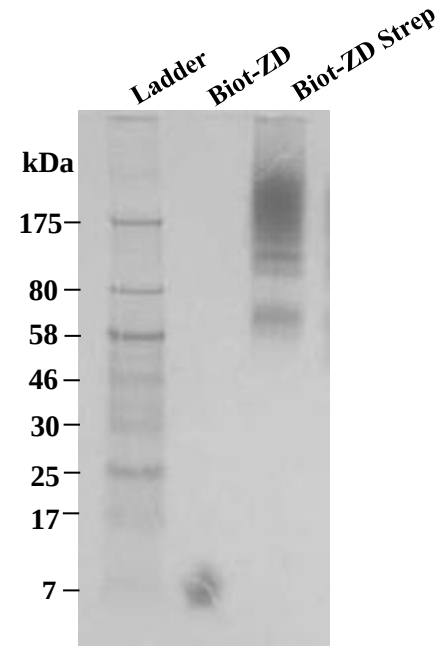


Figure 4.9 Streptavidin gel-shift assay: The first contains the Protein Marker, Broad Range (7-175 kDa), the second lane contains the biotinylated Z domain (Biot. ZD) and the third lane contains the biotinylated Z domain along with streptavidin.

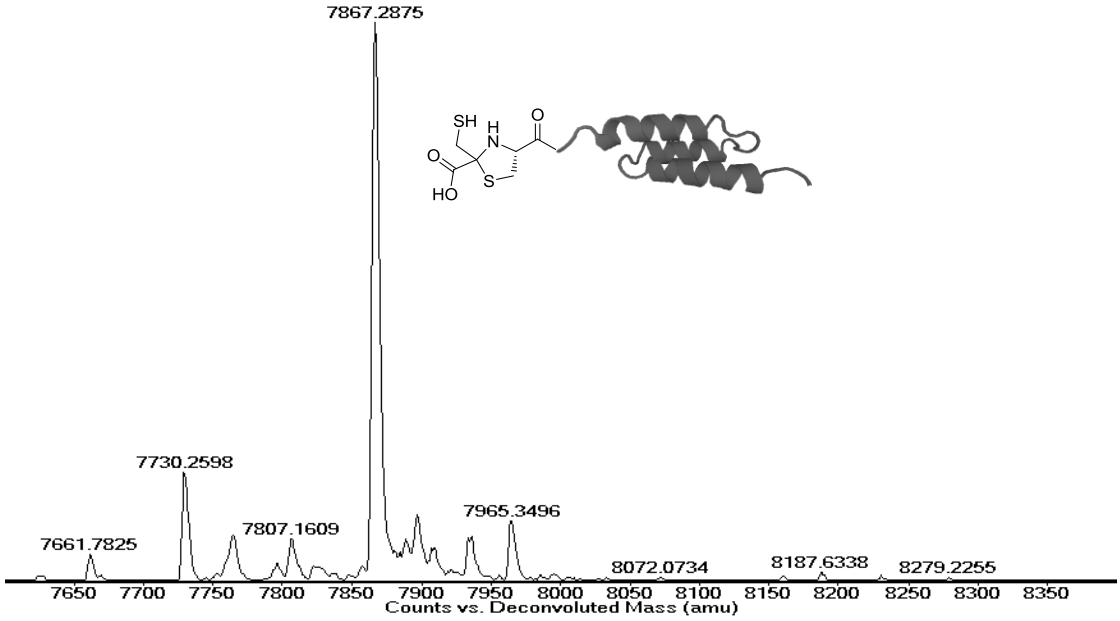


Figure 4.10 Deconvoluted ESI mass spectrum of thiolated T2C Z domain mutant obtained by reaction with thiol derivatized pyruvate analog **6a.**

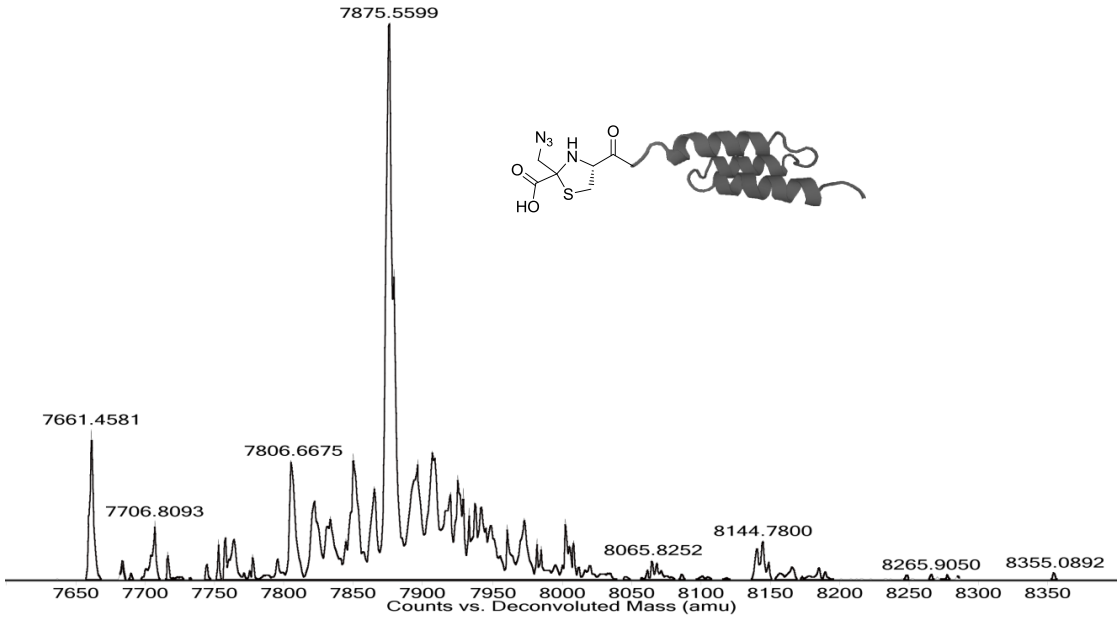


Figure 4.11 Deconvoluted ESI mass spectrum of azido-labeled T2C Z domain mutant obtained by reaction with pyruvate analog **6d**.

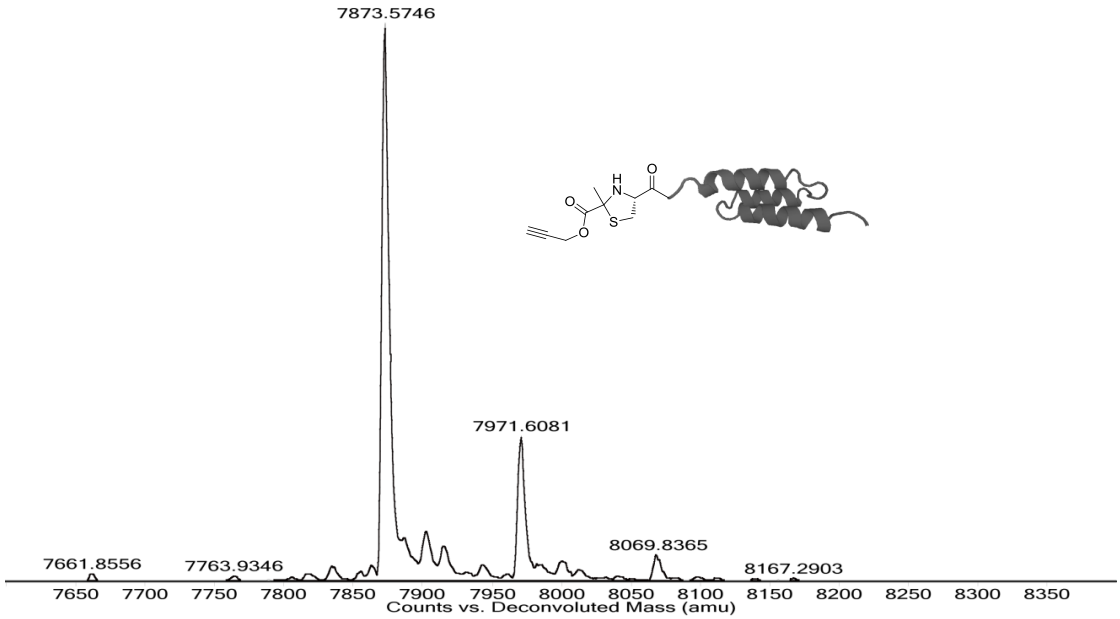


Figure 4.12 Deconvoluted ESI mass spectrum of T2C Z domain mutant functionalized with terminal alkyne obtained by reaction with pyruvate analog **6e**

4.3.4 Amount of analogs and reaction time required for complete modification

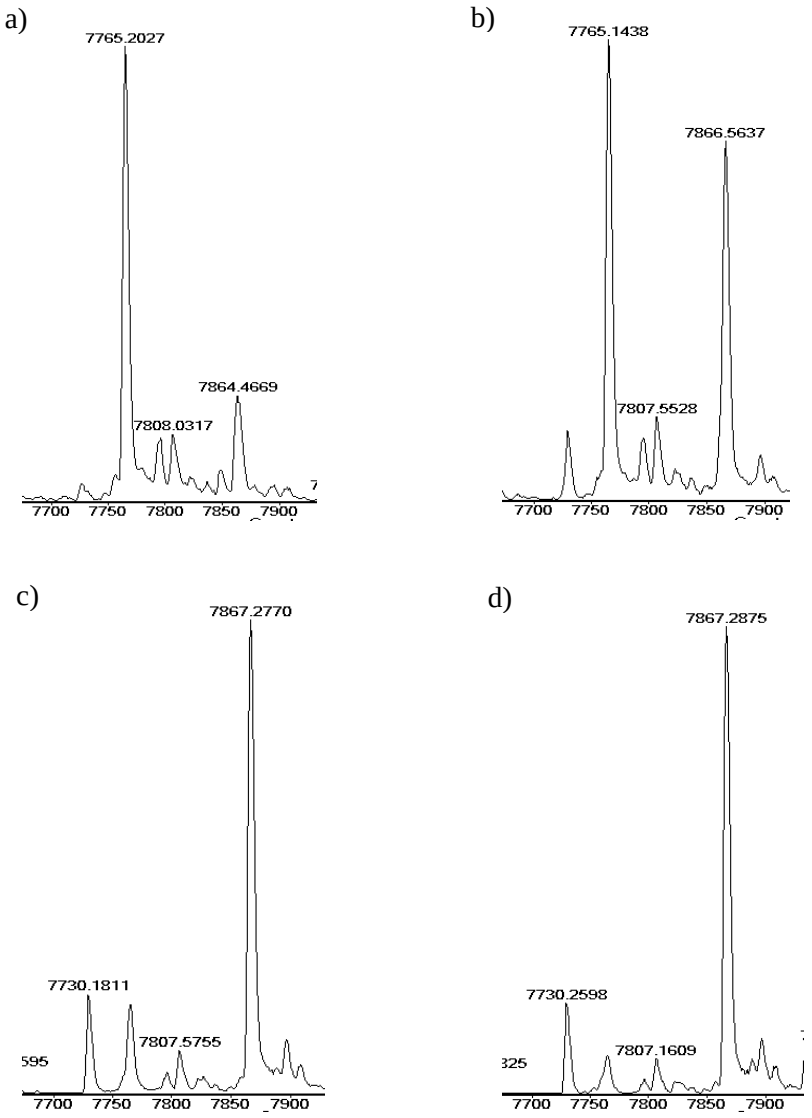


Figure 4.13 Reaction of 22.6 µg (2.9 nmoles) of T2C Z domain mutant using a) 0.1 µmole, b) 1 µmole, c) 5 µmole, and d) 10 µmole of 3-mercaptopyruvic acid.

To investigate the amount of analog required for the complete modification of T2C Z domain mutant using the method presented, we reacted various concentrations of mercaptopyruvate with 2.91 nmoles of the protein. On overnight incubation, reactions

with 0.1 µmol and 1 µmol of mercatopyruvate showed incomplete conversion while approximately 10 µmoles of mercaptopyruvate completely modified the given amount of protein (Figure 4.13 a-d).

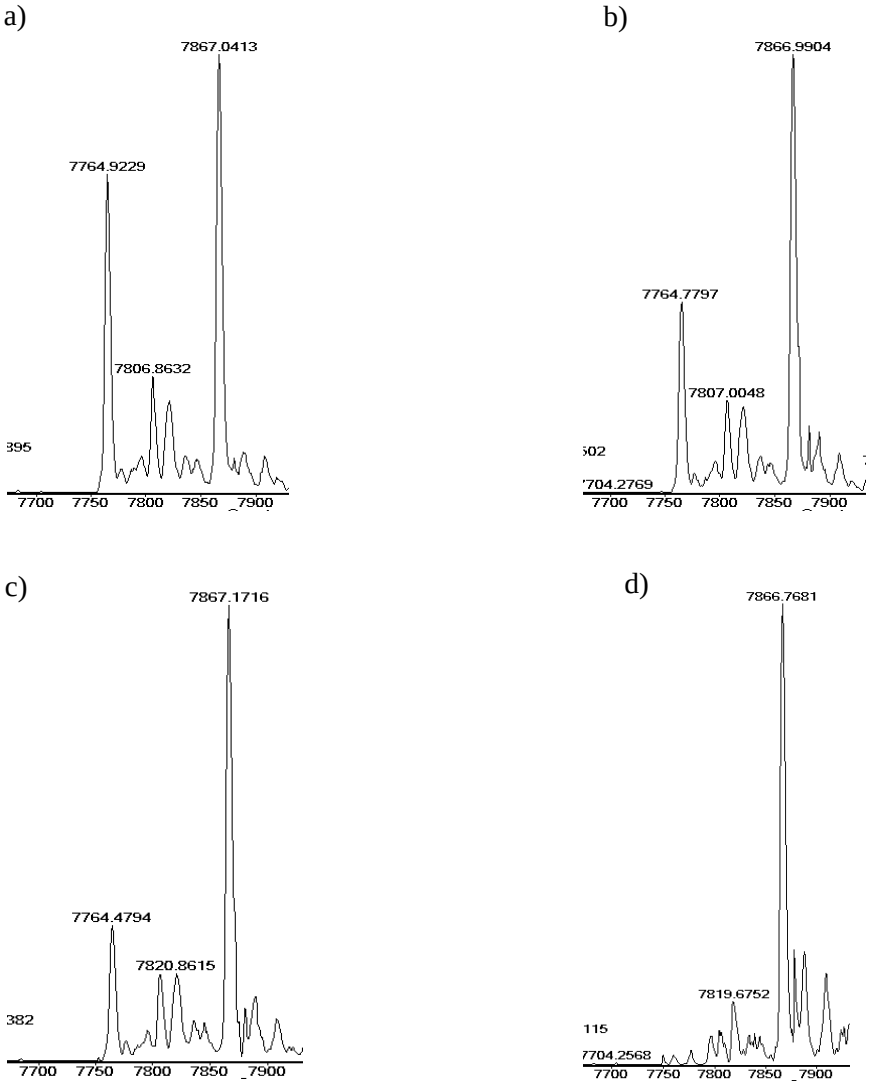


Figure 4.14 Reaction of 22.6 µg (2.91 nmoles) of T2C Z domain mutant using 10 µmole of 3-mercaptopyruvate incubated at room temperature for a) 1 hr., b) 2 hrs., c) 3 hrs., and d) 4 hrs.

For all the aforementioned reactions, an overnight incubation (usually 12-14 hours) was performed to ensure complete modification. However, the reaction between Cys and pyruvic acid is reported to be rapid, taking less than 2 hours in some instances.[121] Hence, to explore the reaction time for T2C Z domain mutant modification using the pyruvate analogs, 2.91 nmoles of the protein was reacted with 10 µmoles of 3-mercaptopyruvic acid for different time interval and analyzed using mass spectrometry. As evident from Figure 4.14 a-b, samples allowed to react for 1-2 hours contained a mixture of both the modified and unmodified protein suggesting incomplete reaction. However, the reaction sample incubated for 4 hours (Fig 4.14d) shows the complete absence of original protein suggesting complete conversion.

4.3.5 Regeneration of the intact unmodified T2C Z domain mutant

The covalent formation of thiazolidine between the N-terminal Cys and the pyruvate analogs was easily reversed to regenerate the unmodified Z domain with free N- terminal cysteine residue by treating with methoxyamine-HCl. The addition of methoxyamine-HCl lowers the pH to ~4 and the nucleophile reacts with the CP to open the five membered thiazolidine ring. The formation of the stable oxime compound shifts the equilibrium towards the formation of the unmodified protein.[112] Easy removal of the excess methoxyamine and the oxime product provides the original unmodified protein (Fig 4.15) that can be reused further for other proteomic application.

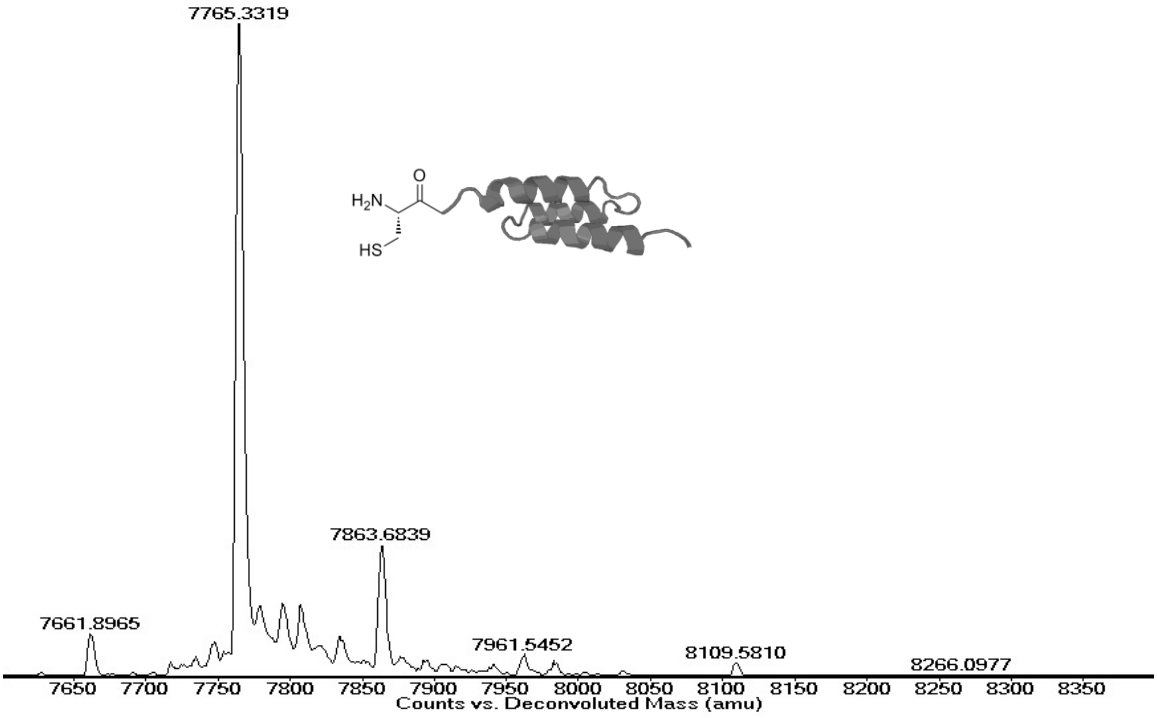


Figure 4.15 Deconvoluted ESI mass spectrum of regenerated T2C Z domain mutant with the treatment with methoxyamine.

4.3.6 In vivo labeling of T2C -Z domain using pyruvate analogs

Pyruvate analogs were directly added to the culture media for *in vivo* modification of T2C Z domain mutant inside the Bl21(DE3) cells. The cells with pET21-Z plasmids were grown either in auto-induction media PA5052 (0.5% glycerol, 0.05% glucose and 2% lactose) or Luria-Bertani (LB) broth with IPTG (Isopropyl β-D-1-thiogalactopyranoside) induction. However, the mass spectrometry analysis of the purified proteins indicated the formation of the Z domain variant with the CP adduct (7835.85 Da). This result clearly indicated the reaction of the newly synthesized T2C Z domain mutant with the cellular

pyruvate. Repetition of the reaction with a very high concentration of pyruvate analogs (~15 mM) to compete against the cellular pyruvate provided the same result.

Pyruvate is a key component in carbohydrate metabolism. For the feasibility of *in vivo* modification, the pyruvate concentration should be maintained just enough for cell growth but not sufficient for reaction with T2C Z domain mutant. Since pyruvate is mostly formed by glycolysis, glucose as the carbon source, especially in complex media like LB and PA5052, directly contributes to the increase in the cellular pyruvate concentration. Hence, other carbon sources that do not directly contribute to pyruvate formation should be used

E. coli can utilize acetate and other compounds as the sole carbon source when glucose in unavailable. The acetate is converted to Acetyl CoA, which is converted to succinate and malate by the glyoxylate cycle.[145] However, this cycle also forms pyruvate since enzymes like phosphoenolpyruvate (PEP) carboxykinase, malic enzyme and oxaloacetate decarboxylase ultimately convert the citric acid cycle intermediates like malate and oxaloacetate directly or indirectly to pyruvate. Nevertheless, using acetate could diminish the pyruvate concentration to such a minimal level that it would not react with the protein N-terminal cysteine in the presence of excess exogenous pyruvate analogs. Hence, M9 minimal media with acetate as the carbon source was used for the growth of BL21(DE3) cells containing the pET21-Z plasmid. As expected, the cell growth was relatively slower. However, the mass spectrometry analysis of the isolated protein still showed the pyruvate adduct. These results suggested that either the concentration of the cellular pyruvate is still relatively high or the pyruvate analogs used in the culture are not cell permeable.

Because of its anti-cancer properties, 3-Bromopyruvic acid has been used for both *in vitro* and *in vivo* studies.[146] Since it is easily taken up by the cell, it was used to monitor its effect on the *in vivo* T2C Z domain mutant modification. The M9 culture with acetate as the carbon source was supplemented with 3-bromopyruvic acid. The expressed protein in this case did not contain the CP adduct, suggesting the lack of reaction with cellular pyruvate. Interestingly, a new modification of the Z domain with peak at 7809.13 Da was obtained (Fig. 4.16). On comparison, the observed mass matched with the thiazolidine analog with hydroxyl group at the C-3 after decarboxylation. Decarboxylation is a spontaneous process in the cell and has been also observed for proteins modified with other pyruvate analogs.

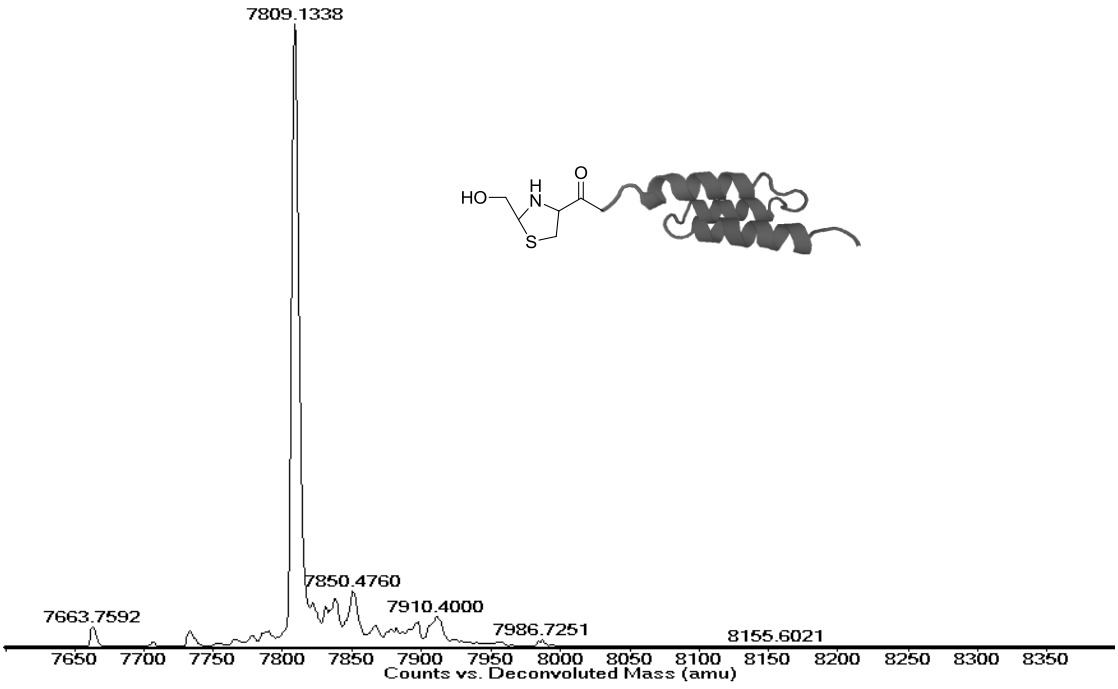


Figure 4.16 Deconvoluted ESI mass spectrum of modified T2C Z domain mutant *in vivo* obtained by using 3-bromopyruvic acid in the M9 culture media with acetate as the carbon source. The structure of the probable modification that matches the observed peak at 7809.13 Da

4.4 Conclusion

Various pyruvate analogs were designed and synthesized. N-terminal cysteine variant of Z domain was obtained by site directed mutagenesis and action of MAP. The synthesized pyruvate analogs were reacted with the N-terminal cysteine to efficiently modify T2C Z domain mutant *in vitro* at physiological condition. This method provides valuable tool in the protein modification of a desired protein with intended modification. The *in vivo* modification studies need further investigation.

4.5 Future Studies

In vivo modification of proteins using pyruvate analogs

In this dissertation we established that proteins with N-terminal Cys can be efficiently modified using pyruvate analogs. The same modification *in vivo* depends on two key factors. The first is the careful design and synthesis of analogs that can easily permeate the cell membrane. The second is the use of an *E. coli* strain lacking enzymes that are responsible for the pyruvate synthesis cultured in M9 minimal media with acetate as the carbon source.

The PEP-pyruvate-oxaloacetate node, which represents the metabolic link between glycolysis/gluconeogenesis and the citric acid cycle, contains a set of enzymes that are responsible for directing the carbon flux into appropriate directions. As a result, this node has been the focal point of several metabolic engineering attempts to improve

the process of biotechnological production of biomass such as succinate, malic acid etc.[147-148] With similar approach engineered strain can be obtained that potentially subsists in minimal amount of pyruvate. Even though there are other enzymes, which are also responsible for pyruvate formation, manipulation of the enzymes shown in Fig 4.17 would be a promising starting point.

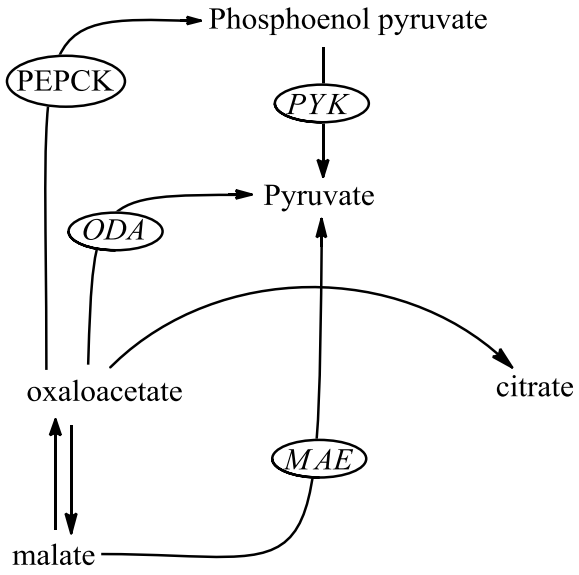


Figure 4.17 Enzymes (enclosed) at the PEP-pyruvate-oxaloacetate node which are responsible for the metabolic formation of pyruvate: PYK (pyruvate kinase), MAE (malic enzyme), ODA (oxaloacetate decarboxylase) and PEPCK (phosphoenolpyruvate carboxykinase).

Synthesis of biologically relevant pyruvate probes

The synthesis of pyruvate analogs should not be limited to small molecular tags. The synthesis of analogs for biological and pharmacological application can be explored. These modifications inlcude lipidation, PEGylation and glycosylation. In fact, some work has already been initiated for the *in vitro* lipidation of T2C Z domain mutant.

Semi-synthesis of proteins by conjugation

Similar to thiazolidine conjugation that utilizes C-terminal aldehyde and N-terminal Cys for the synthesis of proteins, the presented method can also be used for same purpose. The new approach would require a ketoacid moiety at the C-terminal end of one peptide and Cys at the N-terminus of the second peptide. The reaction conditions as shown are very friendly especially for biological samples.

4.6 Experimental methods

4.6.1 General

LB (BD Difco™) and M9 minimal media (Sigma) were prepared according to the manufacturer's standard procedures. M9 minimal medium was supplemented with 0.8 % glycerol (Fisher Scientific) as the carbon source and 0.1 % casamino acids (BD Difco™). LB Agar (BD Difco™) was added at 3.7 % to prepare agar plates. XL1-Blue *E. coli* cells (Stratagene) were used for cloning and maintaining plasmids. All enzymes and biological reagents were purchased from New England Biolabs unless indicated otherwise. *Taq* DNA polymerase was used for polymerase chain reaction (PCR) Synthetic oligonucleotides were obtained from ITD (Integrated DNA Technologies). *E. coli* strain, BL21(DE3) cells were used for protein expression. Quikchange™ II site-directed mutagenesis kit (Stratagene) was used for site-directed mutagenesis. Plasmids were isolated using QIAprep Spin Mini prep Kit (QIAGEN) following the manufacturer's instruction. Ni-NTA metal affinity resin was used for Z domain purification (Invitrogen). DNA and protein concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Fish Scientific).

4.6.2 Construction of recombinant vector pET21-Z

pET21-Z, a plasmid that encodes the Z-domain C-terminal hexa-histidine fusion gene under control of T7 promoter, was generated by inserting the fusion gene between the NdeI and XhoI sites of pET-21a(+) (Novagen). The constructed plasmid was verified by sequencing using Applied Biosystems 3130xl Genetic Analyzer (Hitachi).

4.6.3 Site-directed mutagenesis mutate Thr-Cys at the second position

The genes for the Z-domain variants containing T-C mutation in the second position was generated by PCR of pET-Z with the primer

ZR (5'-GCAGCCGGATCTCAGTGGTGG-3') and

Z2 (5'-ATATACATATGTGTAGTGTAGACAACAAAATCAACAAAGAAC-3') using the protocol described in Quikchange™ kit. The sequence of the new mutation was confirmed by DNA sequencing.

4.6.4 Expression and purification Z domain

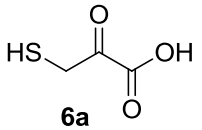
E. coli strain BL21(DE3) was transformed with the plasmids pET-Z containing the Z domain with T-C mutation at the second position. The transformed cells were grown overnight at 37°C either in the PA-5052 auto-induction medium containing 100 μg/mL carbenicillin or high-density-cell IPTG (Isopropyl β-D-1-thiogalactopyranoside) induction method. Cells were grown in LB broth supplemented with 100 μg/ml of carbenicillin at 37 °C until the OD_{600} reached about 0.6. The Z domain expression was induced using 1mM IPTG. Cells were harvested by centrifugation and lysed by sonication in lysis buffer (50mM sodium phosphate, 300 mM NaCl, pH 8.0).

The Z-domain protein in the cell lysate was purified with Ni-NTA metal affinity resin under denaturing conditions according to the manufacturer's protocol (Invitrogen). Purified protein was desalted and concentrated by ultrafiltration using Amicon Ultra 0.5 ml filter 3k devices and analyzed by SDS-PAGE using Coomassie® G-250 stain (Invitrogen, Inc.) or Bio-Rad Silver Stain Plus kit for protein visualization.

4.6.5 General synthetic method

All commercial chemicals were used without further purification. Nuclear magnetic resonance (NMR) spectra were recorded on a Varian INOVA 400 spectrometer. Chemical shifts are reported in δ ppm and coupling constants (*J*) are reported in Hz. High resolution mass spectrometry (ESI-HRMS) data were obtained by using the Fourier-transform ion cyclotron resonance (FTICR) operating in tandem with an LTQ linear ion trap mass spectrometer equipped with an electrospray ionization (ESI) source (LTQ-FT, Thermo).

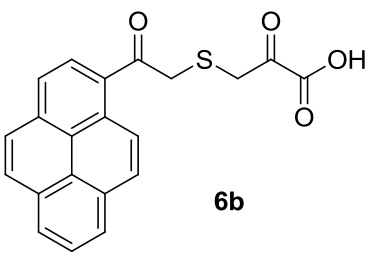
*Synthesis of 3-mercaptopyruvic acid (**6a**)*



6a

3-Mercaptopyruvic acid was prepared according to a reported method.[140] A solution of 3-bromopyruvic acid (2.57 g, 15.38 mmol) in methanol (8ml) was left stirring in an ice bath. To this a cold solution of sodium hydrosulfide (1.72 g, 30.76 mmol) in methanol (8 ml) was added drop-wise in 1.5 hrs. The reaction mixture was left stirring for another 15 min. Addition of ethanol formed precipitate which was collected using Whitman 5 filter paper. The precipitate was washed with additional ethanol to remove the yellow compound and left to dry in vacuum. The dried white precipitate was then dissolved in cold 0.01 M HCl (10 ml) to which ethanol (30 ml) was added. On allowing the mixture to stand at 4 °C for about 4 hours white crystal were formed. The crystals were collected using Whitman 5 filter paper and washed with cold ethanol (200 ml) to yield **6a** (894 mg, 48%). NMR spectra were consistent with literature.[140]

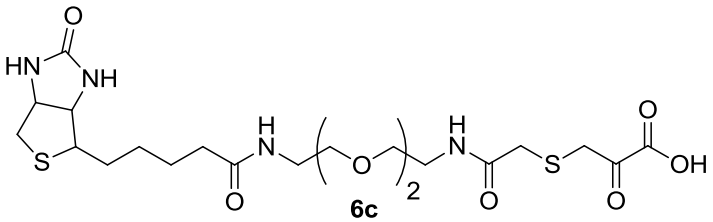
*Synthesis of S-(acetylpyrene)mercaptopyruvic acid (**6b**)*



6b

To a solution of bromoacetyl pyrene (161 mg, 0.5 mmol) in DMF (3 ml), 3-mercaptopyruvic acid (**6a**) (60 mg, 0.5 mmol) and $NaHCO_3$ (168 mg, 2 mmol) were added. The yellowish reaction mixture was left stirring at room temperature for 24 hrs. TLC preformed using 10 % methanol in chloroform showed the absence of the starting material. The reaction mixture was then poured in cold aqueous 1 M HCl (15 ml). The yellow precipitate formed was collected using Whitman 5 filter paper and washed with cold water. The yellow solid was dried to obtain **6b** (104 mg, 57% crude).

*Synthesis of S-(Biotin-PEG_2)mercaptopyruvic acid (**6c**)*



6c

1 M solution each of 3-mercaptopyruivc acid and K_2CO_3 were freshly prepared in water (1 ml). To a solution of biotin-PEG_2-iodoacetamide (10 mg, 18.4 µmol) in water (500 µl), 36.8 µl (2 molar equivalents) of 1 M solution of 3-mercaptopyruvic acid and 55.2 µl (3 molar equivalents) of 1 M solution of K_2CO_3 were added. The reaction was left stirring overnight using IKA MS 3 digital mixer to obtain 6c in quantitative yield. ^{1}H NMR (400 MHz, D_2O) δ 4.50 (dd, J = 8.0, 4.0, 1H, C*H*NH), 4.32 (dd, J = 8.0, 4.0, 1H, C*H*NH), 3.60-3.58 (m, 8H 4xCH_2O), 3.15-3.32 (m, 9H, 2xCH_2NH, 2xCH_2S, C*H*S), 2.88 (dd, J = 16.0, 4, 1H, SC*H*H), 2.67 (d, J = 12.0, 1H, SC*H*H) 2.17 (t, J= 8.0, 2H, CH_2CO)1.46-1.63 (m, 4H,

C*H*$_2$), 1.28-1.34 (m, 2H, C*H*$_2$). ^{13}C NMR (100 MHz, D_2O) δ 177.0 (C=O), 172.3 (C=O), 165.3, 165.3 (C=O), 162.0 (C=O), 69.4, 69.4, 68.9, 68.7 (4 *C*H_2O), 62.1 (*C*H_2NH), 60.2 (*C*H_2NH), 55.3 (S*C*H), 39.7 (S*C*H_2), 39.2 (S*C*H_2), 38.9 (*C*H_2NH) 37.1 (*C*H_2NH), 35.4 (*C*H_2CO), 27.8 (*C*H_2), 27.7(*C*H_2), 25.1(*C*H_2)

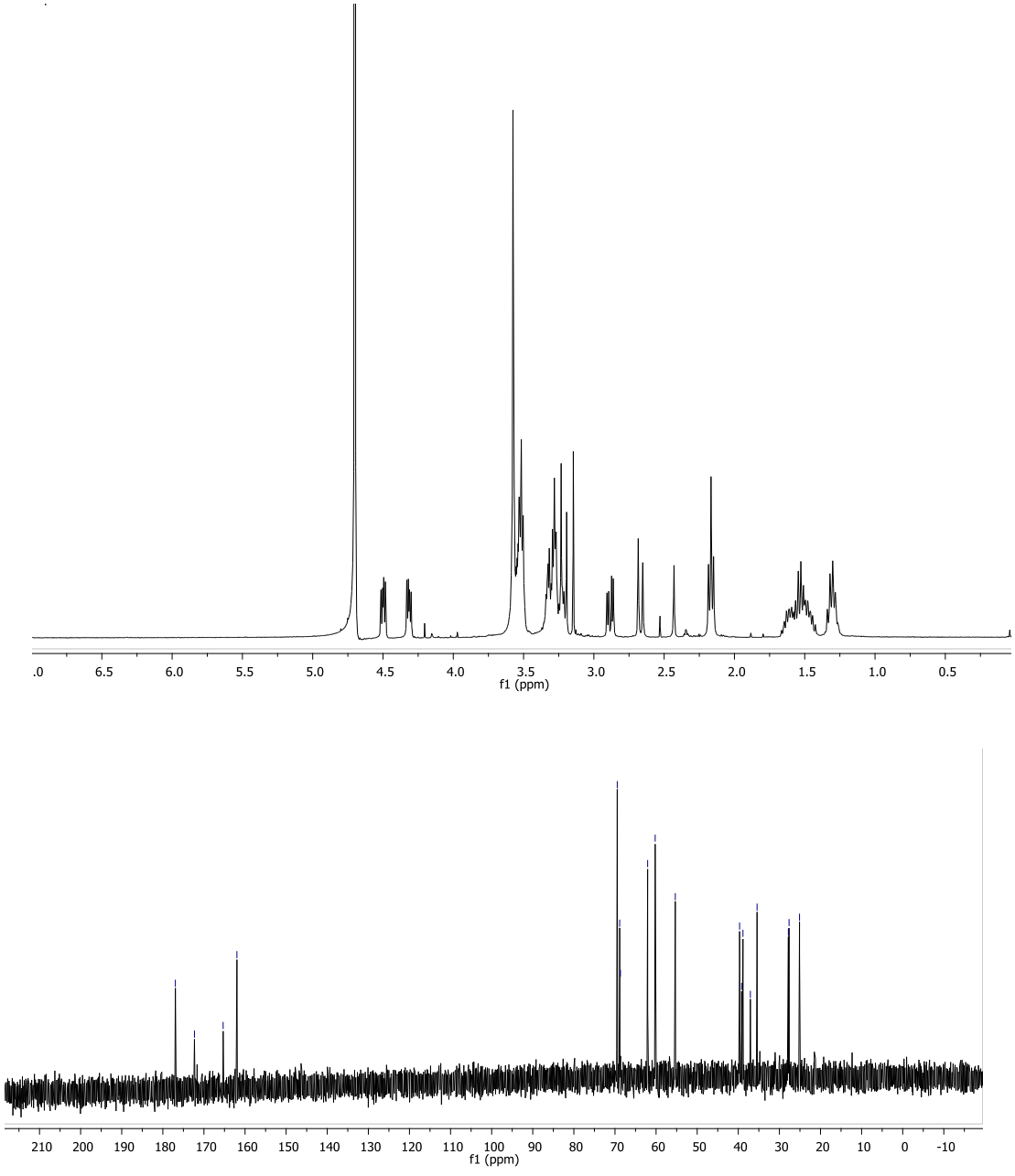
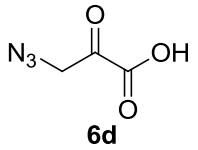


Figure 2.18 ^{1}H and ^{13}C NMR of 6**c**

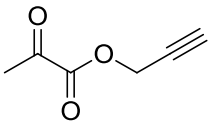
Synthesis of 3-*azido-2-oxo-propanioc acid (**6d**)*



6d

3-azido-2-oxo-propanioc acid was prepared according to a reported method.[141] Sodium azide (0.715 g, 11 mmol) was dissolved in DMSO (20 ml) and was left stirring at room temperature for 40 min. To this a solution of bromopyruvic acid (1.66 g, 10 mmol) in DMSO (20 ml) was added. The reaction was then left stirring at room temperature for 2 hours. The reaction was quenched with water (50 ml). The mixture was then extracted three times with ethyl ether (100 ml). The combined organic layer was then washed with water. Na_2SO_4 was used for drying the and subsequently concentrated *in vacuo* to yield white powder of **6d** (873 mg (76 %). The NMR spectra were consistent with the literature.

*Synthesis of 2-Propynyl 2-oxopropanoate (**6e**)*



6e

2-Propynyl 2-oxopropanoate was synthesized according to a reported method as well.[142] Pyruvi c acid (1.5 g, 17.05 mmol) was mixed with propargyl alcohol (1.15 g, 40.9 mmol) in benzene (60 ml). To which *p*-toluenesulfonic acid (pTsOH) (160 mg, 0.85 mmol) was added. A Dean stark apparatus with benzene in the trap was used for the collection of water. The reaction was refluxed (95 °C) overnight. The presence of water was checked in the trap. The reaction was then cooled and washed with saturated $NaHCO_3$ (2X 50 ml) followed by water (2X 50 ml). The organic layer was dried using $MgSO_4$ and concentrated in vacuo to yield of reddish oil, **6e (**329 mg, 15 %). NMR spectra were consistent with literature.

4.6.6 Reaction of pyruvate analogs with T2C Z domain mutant

50 µl reactions were prepared in 1X PBS buffer with 3 µl of T2C Z domain mutant (10.88 µg/µl), 5 µl of 0.5 mM TCEP and 25 µl of 0.5 mM of pyruvate analogs. Prior to the addition of the pyruvate analogs, the reaction was left stirring only in the presence of TCEP for about 20 min for the complete reduction of disulfide bond to form free N-terminal Cys residue. After the addition of the pyruvate analogs the reactions were left overnight at room temperature. Some analogs which showed low solubility in water were dissolved in DMSO prior to the addition to the reaction mixture. After the completion of reaction, the protein samples were isolated and concentrated using Amicon Ultra 0.5 ml filter 3K devices and analyzed using mass spectrometry.

To find the amount of analog required for the complete modification, four separate 50 µl reactions were set up using the condition mentioned above with each set containing 22.6 µg (2.91 nmol) of Cys-Z domain with 0.1 µmoles, 1 µmoles, 5 µmoles and 10 µmoles of 3-mercatopyruvate. Proteins samples were isolated and concentrated as described above.

Similarly, for the reaction kinetic study the reactions were set up as above but each contained 10 µmoles of the 3-mercatpyruvic acid and were incubated for 1, 2, 3, and 4 hours.

4.6.7 Re-generation of the unmodified protein

50 µl reactions were prepared in 1X PBS buffer with 3 µl of modified T2C Z domain mutant (10.88 µg/µl), and 25 µl of 0.5 mM methoxyamine-HCl. The reaction samples were stirred overnight at room temperature. Excess methoxyamine and the oxime byproduct were removed by ultrafiltration using the Amicon Ultra 0.5 ml filter 3Kdevices. The samples were then analyzed by ESI- mass spectrometry.

4.6.8 Mass spectrometry measurements

The Z-domain samples were analyzed by Agilent technologies 6224 Accurate-Mass TOF LC/MS system equipped with an ESI source. Multiple-charged protein ions were de-convoluted by using the bundled MassHunter Workstation software Qualitative Analysis B.04.00.

4.6.9 Calculation of theoretical mass of modified T2C Z domain mutant

The theoretical mass of T2C Z domain and its derivatives were calculated based upon its sequence CSVDNKINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAKKLND AQAPKGSHHHHH and the chemical structure of the incorporated labels using ChemDraw. The average calculated theoretical mass of Cys-Z domain is 7764.41 Da.

References

1. Yanofsky, C.; Platt, T.; Crawford, I. P.; Nichols, B. P.; Christie, G. E.; Horowitz, H.; VanCleemput, M.; Wu, A. M., The complete nucleotide sequence of the tryptophan operon of Escherichia coli. *Nucleic. Acids Res.* **1981,** *9* (24), 6647-68.

2. Narberhaus, F.; Waldminghaus, T.; Chowdhury, S., RNA thermometers. *FEMS Microbiol. Rev.* **2006,** *30* (1), 3-16.

3. Storz, G.; Vogel, J.; Wassarman, K. M., Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **2011,** *43* (6), 880-91.

4. Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R., Genetic control by a metabolite binding mRNA. *Chem. Biol.* **2002,** *9* (9), 1043-1049.

5. Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (6), 451-63.

6. Roth, A.; Breaker, R. R., The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **2009,** *78*, 305-34.

7. Lin, J. C.; Thirumalai, D., Gene Regulation by Riboswitches with and without Negative Feedback Loop. *Biophys. J.* **2012,** *103* (11), 2320-2330.

8. Welz, R.; Breaker, R. R., Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *Rna-a Publication of the Rna Society* **2007,** *13* (4), 573-582.

9. Nahvi, A.; Barrick, J. E.; Breaker, R. R., Coenzyme B-12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic. Acids Res.* **2004,** *32* (1), 143-150.

10. Weinberg, Z.; Regulski, E. E.; Hammond, M. C.; Barrick, J. E.; Yao, Z.; Ruzzo, W. L.; Breaker, R. R., The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA* **2008,** *14* (5), 822-8.

11. Ames, T. D.; Rodionov, D. A.; Weinberg, Z.; Breaker, R. R., A Eubacterial Riboswitch Class That Senses the Coenzyme Tetrahydrofolate. *Chem. Biol.* **2010,** *17* (7), 681-685.

12. Blount, K. F.; Wang, J. X.; Lim, J.; Sudarsan, N.; Breaker, R. R., Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol.* **2007,** *3* (1), 44-9.

13. Kwon, M.; Strobel, S. A., Chemical basis of glycine riboswitch cooperativity. *RNA* **2008,** *14* (1), 25-34.

14. Mandal, M.; Breaker, R. R., Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **2004,** *11* (1), 29-35.

15. Jenkins, J. L.; Krucinska, J.; McCarty, R. M.; Bandarian, V.; Wedekind, J. E., Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* **2011,** *286* (28), 24626-37.

16. Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R., Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **2004,** *428* (6980), 281-6.

17. Ramesh, A.; Winkler, W. C., Magnesium-sensing riboswitches in bacteria. *RNA Biol* **2010,** *7* (1), 77-83.

18. Gralla, J.; Steitz, J. A.; Crothers, D. M., Direct physical evidence for secondary structure in an isolated fragment of R17 bacteriophage mRNA. *Nature* **1974,** *248* (445), 204-8.

19. Miranda-Rios, J.; Navarro, M.; Soberon, M., A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98* (17), 9736-41.

20. Sudarsan, N.; Barrick, J. E.; Breaker, R. R., Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **2003,** *9* (6), 644-7.

21. Barrick, J. E.; Breaker, R. R., The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **2007,** *8* (11), R239.

22. Garrity, D. B.; Zahler, S. A., Mutations in the gene for a tRNA that functions as a regulator of a transcriptional attenuator in Bacillus subtilis. *Genetics* **1994,** *137* (3), 627-36.

23. Poiata, E.; Meyer, M. M.; Ames, T. D.; Breaker, R. R., A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA* **2009,** *15* (11), 2046-56.

24. Barrick, J. E.; Corbino, K. A.; Winkler, W. C.; Nahvi, A.; Mandal, M.; Collins, J.; Lee, M.; Roth, A.; Sudarsan, N.; Jona, I.; Wickiser, J. K.; Breaker, R. R., New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (17), 6421-6.

25. Dambach, M. D.; Winkler, W. C., Expanding roles for metabolite-sensing regulatory RNAs. *Curr. Opin. Microbiol.* **2009,** *12* (2), 161-9.

26. Kubodera, T.; Watanabe, M.; Yoshiuchi, K.; Yamashita, N.; Nishimura, A.; Nakai, S.; Gomi, K.; Hanamoto, H., Thiamine-regulated gene expression of

Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS Lett.* **2003,** *555* (3), 516-20.

27. Cheah, M. T.; Wachter, A.; Sudarsan, N.; Breaker, R. R., Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **2007,** *447* (7143), 497-500.

28. Blount, K. F.; Breaker, R. R., Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* **2006,** *24* (12), 1558-64.

29. Shiota, T.; Folk, J. E.; Tietze, F., Inhibition of lysine utilization in bacteria by S-(beta-aminoethyl) cysteine and its reversal by lysine peptides. *Arch. Biochem. Biophys.* **1958,** *77* (2), 372-7.

30. Mccord, T. J.; Ravel, J. M.; Skinner, C. G.; Shive, W., Dl-4-Oxalysine, an Inhibitory Analog of Lysine. *J. Am. Chem. Soc.* **1957,** *79* (21), 5693-5696.

31. Ataide, S. F.; Wilson, S. N.; Dang, S.; Rogers, T. E.; Roy, B.; Banerjee, R.; Henkin, T. M.; Ibba, M., Mechanisms of resistance to an amino acid antibiotic that targets translation. *ACS Chem Biol* **2007,** *2* (12), 819-27.

32. Sudarsan, N.; Cohen-Chalamish, S.; Nakamura, S.; Emilsson, G. M.; Breaker, R. R., Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem. Biol.* **2005,** *12* (12), 1325-35.

33. Iwashima, A.; Wakabayashi, Y.; Nose, Y., Formation of pyrithiamine pyrophosphate in brain tissue. *J. Biochem.* **1976,** *79* (4), 845-7.

34. Ellington, A. D.; Szostak, J. W., Invitro Selection of Rna Molecules That Bind Specific Ligands. *Nature* **1990,** *346* (6287), 818-822.

35. Tuerk, C.; Gold, L., Systematic Evolution of Ligands by Exponential Enrichment - Rna Ligands to Bacteriophage-T4 DNA-Polymerase. *Science* **1990,** *249* (4968), 505-510.

36. Robertson, D. L.; Joyce, G. F., Selection Invitro of an Rna Enzyme That Specifically Cleaves Single-Stranded-DNA. *Nature* **1990,** *344* (6265), 467-468.

37. Werstuck, G.; Green, M. R., Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **1998,** *282* (5387), 296-8.

38. Suess, B.; Hanson, S.; Berens, C.; Fink, B.; Schroeder, R.; Hillen, W., Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic. Acids Res.* **2003,** *31* (7), 1853-1858.

39. Suess, B.; Fink, B.; Berens, C.; Stentz, R.; Hillen, W., A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic. Acids Res.* **2004,** *32* (4), 1610-1614.

40. Hunsicker, A.; Steber, M.; Mayer, G.; Meitert, J.; Klotzsche, M.; Blind, M.; Hillen, W.; Berens, C.; Suess, B., An RNA aptamer that induces transcription. *Chem. Biol.* **2009,** *16* (2), 173-80.

41. Tang, J.; Breaker, R. R., Rational design of allosteric ribozymes. *Chem. Biol.* **1997,** *4* (6), 453-9.

42. Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R., A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **2004,** *306* (5694), 275-9.

43. Regulski, E. E.; Breaker, R. R., In-line probing analysis of riboswitches. *Methods Mol. Biol.* **2008,** *419*, 53-67.

44. Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R., Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **2003,** *113* (5), 577-86.

45. McDaniel, B. A.; Grundy, F. J.; Artsimovitch, I.; Henkin, T. M., Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (6), 3083-8.

46. Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T., Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **2008,** *15* (2), 177-82.

47. Rodionov, D. A.; Vitreschak, A. G.; Mironov, A. A.; Gelfand, M. S., Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic. Acids Res.* **2003,** *31* (23), 6748-57.

48. Lakowicz, J. R.; Gryczynski, I. I.; Gryczynski, Z., High Throughput Screening with Multiphoton Excitation. *J. Biomol. Screen* **1999,** *4* (6), 355-362.

49. Wickiser, J. K.; Winkler, W. C.; Breaker, R. R.; Crothers, D. M., The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **2005,** *18* (1), 49-60.

50. Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M., The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **2005,** *44* (40), 13404-14.

51. Serganov, A.; Huang, L.; Patel, D. J., Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* **2008,** *455* (7217), 1263-7.

52. Begley, T. P.; Downs, D. M.; Ealick, S. E.; McLafferty, F. W.; Van Loon, A. P.; Taylor, S.; Campobasso, N.; Chiu, H. J.; Kinsland, C.; Reddick, J. J.; Xi, J., Thiamin biosynthesis in prokaryotes. *Arch. Microbiol.* **1999,** *171* (5), 293-300.

53. Serganov, A.; Polonskaia, A.; Phan, A. T.; Breaker, R. R.; Patel, D. J., Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **2006,** *441* (7097), 1167-71.

54. Thore, S.; Leibundgut, M.; Ban, N., Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **2006,** *312* (5777), 1208-11.

55. Edwards, T. E.; Ferre-D'Amare, A. R., Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **2006,** *14* (9), 1459-68.

56. Patnaik, A. K.; Rao, N. S.; Kumar, P.; Sharma, A. K.; Garg, B. S.; Gupta, K. C., High-loading supports for oligonucleotide synthesis. *Helv. Chim. Acta* **2000,** *83* (2), 322-327.

57. Maier, G. D.; Metzler, D. E., Structures of Thiamine in Basic Solution. *J. Am. Chem. Soc.* **1957,** *79* (16), 4386-4391.

58. Page, P. C.; McKenzie, M. J.; Gallagher, J. A., Novel synthesis of bis(phosphonic acid)-steroid conjugates. *J. Org. Chem.* **2001,** *66* (11), 3704-8.

59. Saady, M.; Lebeau, L.; Mioskowski, C., Selective Monodeprotection of Phosphate, Phosphite, Phosphonate, and Phosphoramide Benzyl-Esters. *J. Org. Chem.* **1995,** *60* (9), 2946-2947.

60. Winkler, W.; Nahvi, A.; Breaker, R. R., Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **2002,** *419* (6910), 952-6.

61. Ware, W. R., Oxygen Quenching of Fluorescence in Solution - an Experimental Study of Diffusion Process. *J. Phys. Chem.* **1962,** *66* (3), 455-&.

62. Fischkoff, S.; Vanderkooi, J. M., Oxygen diffusion in biological and artificial membranes determined by the fluorochrome pyrene. *J. Gen. Physiol.* **1975,** *65* (5), 663-76.

63. Ribou, A. C.; Salmon, J. M.; Vigo, J.; Goyet, C., Measurements of calcium with a fluorescent probe Rhod-5N: Influence of high ionic strength and pH. *Talanta* **2007,** *71* (1), 437-42.

64. Vaughan, W. M.; Weber, G., Oxygen quenching of pyrenebutyric acid fluorescence in water. A dynamic probe of the microenvironment. *Biochemistry* **1970,** *9* (3), 464-73.

65. Weber, G., Polarization of the fluorescence of macromolecules. II. Fluorescent conjugates of ovalbumin and bovine serum albumin. *Biochem. J.* **1952,** *51* (2), 155-67.

66. Dai, Z.; Dulyaninova, N. G.; Kumar, S.; Bresnick, A. R.; Lawrence, D. S., Visual snapshots of intracellular kinase activity at the onset of mitosis. *Chem. Biol.* **2007,** *14* (11), 1254-60.

67. Levi, J.; Cheng, Z.; Gheysens, O.; Patel, M.; Chan, C. T.; Wang, Y.; Namavari, M.; Gambhir, S. S., Fluorescent fructose derivatives for imaging breast cancer cells. *Bioconjug. Chem.* **2007,** *18* (3), 628-34.

68. Skrzypczynski, Z.; Wayland, S., A modular approach to the synthesis of new reagents useful in the chemical synthesis of modified DNA probes: derivatives of 3-(tert-butyldimethylsiloxy)glutaric anhydride as versatile building blocks in the synthesis of new phosphoramidites and modified solid supports. *Bioconjug. Chem.* **2004,** *15* (3), 583-93.

69. Beylin, V. G.; Goel, O. P., A Convenient Synthesis of Tert-Butyl N-(2-Bromoethyl)Carbamate. *Org. Prep. Proced. Int.* **1987,** *19* (1), 78-80.

70. Pennisi, E., Genomics. ENCODE project writes eulogy for junk DNA. *Science* **2012,** *337* (6099), 1159, 1161.

71. Black, D. L., Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **2003,** *72*, 291-336.

72. Leff, S. E.; Rosenfeld, M. G.; Evans, R. M., Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* **1986,** *55*, 1091-117.

73. Rosenfeld, M. G.; Lin, C. R.; Amara, S. G.; Stolarsky, L.; Roos, B. A.; Ong, E. S.; Evans, R. M., Calcitonin mRNA polymorphism: peptide switching associated with alternative RNA splicing events. *Proc. Natl. Acad. Sci. U. S. A.* **1982,** *79* (6), 1717-21.

74. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **2005,** *44* (45), 7342-72.

75. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene expression. *Science* **1994,** *263* (5148), 802-5.

76. Hochuli, E.; Bannwarth, W.; Dobeli, H.; Gentz, R.; Stuber, D., Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio-Technology* **1988,** *6* (11), 1321-1325.

77. Ericsson, E. M.; Enander, K.; Bui, L.; Lundstrom, I.; Konradsson, P.; Liedberg, B., Site-specific and covalent attachment of his-tagged proteins by chelation assisted photoimmobilization: a strategy for microarraying of protein ligands. *Langmuir* **2013,** *29* (37), 11687-94.

78. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **2010,** *79*, 413-44.

79. Feeney, R. E.; Yamasaki, R. B.; Geoghegan, K. F., Chemical Modification of Proteins - an Overview. *Advances in Chemistry Series* **1982,** (198), 3-55.

80. Dixon, H. B. F., N-Terminal Modification of Proteins - a Review. *J. Protein Chem.* **1984,** *3* (1), 99-108.

81. Fodje, M. N.; Al-Karadaghi, S., Occurrence, conformational features and amino acid propensities for the pi-helix. *Protein Eng.* **2002,** *15* (5), 353-8.

82. Goddard, D. R.; Schubert, M. P., The action of iodoethyl alcohol on thiol compounds and on proteins. *Biochem. J.* **1935,** *29* (5), 1009-11.

83. Kaiser, E. T.; Kezdy, F. J., Amphiphilic secondary structure: design of peptide hormones. *Science* **1984,** *223* (4633), 249-55.

84. Gundlach, H. G.; Moore, S.; Stein, W. H., The reaction of iodoacetate with methionine. *J. Biol. Chem.* **1959,** *234* (7), 1761-4.

85. Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M., Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat. Methods* **2008,** *5* (6), 459-60.

86. Rizzi, S. C.; Hubbell, J. A., Recombinant protein-co-PEG networks as cell-adhesive and proteolytically degradable hydrogel matrixes. Part I: Development and physicochemical characteristics. *Biomacromolecules* **2005,** *6* (3), 1226-38.

87. Schelte, P.; Boeckler, C.; Frisch, B.; Schuber, F., Differential reactivity of maleimide and bromoacetyl functions with thiols: application to the preparation of liposomal diepitope constructs. *Bioconjug. Chem.* **2000,** *11* (1), 118-23.

88. Smith, M. E.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R., Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.* **2010,** *132* (6), 1960-5.

89. Morpurgo, M.; Veronese, F. M.; Kachensky, D.; Harris, J. M., Preparation of characterization of poly(ethylene glycol) vinyl sulfone. *Bioconjug. Chem.* **1996,** *7* (3), 363-8.

90. Morales-Sanfrutos, J.; Lopez-Jaramillo, J.; Ortega-Munoz, M.; Megia-Fernandez, A.; Perez-Balderas, F.; Hernandez-Mateo, F.; Santoyo-Gonzalez, F., Vinyl sulfone: a versatile function for simple bioconjugation and immobilization. *Org. Biomol. Chem.* **2010,** *8* (3), 667-75.

91. Shiu, H. Y.; Chan, T. C.; Ho, C. M.; Liu, Y.; Wong, M. K.; Che, C. M., Electron-deficient alkynes as cleavable reagents for the modification of cysteine-containing peptides in aqueous medium. *Chemistry* **2009,** *15* (15), 3839-50.

92. Smith, H. B.; Hartman, F. C., Restoration of activity to catalytically deficient mutants of ribulosebisphosphate carboxylase/oxygenase by aminoethylation. *J. Biol. Chem.* **1988,** *263* (10), 4921-5.

93. Fontana, A.; Scoffone, E.; Benassi, C. A., Sulfenyl halides as modifying reagents for polypeptides and proteins. II. Modification of cysteinyl residues. *Biochemistry* **1968,** *7* (3), 980-6.

94. Davis, B. G., Synthesis of glycoproteins. *Chem. Rev.* **2002,** *102* (2), 579-602.

95. Ellman, G. L., Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959,** *82* (1), 70-7.

96. Boccu, E.; Veronese, F. M.; Fontana, A.; Benassi, C. A., Sulfenyl halides as modifying reagents for polypeptides and proteins. Quantitative evaluation of tryptophan and cystei residues in proteins. *Eur. J. Biochem.* **1970,** *13* (1), 188-92.

97. Wender, P. A.; Goun, E. A.; Jones, L. R.; Pillow, T. H.; Rothbard, J. B.; Shinde, R.; Contag, C. H., Real-time analysis of uptake and bioactivatable cleavage of luciferin-transporter conjugates in transgenic reporter mice. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (25), 10340-5.

98. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.

99. Lu, W. Y.; Qasim, M. A.; Kent, S. B. H., Comparative total syntheses of turkey ovomucoid third domain by both stepwise solid phase peptide synthesis and native chemical ligation. *J. Am. Chem. Soc.* **1996,** *118* (36), 8518-8523.

100. Dawson, P. E.; Kent, S. B., Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000,** *69*, 923-60.

101. Camarero, J. A.; Shekhtman, A.; Campbell, E. A.; Chlenov, M.; Gruber, T. M.; Bryant, D. A.; Darst, S. A.; Cowburn, D.; Muir, T. W., Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (13), 8536-41.

102. Schwarzer, D.; Cole, P. A., Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.* **2005,** *9* (6), 561-9.

103. Liu, C. F.; Tam, J. P., Peptide segment ligation strategy without use of protecting groups. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91* (14), 6584-8.

104. Liu, C. F.; Tam, J. P., Chemical Ligation Approach to Form a Peptide-Bond between Unprotected Peptide Segments - Concept and Model Study. *J. Am. Chem. Soc.* **1994,** *116* (10), 4149-4153.

105. Liu, C. F.; Rao, C.; Tam, J. P., Orthogonal ligation of unprotected peptide segments through pseudoproline formation for the synthesis of HIV-1 protease. *J. Am. Chem. Soc.* **1996,** *118* (2), 307-312.

106. Zhang, L.; Torgerson, T. R.; Liu, X. Y.; Timmons, S.; Colosia, A. D.; Hawiger, J.; Tam, J. P., Preparation of functionally active cell-permeable peptides by single-step ligation of two peptide modules. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (16), 9184-9.

107. Tam, J. P.; Yu, Q.; Miao, Z., Orthogonal ligation strategies for peptide and protein. *Biopolymers* **1999,** *51* (5), 311-32.

108. Schubert, M. P., Combination of thiol acids with methylglyoxal. *J. Biol. Chem.* **1935,** *111* (3), 671-678.

109. Schubert, M. P., Compounds of thiol acids with aldehydes. *J. Biol. Chem.* **1936,** *114* (1), 341-350.

110. Ratner, S. C., H.T., The action of Formaldehyde upon Cysteine. *J. Am. Chem. Soc.* **1937,** *59* (1), 200-206.

111. Sheehan, J. C.; Yang, D. D. H., A New Synthesis of Cysteinyl Peptides. *J. Am. Chem. Soc.* **1958,** *80* (5), 1158-1164.

112. Wohr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X. C.; Mutter, M., Pseudo-prolines as a solubilizing, structure-disrupting protection technique in peptide synthesis. *J. Am. Chem. Soc.* **1996,** *118* (39), 9218-9227.

113. Mutter, M.; Nefzi, A.; Sato, T.; Sun, X.; Wahl, F.; Wohr, T., Pseudo-Prolines (Psi-Pro) for Accessing Inaccessible Peptides. *Pept. Res.* **1995,** *8* (3), 145-153.

114. Nagasawa, H. T.; Goon, D. J.; Muldoon, W. P.; Zera, R. T., 2-Substituted thiazolidine-4(R)-carboxylic acids as prodrugs of L-cysteine. Protection of mice against acetaminophen hepatotoxicity. *J. Med. Chem.* **1984,** *27* (5), 591-6.

115. Porta, P.; Aebi, S.; Summer, K.; Lauterburg, B. H., L-2-Oxothiazolidine-4-Carboxylic Acid, a Cysteine Prodrug - Pharmacokinetics and Effects on Thiols in Plasma and Lymphocytes in Human. *J. Pharmacol. Exp. Ther.* **1991,** *257* (1), 331-334.

116. Schubert, M. P., Reactions of Semimercaptals with amino compounds. *J. Biol. Chem.* **1937,** *121*, 539-548.

117. Dann, J. R.; Oliver, G. L.; Gates, J. W., The Reaction of Cystine and Lanthionine with Aqueous Calcium Hydroxide - the Identification of 2-Methylthiazolidine-2,4-Dicarboxylic Acid. *J. Am. Chem. Soc.* **1957,** *79* (7), 1644-1649.

118. Kredich, N. M.; Foote, L. J.; Keenan, B. S., Stoichiometry and Kinetics of Inducible Cysteine Desulfhydrase from Salmonella-Typhimurium. *J. Biol. Chem.* **1973,** *248* (17), 6187-6196.

119. Dugaiczyk, A.; Malecki, M. T.; Eiler, J. J., The effect of cysteine on L-alpha-glycerophosphate and lactate dehydrogenase reactions. *J. Biol. Chem.* **1968,** *243* (9), 2236-40.

120. Guarneros, G.; Ortega, M. V., Cysteine desulfhydrase activities of Salmonella typhimurium and Escherichia coli. *Biochim. Biophys. Acta* **1970,** *198* (1), 132-42.

121. Wlodek, L.; Czubak, J., Formation of 2-methyl-2,4-thiazolidinedicarboxylic acid from L-cysteine in rat tissues. *Acta Biochim. Pol.* **1984,** *31* (3), 279-88.

122. Rose, K.; Simona, M. G.; Savoy, L. A.; Regamey, P. O.; Green, B. N.; Clore, G. M.; Gronenborn, A. M.; Wingfield, P. T., Pyruvic acid is attached through its central carbon atom to the amino terminus of the recombinant DNA-derived DNA-binding protein Ner of bacteriophage Mu. *J. Biol. Chem.* **1992,** *267* (27), 19101-6.

123. Gentle, I. E.; De Souza, D. P.; Baca, M., Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug. Chem.* **2004,** *15* (3), 658-663.

124. Bernal-Perez, L. F.; Sahyouni, F.; Prokai, L.; Ryu, Y., RimJ-mediated context-dependent N-terminal acetylation of the recombinant Z-domain protein in Escherichia coli. *Mol. BioSyst.* **2012,** *8* (4), 1128-30.

125. King, F. E.; Clarklewis, J. W.; Wade, R., Syntheses from Phthalimido-Acids .8. Synthesis of Glutathione by a New Route to Cysteinyl-Peptides. *J. Chem. Soc.* **1957,** (Feb), 880-885.

126. Kemp, D. S.; Carey, R. I., Boc-L-Dmt-Oh as a Fully N,S-Blocked Cysteine Derivative for Peptide-Synthesis by Prior Thiol Capture - Facile Conversion of N-Terminal Boc-L-Dmt-Peptides to H-Cys(Scm)-Peptides. *J. Org. Chem.* **1989,** *54* (15), 3640-3646.

127. Villain, M.; Vizzavona, J.; Rose, K., Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. *Chem. Biol.* **2001,** *8* (7), 673-9.

128. Bang, D.; Kent, S. B., A one-pot total synthesis of crambin. *Angew. Chem. Int. Ed. Engl.* **2004,** *43* (19), 2534-8.

129. Dirksen, A.; Dawson, P. E., Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug. Chem.* **2008,** *19* (12), 2543-8.

130. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G., Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **2003,** *125* (39), 11782-3.

131. Prescher, J. A.; Dube, D. H.; Bertozzi, C. R., Chemical remodelling of cell surfaces in living animals. *Nature* **2004,** *430* (7002), 873-7.

132. Song, W.; Wang, Y.; Qu, J.; Lin, Q., Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells. *J. Am. Chem. Soc.* **2008,** *130* (30), 9654-5.

133. Tolbert, T. J.; Wong, C. H., New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew. Chem. Int. Ed. Engl.* **2002,** *41* (12), 2171-4.

134. Adams, J. M., On the release of the formyl group from nascent protein. *J. Mol. Biol.* **1968,** *33* (3), 571-89.

135. Bradshaw, R. A.; Brickey, W. W.; Walker, K. W., N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. *Trends Biochem. Sci.* **1998,** *23* (7), 263-7.

136. Giglione, C.; Pierre, M.; Meinnel, T., Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. *Mol. Microbiol.* **2000,** *36* (6), 1197-205.

137. Xiao, Q.; Zhang, F.; Nacev, B. A.; Liu, J. O.; Pei, D., Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry* **2010,** *49* (26), 5588-99.

138. Gonzalez-Ros, J. M.; Calvo-Fernandez, P.; Sator, V.; Martinez-Carrion, M., Pyrenesulfonyl azide as a fluorescent label for the study of protein-lipid boundaries of acetylcholine receptors in membranes. *J. Supramol. Struct.* **1979,** *11* (3), 327-38.

139. Hofmann, K.; Wood, S. W.; Brinton, C. C.; Montibeller, J. A.; Finn, F. M., Imino-Biotin Affinity Columns and Their Application to Retrieval of Streptavidin. *Proc. Natl. Acad. Sci. U. S. A.* **1980,** *77* (8), 4666-4668.

140. Tanabe, S.; Ogasawara, Y.; Nawata, M.; Kawanabe, K., Preparation of a Sulfurtransferase Substrate, Sodium 3-Mercaptopyruvate, from 3-Bromopyruvic Acid and Sodium Hydrosulfide. *Chem. Pharm. Bull. (Tokyo)* **1989,** *37* (10), 2843-2845.

141. Chorev, M. S. S., Chestnut Hill, Massachusetts, 02467, US), AKTAS, Bertal Huseyin (35 Temple Street, Newton, Massachusetts, 02465, US), HALPERIN, José A. (1443 Beacon Street, Brookline, Massachusetts, 02446, US), WAGNER, Gerhard (111 Princeton Street, Chestnut Hill, Massachusetts, 02467, US), Compounds for the Inhinition of Cellular Proliferation 2012.

142. Heaney, F.; Fenlon, J.; O'Mahony, C.; McArdle, P.; Cunningham, D., Alpha-oximono-esters as precursors to heterocycles--generation of oxazinone N-oxides and cycloaddition to alkene dipolarophiles. *Org. Biomol. Chem.* **2003,** *1* (23), 4302-16.

143. Chowdhury, S. K.; Katta, V.; Beavis, R. C.; Chait, B. T., Origin and removal of adducts (molecular mass = 98 u) attached to peptide and protein ions in electrospray ionization mass spectra. *J. Am. Soc. Mass Spectrom.* **1990,** *1* (5), 382-8.

144. Chaiet, L.; Wolf, F. J., The Properties of Streptavidin, a Biotin-Binding Protein Produced by Streptomycetes. *Arch. Biochem. Biophys.* **1964,** *106*, 1-5.

145. Lorenz, M. C.; Fink, G. R., Life and death in a macrophage: role of the glyoxylate cycle in virulence. *Eukaryot Cell* **2002,** *1* (5), 657-62.

146. Chen, Z.; Zhang, H.; Lu, W.; Huang, P., Role of mitochondria-associated hexokinase II in cancer cell death induced by 3-bromopyruvate. *Biochim. Biophys. Acta* **2009,** *1787* (5), 553-60.

147. Zelle, R. M.; de Hulster, E.; van Winden, W. A.; de Waard, P.; Dijkema, C.; Winkler, A. A.; Geertman, J. M.; van Dijken, J. P.; Pronk, J. T.; van Maris, A. J., Malic acid production by Saccharomyces cerevisiae: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Appl. Environ. Microbiol.* **2008,** *74* (9), 2766-77.

148. Lee, J.; Sim, S. J.; Bott, M.; Um, Y.; Oh, M. K.; Woo, H. M., Succinate production from CO(2)-grown microalgal biomass as carbon source using engineered Corynebacterium glutamicum through consolidated bioprocessing. *Sci. Rep.* **2014,** *4*, 5819.

Curriculum Vitae

Pradeep Budhathoki

8449 Miami Springs Dr.
Fort Worth TX 76123
214-686-9187
pradeepth35@gmail.com

Department of Chemistry
TCU Box 298860,
Fort Worth, TX 76129
214-686-9187
p.s.budhathoki@tcu.edu

Education

- Ph. D. – Texas Christian University, Fort Worth, TX (2015)
- B.S., Chemistry – Texas Christian University, Fort Worth, TX (2009)

Research experience

- *2009-2015 Graduate Research Assistant*
  Department of Chemistry, Texas Christian University
  Designed and synthesized various metabolite analogs to monitor binding interactions with their corresponding riboswitches
  Designed and synthesized pyruvate analogs to reversibly modify N-terminal cysteine containing proteins in vitro and in vivo.
- *2008-2009 Undergraduate Research*
  Department of Chemistry, Texas Christian University
  In vitro selection of peptides that recognize riboswitch conformational changes from combinatorial phage display libraries.

Skills

- *Biophysical methods*: fluorescence and UV-Vis spectroscopy for investigating binding constants, isothermal titration calorimetry.
- *Molecular biology*: DNA cloning, DNA sequencing, isolation of genomic and plasmid DNA, agarose gel electrophoresis, RNA transcription and isolation, site-directed mutagenesis.
- *Protein biochemistry*: protein expression, protein purification using size exclusion chromatography, FPLC and affinity chromatography, selective labeling of proteins, PAGE, Western blots, dialysis, mass spectrometry analysis of proteins,
- *Organic Synthesis*: Design and synthesis of small molecules, purification using silica gel chromatography and preparative HPLC, NMR, ESI Mass spectrometry, HPLC-MS.
- *Team and Management*: Supervised 7 undergraduate and 4 high school students, maintenance of all lab equipment, creating inventory for chemicals, cells, plasmids, enzymes and glycerol stocks, ordering chemicals and reagents.
- *Teaching*: Taught several semesters of biochemistry lab course on topics that included fundamentals on laboratory techniques, bacterial transformation, protein extraction and purification, proteomics, DNA sequencing and fingerprinting.

Symposium and Meeting

- The 28th Annual Symposium of the Protein society July27-30, 2014 San Diego, CA
- 47th ACS DFW Meeting in Miniature Texas Wesleyan University, Fort Worth, TX Saturday, April 26, 2014 (*1st Place Graduate level Presenter*)
- Annual Student Research Symposium Texas Christian University April, 11th 2014
- 44th ACS DFW Meeting in Miniature Tarleton State University Stephenville, Texas Saturday, April 16. 2011
- Annual Student Research Symposium Texas Christian University April 2011 (*Best Poster Winner*)
- Annual Student Research Symposium Texas Christian University April 2010

Awards

- Shale Barnett Scholarship 2012
- Graduate Student Teaching Award 2011-2012
- Biochemistry Research with Dr. Youngha Ryu 2009-present
- DFW Honors Outstanding Chemistry Seniors 2009
- Dean's Honor List 2008

Publications

- Budhathoki, P., Bernal-Perez, L. F., Annunziata, O., and Ryu, Y. (2012) Rationally-designed fluorescent lysine riboswitch probes. Org. Biomol. Chem. 10, 7872−7874
- Reversible modification of the N-terminal cysteine residue of proteins in vitro using pyruvic acid analogs (Manuscript)

Abstract

SMALL MOLECULES FOR RIBOSWITCH DETECTION AND N-TERMINAL PROTEIN MODIFICATION

By

Pradeep Budhathoki, Ph.D. 2015

Department of Chemistry

Texas Christian University

Dissertation Advisor: Dr. Youngha Ryu

Riboswitches are metabolite sensing mRNA molecules that regulate gene expression in bacteria, archea, fungi, and plants. These recently characterized molecules specifically bind to an associated metabolite, inducing conformational changes that ultimately lead to negative or positive regulation of the genes associated with the biosynthesis or transport of the bound metabolite**.** To date, more than 20 distinct classes of riboswitches have been found for different cellular metabolites including various amino acids, sugars, coenzymes, nucleotide bases, and metal ions. In line probing, transcription termination assay and isothermal titration calorimetry have been used to study riboswitches (Chapter 1). However, due to its distinct advantages, fluorescence based approach is useful for probing riboswitches. Based on the three dimensional structure of metabolite-bound riboswitch, fluorescent probes were designed, synthesized and evaluated (chapter 2). Two fluorescent lysine amide analogs, in which the carboxyl

end of lysine was covalently attached to dansyl or NBD groups through an ethylene glycol-based linker showed high binding affinity to the lysine riboswitch *in vitro*.

Protein modification provides an important tool for studying protein structure and function. Due to its low abundance and nucleophilicity, cysteine is useful for selective and site-specific protein modifications (Chapter 3). The N-terminal cysteine residue of proteins can selectively react with pyruvate analogs at pH 7 to form their corresponding 2-methyl-2,4-thiazolidinedicarboxylic acid derivatives (chapter 4). Aminoxy-containing reagents such as methoxylamine can easily reverse this reaction to regenerate the free cysteine residue. Several pyruvate analogs were synthesized with affinity (biotin) and fluorescent (pyrene) labels, which are useful for protein purification and detection, respectively. Also synthesized were pyruvate analogs with unique chemical handles such as azide and propargyl groups. These functional groups can be used to further modify proteins using bioorthogonal reactions without interfering with native biochemical reactions. All of the prepared pyruvate analogs were successfully tested for the selective and efficient N-terminal modification of the Z-domain as a model protein. The modification was easily removed to regenerate the intact unmodified proteins.