

STUDIES ON PROTEIN N-TERMINAL ACETYLATION IN BACTERIA

by

Lina Fernanda Bernal-Perez

Bachelor of Science, 2005
Texas Wesleyan 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

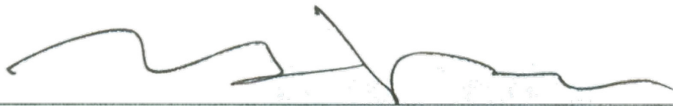
December 2012

STUDIES ON PROTEIN N-TERMINAL ACETYLTATION IN BACTERIA

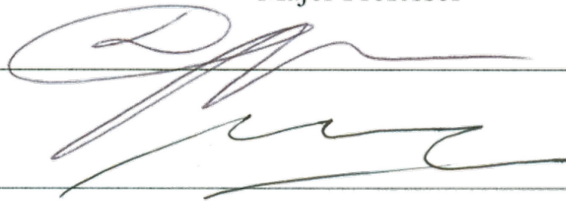
by

Lina Fernanda Bernal-Perez

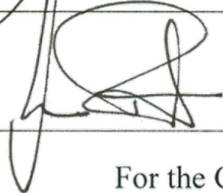
Dissertation approved



Major Professor



Jindhan Akkarajin



For the College of Science and Engineering

Copyright by
Lina Fernanda Bernal-Perez
2012

..... *to my family who is always by my side*

ACKNOWLEDGEMENTS

Undoubtedly, this is one of the most important parts of my dissertation, because I have the opportunity to thank many people who have helped me through my way on pursuing a doctoral degree. I would like to express my deepest gratitude to my advisor, Dr. Youngha Ryu, who has offered me an invaluable amount of knowledge throughout these continuous five years at TCU. Without his support and understanding, this journey would not have been intellectually nurturing or worthwhile. I am truly indebted and thankful to my committee members, Drs. Jean-Luc Montchamp, Sergei V. Dzyuba, and Giridhar R. Akkaraju, for their continuous support and exceptional advices. Their commitment as great scientists and professors has given me motivation throughout my studies to become someone successful and influential like them.

I appreciate Dr. Laszlo Prokai for his help with protein studies using mass spectrometry. I would like to thank Dr. Dean Williams for his help with sequencing. I am also thankful to Drs. Jongdoo Lim and Kristian Schlick as well as Mr. Alan Enciso Barros for their help with the mass spectrometer.

I am grateful for an opportunity that allowed me to participate in research at the Air Force Research Laboratory in Dayton, Ohio for two summers alongside many great scientists. I want to give a special thanks to my labmate, Pradeep Budhathoki, for his friendship, support, and assistance in the lab. I would like to thank a past fellow group member, Aery Lee, for all her support during her time at TCU.

Very importantly, I would like to acknowledge two great people I met during my career and who I admire very much: Drs. Ricardo Rodriguez and Geoffrey Owen. Both of them have inspired me to continue my career as a scientist. I also thank all undergraduate students I had the

opportunity to work with for their help, interest, dedication, and company in the lab: Diana Tran, Trang Le and particularly Fatima Sahyouni.

Of course I deeply owe to my dearest family: my mother, Matilde Perez Vasquez, my father, Carlos Arturo Bernal Ayala and my brother, Freddy Gabriel Bernal Perez, for their encouragement, understanding and unconditional help throughout my life and for teaching me that dedication leads to great success. A special thanks is extended to my aunt, Graciela Perez McCarthy, for her support during my career.

I am grateful to Conrad James Herrera for his unconditional help. He has been inspiring me to always look ahead and expect greatness. In fact, I would not be the person I am if he had not been a part of my life.

I would also like to extend a heartfelt thanks to my dearest friends, especially Bernat Martinez Ortega, Viviana Costa, Henry Fisher, Alex Adair, Diana Urbina, Jeremy Lantz and Maria Luisa Pineda. Without them, I would not have the energy that makes me happy and keeps me motivated to pursue my dreams.

Finally, I would like to thank Texas Christian University for funding our research.

“Those who do not remember the past are condemned to repeat it.” ...George Santayana.

PUBLICATIONS INCLUDED IN THIS DISSERTATION

The text and figures in Chapter 2, in part or in full, is a reprint of the material as it appears in – 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, 1128-1130. Reproduced by permission of The Royal Society of Chemistry.

The text and figures in Chapter 4, in part or in full, is a reprint of the material as it appears in – Bernal-Perez, L.F.; Prokai, L.; Ryu, Y. "Selective N-terminal fluorescent labeling of proteins: a method to distinguish protein N-terminal acetylation." *Anal. Biochem.* **2012**, 428, 13-15.

TABLE OF CONTENTS

Acknowledgements	iv
Publications included in this dissertation.....	vi
List of figures.....	xi
List of tables.....	xix
List of schemes	xx
List of abbreviations.....	xxi

CHAPTER 1

N-TERMINAL ACETYLATION OF PROTEINS	1
1.1 Protein posttranslational modifications	1
1.2 N-terminal methionine cleavage	2
1.2.1 The methionine aminopeptidases	3
1.3 Protein N-acetylation	5
1.4 Protein N-terminal acetylation	6
1.4.1 Eukaryotic N-terminal acetyltransferases	7
1.4.2 Prokaryotic N-terminal acetyltransferases	9
1.4.3 Sequence dependence of N-terminal acetylation.....	10
1.5 Biological importance of N-terminal acetylation.....	11
1.5.1 Protein stability.....	11
1.5.2 Protein folding and activity.....	11
1.5.3 Protein targeting.....	12
1.6 N-terminal acetylation of recombinant proteins	14
1.7 Mass spectrometry (MS) analysis of N-terminal acetylation	14
1.8 Conclusions	15

CHAPTER 2

CONTEXT-DEPENDENT N ^α -ACETYLATION OF THE RECOMBINANT Z-DOMAIN PROTEIN IN <i>ESCHERICHIA COLI</i>	16
2.1 Introduction	16
2.2 Results and discussion	17
2.2.1 Identification of the putative NAT of the Z-domain by photocross-linking.....	17
2.2.2 Unacetylated Z-domain using different <i>E.coli</i> strains and plasmids	23
2.2.3 T7 RNA Polymerase does not contribute to acetylation of the Z-domain	39
2.2.4 N ^α -Acetylation of the Z-domain protein is influenced by RimJ	41
2.3 Conclusions	64
2.4 Experimental section.....	65
2.4.1 General	65
2.4.2 Construction of plasmids	65
2.4.3 Protein expression and purification	66
2.4.4 Expression of the Z-domain Bpa mutants	67
2.4.5 Photochemical cross-linking experiment	68
2.4.6 Deletion of the <i>rimJ</i> gene from the BL21(DE3) <i>E. coli</i> strain	68
2.4.7 Mass spectrometry	69
2.4.8 Theoretical mass calculations	69
2.4.9 Trypsin proteolysis and high resolution ESI-MS analysis	70

CHAPTER 3

SEQUENCE REQUIREMENTS ON THE CLEAVAGE OF INITIATING METHIONINE & THE N ^α -ACETYLATION OF THE Z DOMAIN IN ESCHERICHIA COLI	71
3.1 Introduction	71
3.2 Results and discussion	72
3.2.1 Second amino acid substitutions	96
3.2.2 Third amino acid substitutions	99
3.3 Conclusions	101
3.4 Materials and methods	102
3.4.1 General	102
3.4.2 Site-directed saturation mutagenesis of the Z-domain gene	102
3.4.3 Expression and purification of the Z-domain variants	103
3.4.4 Mass spectrometry	103
3.4.5 Theoretical mass calculations	103

CHAPTER 4

FLUORESCENCE LABELING OF PROTEINS TO DISTINGUISH N ^α -ACETYLATION ...	105
4.1 Introduction	105
4.2 Results and discussion	107
4.2.1 OPA reactions	107
4.2.2 NBD-Cl reactions	109
4.2.2.1 Studies on time-dependent changes in fluorescence of the NBD-Cl reactions	111
4.2.2.2 NBD-Cl Fluorescence as a function of different Z-domain protein concentrations	113
4.2.2.3 Generality of the NBD-Cl method	114

4.3 Conclusions	118
4.4 Experimental section	119
4.4.1 General	119
4.4.2 Reagents and chemicals	119
4.4.3 T α 1-L12 protein expression and purification	119
4.4.4 Mass spectrometry characterization of T α 1-L12.....	120
4.4.5 T α 1-L12 Theoretical mass calculations	120
4.4.6 OPA-derivatization reactions	120
4.4.7 NBD-Cl derivatization reactions	121
4.4.8 Fluorescent measurements	121
References.....	122

VITA

ABSTRACT

LIST OF FIGURES

Figure 1.1 Ribbon structures of different methionine aminopeptidases types 1 and 2 showing in red α -helices, blue β -strands, purple metal center, green subdomain insertion unique to type 2 enzymes and yellow N-terminal extensions present in the eukaryotic enzymes: a) <i>E. coli</i> MAP (PDB accession code: 2MAT). b) <i>P. furiosus</i> MAP-2a (PDB: 1XGS). c) Human MAP-2b (PDB:1BN5).....	3
Figure 1.2 Cartoon of the different types of eukaryotic NATs.....	9
Figure 1.3 Examples of N ^α -acetylation molecular functions: a) Protein stability: N-terminal acetylation creates specific degradation signals in yeast that destabilizes the protein. b) Protein folding and activity: N ^α -acetylation is required for the activation of the enzymatic complex, which involves the interaction between E3 ligase Ubc12 and the E2 ligase Dcn1 to subsequently activate neddylation of downstream substrates. c) Protein targeting: N ^α -acetylation targets the membrane with the small GTPase Arl3 in yeast. Also, N ^α -acetylation is involved in cytosolic retention by inhibiting protein translocation to the endoplasmic reticulum.	13
Figure 2.1 Silver stained SDS-PAGE analysis of the photo cross-linked Z-domain Bpa mutants in <i>E. coli</i>	19
Figure 2.2 Mass-transformed ESI-MS spectrum of the Z-domain Bpa mutant that was expressed in BL21(DE3)cells co-transformed with pET-Z(S3TAG) and pSup-BpaRS-6TRN.....	20
Figure 2.3 Mass-transformed ESI-MS spectrum of the Z-domain Bpa mutant that was expressed in BL21(DE3)cells co-transformed with pET-Z(V4TAG) and pSup-BpaRS-6TRN.....	21
Figure 2.4 Mass-transformed ESI-MS spectrum of the Z-domain Bpa mutant that was expressed in BL21(DE3) cells co-transformed with pET-Z(D5TAG) and pSup-BpaRS-6TRN.....	22

Figure 2.5 Mass-transformed ESI-MS spectra of the Z-domain proteins transformed with pBAD-Z(Biotin) in different <i>E. coli</i> strains: a) AVB100 and b) JW 4335.....	25
Figure 2.5 Cont. Mass-transformed ESI-MS spectra of the Z-domain proteins co-transformed with pBAD-Z(Biotin) in different <i>E. coli</i> strains: c) JW 1053 and d) JW 1423..	26
Figure 2.5 Cont. Mass-transformed ESI-MS spectra of the Z-domain proteins co-transformed with pBAD-Z(Biotin) in different <i>E. coli</i> strains: e) JW 2293 and f) JW 2294.....	27
Figure 2.5 Cont. Mass-transformed ESI-MS spectra of the Z-domain proteins transformed with pBAD-Z(Biotin) in different <i>E. coli</i> strains: g) JW 4030 and h) LCB90..	28
Figure 2.6 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in XL1-BLUE cells transformed with pBAD-Z	30
Figure 2.7 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109 cells transformed with pBADZ	31
Figure 2.8 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells transformed with pBAD-Z.....	32
Figure 2.9 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells transformed with pET-Z	33
Figure 2.10 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in AD494(DE3) cells transformed with pET-Z	34
Figure 2.11 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in K12(WT) cells transformed with pBAD-Z	35
Figure 2.12 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in B(WT) cells transformed with pBAD-Z	36
Figure 2.13 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in MG1655(Seq) cells transformed with pBAD-Z	37

Figure 2.14 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells transformed with pBAD-Z	38
Figure 2.15 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in XL1-BLUE cells co-transformed with pBAD-Z and pRep2-YC	39
Figure 2.16 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109 cells co-transformed with pBAD-Z and pRep2-YC	40
Figure 2.17 SDS-PAGE analysis of the Z-domain in <i>E. coli</i> cell lysates and the protein purified by immobilized Ni ion-affinity chromatography. Lanes 1 and 2: the Z-domain purified from BL21(DE3) <i>E. coli</i> cells co-transformed with pET-Z and pSup-JYRS-6TRN, and its corresponding cell lysate, respectively; Lanes 3 and 4: the Z-domain purified from BL21(DE3) <i>E. coli</i> cells transformed with pET-Z alone and its corresponding cell lysate, respectively; Lanes 5 and 6: the Z-domain purified from BL21(DE3) <i>E. coli</i> cells co-transformed with pBAD-Z and pACYCDuet-RimJ, and its corresponding cell lysate, respectively	42
Figure 2.18 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pET-Z and pSup-JYRS-6TRN	44
Figure 2.19 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells transformed with pET-Z	45
Figure 2.20 Mass-transformed ESI-MS spectra of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pET-Z and pACYCDuet-RimJ	46
Figure 2.21 Mass-transformed ESI-MS spectra of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pET-Z and pACYCDuet	47
Figure 2.22 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells co-transformed with pET-Z and pACYCDuet	48

Figure 2.23 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells co-transformed with pET-Z and pSup-JYRS-6TRN	49
Figure 2.24 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells co-transformed with pET-Z and pACYCDuet-RimJ	50
Figure 2.25 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells transformed with pBAD-Z	51
Figure 2.26 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells co-transformed with pBAD-Z and pACYCDuet	52
Figure 2.27 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells co-transformed with pBAD-Z and pSup-JYRS-6TRN	53
Figure 2.28 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells co-transformed with pBAD-Z and pACYCDuet-RimJ	54
Figure 2.29 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pBAD-Z and pACYCDuet-RimJ	55
Figure 2.30 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) $\Delta rimJ::kan$ cells co-transformed with pBAD-Z and pACYCDuet-RimJ.....	56
Figure 2.31 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) $\Delta rimJ::Kan$ cells co-transformed with pET-Z and pACYCDuet-RimJ	57
Figure 2.32 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) $\Delta rimJ::Kan$ cells co-transformed with pET-Z and pSup-JYRS-6TRN	59
Figure 2.33 High Resolution ESI-MS analysis of the N-terminal peptide fragments obtained by the trypsin proteolysis of the Z-domain, which was expressed in BL21(DE3) <i>E. coli</i> cells in the presence of pET-Z and pSup-JYRS-6TRN.	61

Figure 2.34 High Resolution ESI-MS analysis of the N-terminal peptide fragments obtained by the trypsin proteolysis of the Z-domain, which was expressed in BL21(DE3) <i>E. coli</i> cells in the presence of pET-Z.	62
Figure 2.35 High Resolution ESI-MS analysis of the N-terminal peptide fragments obtained by the trypsin proteolysis of the Z-domain, which was expressed in BL21(DE3) <i>E. coli</i> cells in the presence of pBAD-Z and pACYCDuet-RimJ	63
Figure 3.1 Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: a) glycine and b) alanine	74
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: c) proline and d) serine	75
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: e) threonine and f) cysteine	76
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: g) valine and h) leucine	77
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: i) isoleucine and j) methionine	78
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: k) asparagine and l) aspartic acid	79
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: m) glutamine and n) glutamic acid	80
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: o) histidine and p) lysine	81
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: q) arginine and r) phenylalanine	82

Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: s) tyrosine and t) tryptophan	83
Figure 3.2 Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: a) glycine and b) alanine	85
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: c) proline and d) serine	86
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: e) threonine and f) cysteine	87
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: g) valine and h) leucine	88
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: i) isoleucine and j) methionine	89
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: k) asparagine and l) aspartic acid	90
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: m) glutamine and n) glutamic acid	91
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: o) histidine and p) lysine	92
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: q) arginine and r) phenylalanine	93
Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: s) tyrosine and t) tryptophan	94
Figure 3.3 Distribution of different N-terminal end forms for the Z-domain variants containing 20 different amino acids: a) in the second position and b) in the third position	97

Figure 4.1 Fluorescence spectra for the OPA conjugates of the Ac-Z form and the Z form. Each protein (1.3 μ M) was incubated with 50 mM OPA in 100 mM borate buffer, pH 8.0 including 50 mM 2-ME for 5 minutes; $\lambda_{\text{ex}} = 340$ nm.	108
Figure 4.2 Fluorescence spectra for the OPA conjugates of the Ac-Z form and the Z form. Each protein (1.3 μ M) was incubated with 50 mM OPA in 50 mM sodium citrate buffer, pH 7.0 including 50 mM 2-ME for 5 minutes; $\lambda_{\text{ex}} = 340$ nm.	109
Figure 4.3 Fluorescence spectra for the NBD-Cl conjugates of the Ac-Z form and the Z form. Each protein (6 μ M) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA, pH 7.0 for 7 hours.....	111
Figure 4.4 Time-dependent NBD-Cl fluorescence intensity changes of the Ac-Z form and the Z form. Each protein (6 μ M) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA at pH 7.0 the indicated times. At each time, an aliquot was removed from the reaction mixture and its fluorescence intensity was measured at 535nm. Results are the average of three independent trials per point.....	112
Figure 4.5 Effects of the unacetylated Z-domain concentrations on the NBD-Cl fluorescence emission intensity. The indicated concentrations of protein were incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1 mM EDTA at pH 7.0 for 5.5 h. The fluorescence emission intensity was recorded at 535 nm ($\lambda_{\text{ex}} = 465$ nm). Results are the average of three independent trials per data point.....	113
Figure 4.6 Mass-transformed ESI-MS spectra of the thymosin proteins that were expressed in: a) JM109(DE3) cells transformed with pET26-T α 1-L12. b) BL21(DE3) cells co-transformed with pET26-T α 1-L12 and pACYCDuet-RimJ	115

Figure 4.7 Fluorescence spectra for the NBD-Cl conjugates of the N^α-acetylated thymosin α1-L12 fusion protein (The Ac-T form) and its unacetylated form (the T form). Each protein (6 μM) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA, pH 7.0 for 5 hours..... 116

Figure 4.8 Time-dependent NBD-Cl fluorescence intensity changes of the Ac-T form and the T form. Each protein (6 μM) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA at pH 7.0 the indicated times. At each time, an aliquot was removed from the reaction mixture and its fluorescence intensity was measured at 535nm. Results are the average of three independent trials per point..... 117

LIST OF TABLES

Table 1.1 The composition and substrate specificity of the N-acetyltransferases.....	8
Table 2.1 Proteomic Analysis of the photo cross-linked products of Z-domain	17
Table 2.2 ESI-MS analysis of unacetylated Z-domain proteins expressed from different expression plasmids and <i>E. coli</i> strains	24
Table 2.3 ESI-MS analysis of the Z-domain proteins expressed from different expression plasmids and wild type <i>E. coli</i> strains	43
Table 2.4 <i>E.coli</i> strains and plasmids used for the expression of the Z-domain proteins	67
Table 2.5 Molecular weight difference between the wild type amino acid residue and Bpa	70
Table 3.1. Observed masses and their relative intensities of the different N-terminal end forms for each Z-domain variant differing by the second amino acid residue	84
Table 3.2. Observed masses and their relative intensities of the different N-terminal end forms for each Z-domain variant differing by the third amino acid residue	95
Table 3.3 The Z-domain site directed-mutagenesis oligonucleotide primers	102

LIST OF SCHEMES

Scheme 1.1 Two categories of protein posttranslational modifications: a) Covalent modification and b) cleavage of protein backbone.....	2
Scheme 1.2 Acetyl CoA as acetyl donor for protein N-acetylation. a) Acetylation of the N-terminal amino group of proteins. b) Acetylation of the ϵ -NH ₂ of lysine residues.....	5
Scheme 1.3 a) N-terminal acetylation of eukaryotic proteins can proceed after methionine removal by the action of MAPs or intact Met residue. b) N-terminal acetylation of prokaryotic proteins occur after deformylation and hydrolytic removal of the N-terminal methionine.....	7
Scheme 2.1 Strategy to identify putative N-acetyltransferase of the Z-domain protein by photochemical cross-linking	18
Scheme 4.1 OPA and SH selectively react with primary amines to yield fluorescent adducts.	106
Scheme 4.2 Isoindole degradation a) obtained from 2-mercaptoethanol and b) autoxidation....	106
Scheme 4.3 NBD-Cl reacts with primary amines to yield fluorescent adducts	106
Scheme 4.4 The reaction of NBD-Cl in citrate buffer, pH 7: the unacetylated Z-domain (The Z form), b) N ^{α} -acetylated Z-domain (The Ac-Z form).....	110

LIST OF ABBREVIATIONS

2-ME = 2-mercaptoethanol

Ack = acetate kinase

Acs = acetyl-CoA synthetase

Ac-Met-Z form = N^α-acetylated Z-domain with the Met₁ residue

Ac-T form = N^α-acetylated form of Tα1-L12

Ac-Z form = N^α-acetylated Z-domain without Met₁

ARFRP1 = ARF-related protein

Bpa = p-benzoyl-L-phenylalanine

CAT = chloramphenicol acetyltransferase

CPY = carboxypeptidase Y

EDTA = ethylenediaminetetraacetic acid

EF-Tu = elongation factor Tu

ER = endoplasmic reticulum

ESI = electrospray ionization

fMet₁ = N-formylmethionine

FTICR = fourier-transform ion cyclotron resonance

GNAT = GCN5 N-acetyltransferase

GNMT = glycine N-methyltransferase

IMAC = immobilized metal ion affinity chromatography

m/z = mass-to-charge ratios

MALDI = matrix-assisted laser desorption/ionization

MAP = methionine aminopeptidase

Met₁ = N-terminal methionine

Met-Z form = Z-domain with the Met₁ residue

MS = mass spectrometry

MS/MS = tandem mass spectrometry

α -MSH = α -melanocyte stimulating hormone

NAT = N-acetyltransferases

NBD-Cl = 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole

N-TIMP-1 = mammalian tissue inhibitors of matrix metalloproteinases

N ^{α} = N-terminal

OPA = O-phthalaldehyde

P2 = penultimate position

P3 = antepenultimate position

PCR = polymerase chain reaction

Pta = phosphotransacetylase

PTM = co- and post- translational modifications of proteins

SDS-PAGE = sodium dodecyl sulfate polyacrylamide gel electrophoresis

SLDs = stathmin-like domains

T form = unacetylated form of T α 1-L12 without the initiating fMet₁ residue

TIMPs = tissue inhibitors of metalloproteinases

T α = thymosin α

T α 1-L12 = thymosin α 1-L12 fusion

Ubc12 = E2 ubiquitin-conjugating enzyme

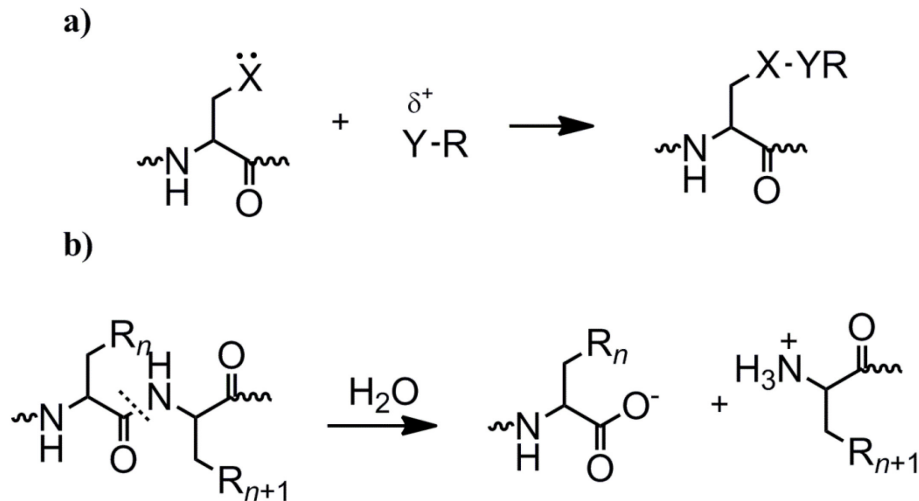
Z form = the unacetylated Z-domain without Met₁

CHAPTER 1. N-TERMINAL ACETYLATION OF PROTEINS.

1.1 Protein posttranslational modifications

Proteome, the inventory of all proteins in an organism or cell, encompass beyond that of the genome since proteomic diversity is further generated either by mRNA splicing or by covalent modification of proteins at one or more sites.¹ This dissertation emphasizes the latter modification, co- and post- translational modifications of proteins (collectively abbreviated as PTM). After the information encoded in DNA is translated into proteins via RNA, proteins can be further altered by a myriad of chemical modifications, which affect their folding, stability, activity, intracellular localization, antigenicity, and interactions with other cellular molecules such as nucleic acids, cofactors, lipids and other proteins. Over 300 protein modifications have been identified that add up to the array of approximately greater than 1,000,000 distinct protein forms, more than 2-3 orders of magnitude more complex than the genome would anticipate.²

PTMs are often divided into two broad categories as shown in Scheme 1.1. The first group includes all enzyme-catalyzed covalent additions of different moieties such as phosphoryl, acetyl, methyl or glycosyl groups into a particular protein. In the second group of PTM, the protein substrate is specifically cleaved from either the N or C termini by the action of proteases. The two major types of protein modifications that will be discussed in this dissertation include methionine cleavage and N-acetylation.



Scheme 1.1 Two categories of protein posttranslational modifications: a) Covalent modification and b) cleavage of protein backbone.¹

1.2 N-terminal methionine cleavage

Protein synthesis is initiated with methionine in the eukaryotic cytoplasm or N-formylmethionine (fMet₁) in prokaryotes, mitochondria and chloroplasts. The N-formyl group can be removed by peptide deformylase leaving the protein with an N-terminal methionine (Met₁) residue. Subsequently, the Met₁ residue is removed for some proteins.³ The enzyme involved in this process is methionine aminopeptidase (MAP).⁴ The Met₁ residue is cleaved co-translationally before the polypeptide chains are completely synthesized.⁵ The Met₁ cleavage is conserved from bacteria to eukaryotes. The rules dictating methionine retention, or removal, are similar in all organisms.⁶ Met₁ cleavage is necessary for the activity, stability, degradation, maturity, and subcellular localization of diverse proteins.^{3,7} The efficiency of Met₁ cleavage depends on the N-terminal peptide sequence and the specificity of MAP.⁵

1.2.1 The methionine aminopeptidases

MAPs are dinuclear metalloproteases that catalyze the hydrolytic cleavage of the N-terminal methionine from some proteins. The crystallographic analysis of different MAP structures has revealed their “pita-bread” fold containing a metal center and a well-defined substrate pocket as shown in Figure 1.1.⁸⁻¹⁰ There are two types of MAPs: MAP1 and MAP2, which are found in eubacteria and archaea, respectively. Eukaryotes contain both types. Unlike MAP1, MAP2 has an insertion of about 60 amino acids long α helical domain within its catalytic region. This insertion sequence shares no sequence homology with any other known protein.¹¹ MAPs are further divided into two subclasses: a or b, depending on the absence or presence of N-terminal extension, respectively. The N-terminal extension in the type 1b contains two zinc finger motifs, while that in the type 2b has alternate stretches of polyacidic and polybasic residues.¹² The recently discovered type 1c and 2c subclasses have 40 amino acids long N-terminal extensions but no zinc finger motif is found.^{13,14}

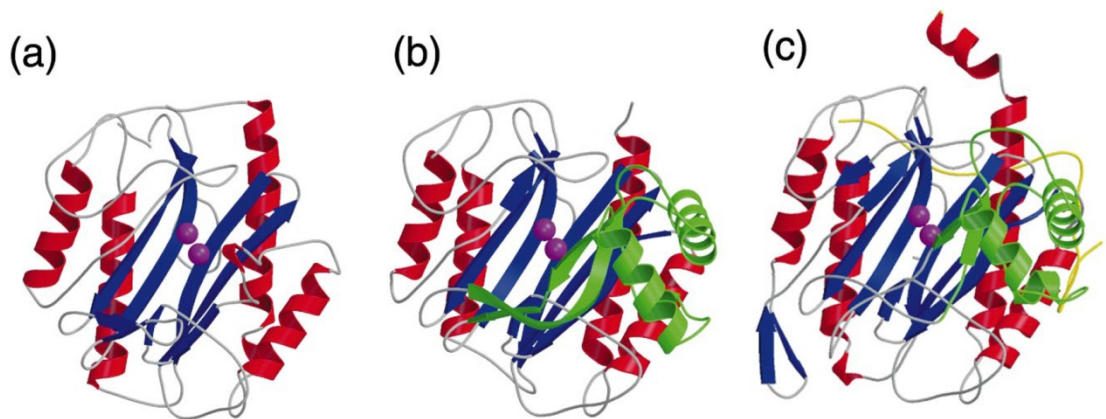


Figure 1.1 Ribbon structures of different methionine aminopeptidases types 1 and 2 showing in red α -helices, blue β -strands, purple metal center, green subdomain insertion unique to type 2 enzymes and yellow N-terminal extensions present in the eukaryotic enzymes: a) *E. coli* MAP (PDB accession code: 2MAT). b) *P. furiosus* MAP-2a (PDB: 1XGS). c) Human MAP-2b (PDB:1BN5).⁴ (Reproduced with permission of Elsevier.)

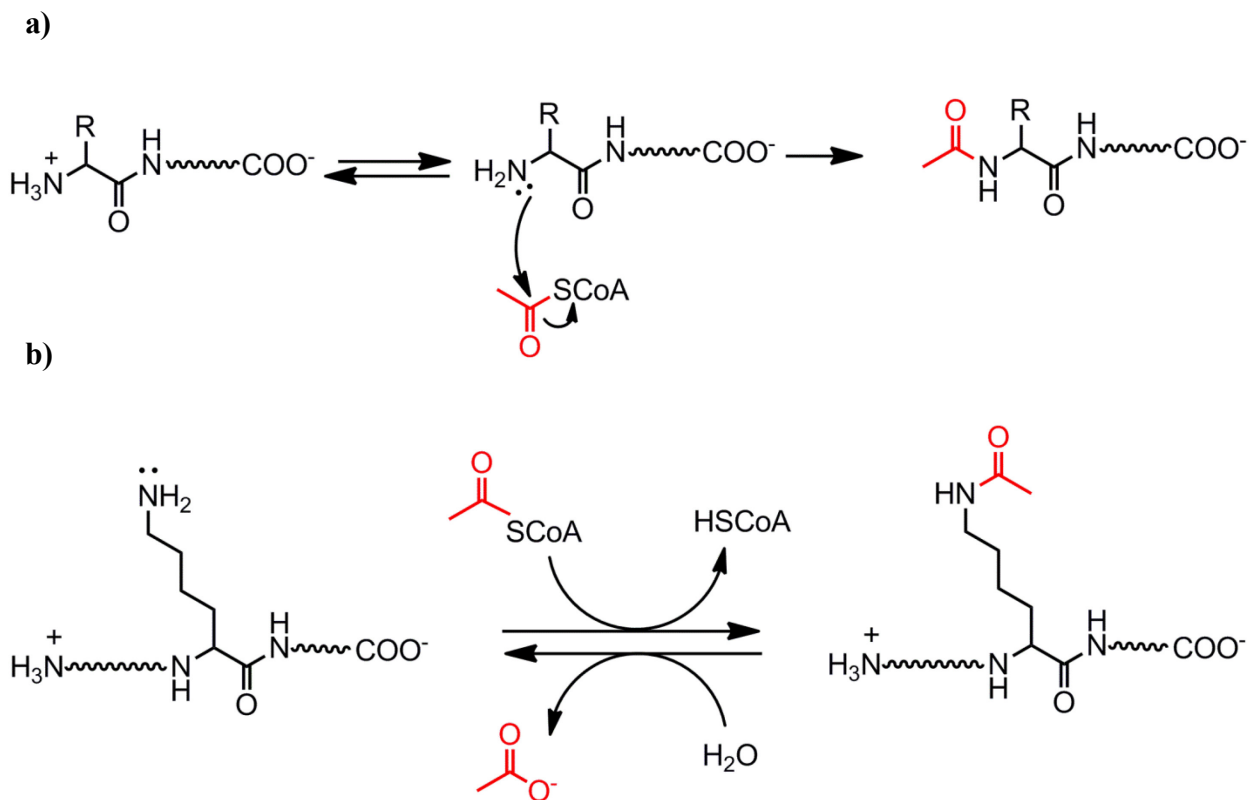
MAPs are activated *in vitro* by different divalent metal ions, such as Co^{2+} , Ni^{2+} , Zn^{2+} , Mn^{2+} , and Fe^{2+} , but it is not yet clear which metal ions are present in physiological conditions.¹⁵⁻¹⁸ Early studies on the crystal structures of MAP1 and MAP2 revealed the presence of two cobalt cations per polypeptide chain.^{8,9} However, it is now thought that prokaryotic MAPs are mononuclear iron enzymes.¹⁹ On the other hand, human MAP2 produced in *E. coli* appears to be a manganese enzyme.²⁰ For this reason, the attached metal ion may vary. A mixture of metal ions may exist according to their availability at specific physiological conditions, or the identity of metal ion may depend on the MAP concentration.²¹

The removal of Met_1 is an essential process. For example, in *E. coli* and *Salmonella typhimurium* the deletion of the MAP genes is lethal.^{22,23} In yeast, the deletion of MAP1 or MAP2 results in a slow-growth phenotype and the disruption of both genes is fatal.^{24,25}

E. coli and *S. typhimurium* MAPs are unable to act on N-formylmethionyl termini, neither *in vitro* nor *in vivo*.²⁶ The MAP-catalyzed Met_1 removal is sequence-specific. In *E. coli*, proteins undergo complete Met_1 cleavage when small amino acid residues such as Ala, Cys, Gly, Pro and Ser occupy the penultimate position (P2) and any amino acid (except proline) in the antepenultimate position (P3). On the other hand, Met_1 is retained when the P2 residue is any amino acid other than Ala, Cys, Gly, Pro, Ser, Thr, and Val and the P3 residue is Pro.¹¹ Furthermore, the Met_1 excision varies if the P2 residue is Thr or Val, and it also depends on the identity of the P3-P6 residues as well as protein expression and cell culture conditions.¹¹

1.3 Protein N-acetylation

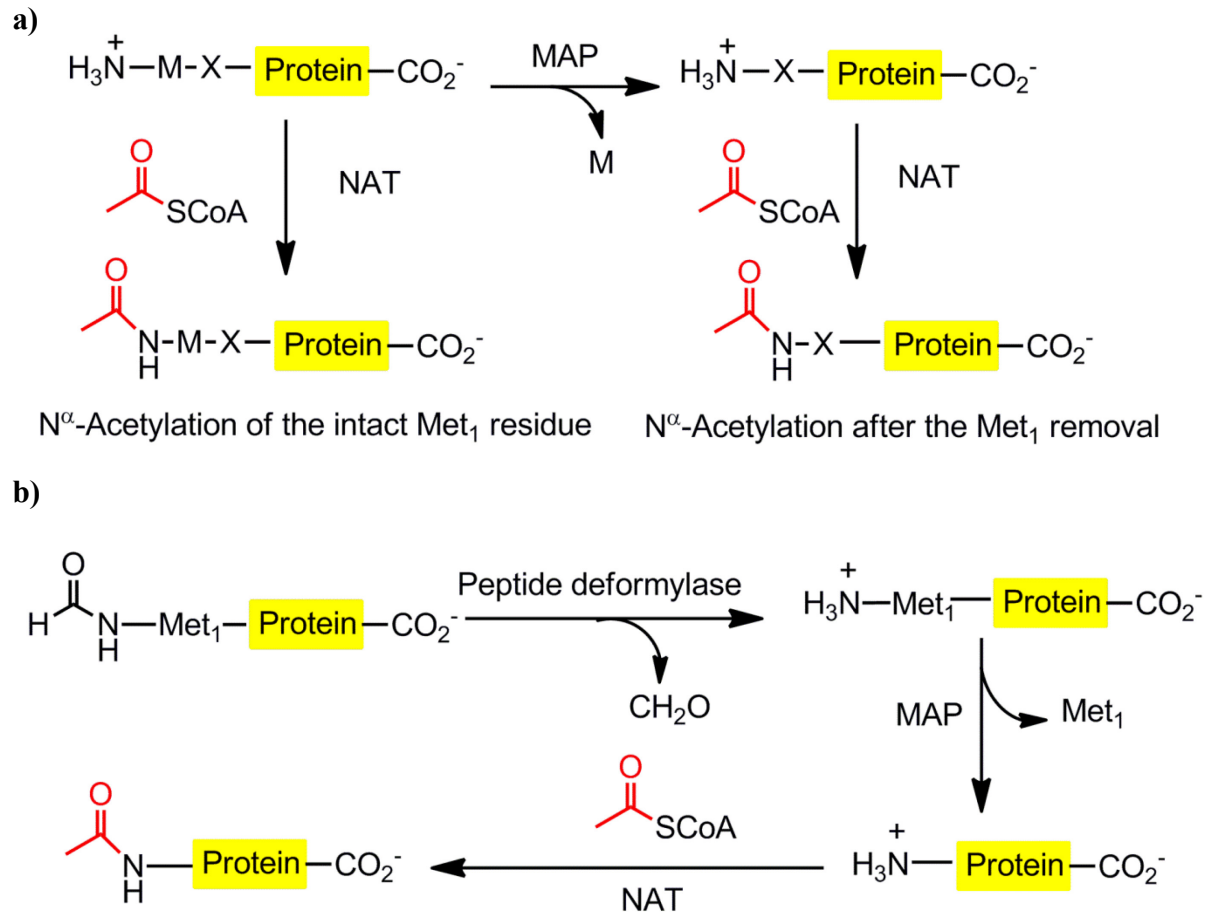
Protein N-acetylation is one of the 300 reported covalent post-translational modifications, which adds complexity to the proteome by producing a large number of distinct protein isoforms.²⁷ Catalyzed by a range of acetyltransferases, protein N-acetylation involves the transfer of acetyl groups from acetyl-coenzyme A to the α - amino group of the N-terminal residue or to the ϵ - amino group of lysine residues (Scheme 1.2).² The N-terminal (N^α) acetylation is irreversible, occurring post-translationally or co-translationally after the polypeptide chain has grown with around 20-50 residues long. The lysine acetylation is reversible and takes place post-translationally.² This dissertation is focused on the N-terminal acetylation of proteins.



Scheme 1.2 Acetyl CoA as acetyl donor for protein N-acetylation. a) Acetylation of the N-terminal amino group of proteins. b) Acetylation of the ϵ -NH₂ of lysine residues.²

1.4 Protein N-terminal acetylation

Protein N^α-acetylation occurs in all three domains of life. This modification is very frequent in eukaryotes but significantly less frequent in prokaryotes.²⁸ Recent studies have demonstrated that N^α-acetylation is more prevalent in archaeal and bacterial organisms than previously thought.^{32,33,29} N^α-acetylation occurs in approximately 50-70% of proteins in *Saccharomyces cerevisiae*,^{30,31} 70-90% in human cells,^{5,32} 70-75% in *Arabidopsis thaliana*,^{33,34} 14-19% in haloarchaea³⁵ and 16% in bacteria.³⁴ In eukaryotes, N^α-acetylation generally occurs co-translationally on the Met₁ residue or on the N-terminal residue after the cleavage of Met₁ by MAPs (Scheme 1.3a). One exception was reported in *A. thaliana*, in which protein N^α-acetylation occurs post-translationally in the chloroplast.³⁴ In contrast, prokaryotic proteins are N^α-acetylated post-translationally on the N-terminal residue after deformylation and consequent removal of Met₁ (Scheme 1.3b).



Scheme 1.3 a) N-terminal acetylation of eukaryotic proteins can proceed after methionine removal by the action of MAPs or intact Met residue. b) N-terminal acetylation of prokaryotic proteins occur after deformylation and hydrolytic removal of the N-terminal methionine.²

1.4.1. Eukaryotic N-terminal acetyltransferases

Protein N^α -acetylation in eukaryotes is catalyzed by various NAT complexes, which are composed of different subunits and have diverse substrate specificities (Figure 1.2, Table 1.1). Five conserved NATs have been identified from yeast to humans: NatA-NatE. Higher eukaryotes also contain NatF, which is probably responsible for higher level of N^α -acetylation in humans than yeast. Only after the Met₁ removal by MAPs, NatA acetylates N-terminal Ser, Thr, Ala, Gly, Cys or Val residue.³⁰ NatB and NatC complexes catalyze the N^α -acetylation of the Met₁

residue, which is influenced by the identity of the penultimate residue. NatB acetylates Met₁ followed by an acidic or Asn residue.³⁶ NatC acetylates Met₁ followed by a large hydrophobic residue such as Ile, Leu, Tyr or Phe.³⁷ In yeast, NatD acetylates the N-terminal Ser residue of histones H2 and H4.³⁸ NatD recognizes the first 23-51 residues for its effective acetylation whereas the other NATs distinguish the first 2-5 residues of the substrate.³⁸ NatE acetylates the Met₁ residues followed by Leu, Ala, Lys, or Met *in vitro*.³⁹ The recently discovered NatF acetylates Met₁ followed by Lys, Leu, Ile, Trp or Phe.³¹

NATs are essential for normal cell functions. In humans, the deletion of different subunits of any NAT complexes results in cell death or defects in cell division.^{37,40} According to a recent study on male infant lethality, a mutation of the NatA catalytic subunit (S37P variant) impaired N-terminal acetylation, causing a human genetic disorder.⁴¹ In yeast, mutations in the Nat1 and Ard1 catalytic subunits of NatA resulted in slow growth and failure to acetylate the same subset of proteins that are normally acetylated.⁴²

Table 1.1 The composition and substrate specificity of the N-acetyltransferases.⁴³

Nat	Subunits	Substrates
NatA	Naa10p(Ard1p) Naa15p (Nat1p)	Ser, Ala, Gly, Thr, Val, Cys
NatB	Naa20p(Nat3p) Naa25p (Mdm20p)	Met-Glu-, Met-Asp, Met-Asn-, Met-Gln-
NatC	Naa30p (Mak3p) Naa35p (Mak10p) Naa38p (Mak31p)	Met-Leu-, Met-Ile-, Met-Trp-, Met-Phe-
NatD	Naa40p (Nat4p)	Ser-Gly-Gly-, Ser-Gly-Arg-
NatE	Naa50p (Nat5p)	Met-Leu-, Met-Ala-, Met-Lys-, Met-Met-
NatF	Naa60p	Met-Lys-, Met-Leu-, Met-Ile-, Met-Trp-, Met-Phe-

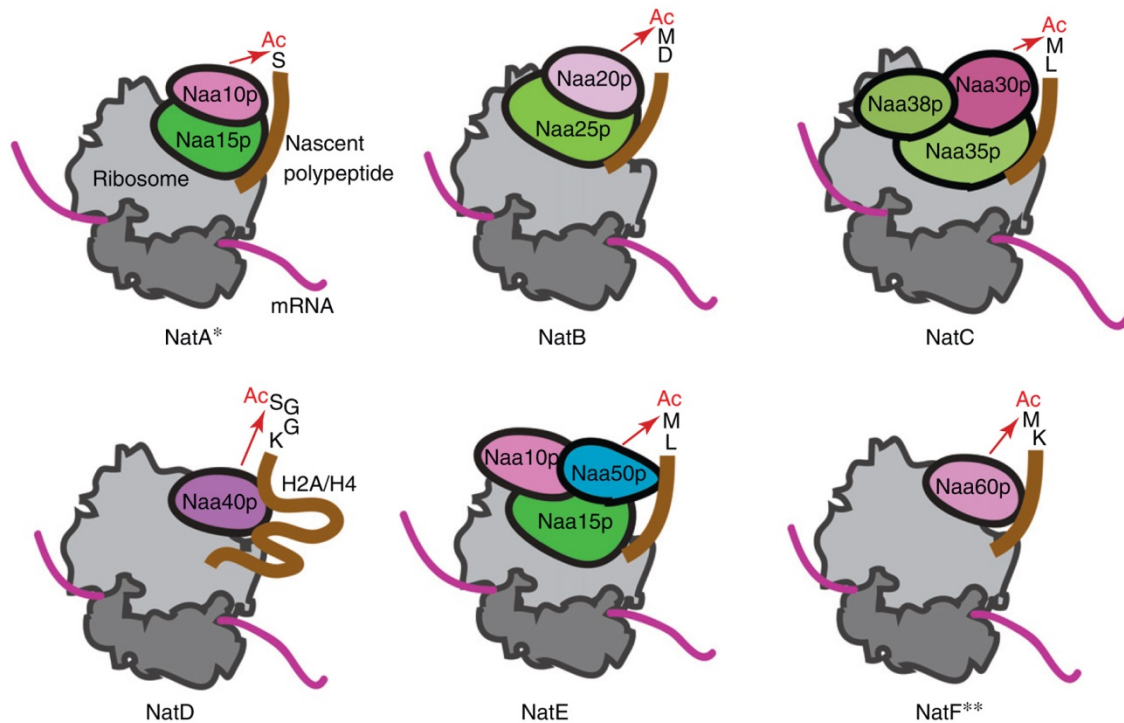


Figure 1.2 Cartoon of the different types of eukaryotic NATs.⁴⁴ (Reproduced with permission of Elsevier.)

1.4.2 Prokaryotic N-terminal acetyltransferases

In *E. coli*, five N^α-acetylated proteins have been identified: elongation factor EF-Tu,⁴⁵ chaperone secB,⁴⁶ and the three ribosomal proteins S5, S18 and L7/L12.^{47,48} EF-Tu promotes the binding of aminoacyl-tRNA to ribosomes during protein biosynthesis and it has an N-terminal serine residue.^{45,49} SecB is involved in the protein translocation pathway. This protein has the N-terminal serine residue followed by glutamate. It was suggested that the N^α acetyl moiety in SecB stabilizes the tetramer form, resulting in a more compact particle.^{46,50} RimJ, RimI and RimL are the three known NATs that acetylate the ribosomal proteins S5, S18 and L7/L12 in *E.*

coli with the Ala-His-, Ala-Arg-, and Ser-Ile- termini, respectively.^{47,48} The NATs of the EF-Tu and Sec B have not been reported.

The prokaryotic Rim NATs and eukaryotic NATs belong to the GCN5 N-acetyltransferase (GNAT) family, which acetylate a diverse group of substrates with primary amines.⁵¹ One characteristic of GNAT proteins is their low sequence homology between family members with different target substrates.⁵² Unlike eukaryotic NATs that consist of two or more different subunits, prokaryotic NATs are homodimeric.

1.4.3 Sequence dependence of N-terminal acetylation

A number of studies on protein N^α-acetylation in different organisms have attempted to relate the N^α-acetylation patterns based on the frequency of the amino acids found at the protein N-termini. For example, in yeast the removal of Met₁ by MAPs frequently resulted in N-terminal acetylation of Ala, Val, Ser, Thr and Cys residues.⁵³ A study on human and archaeal proteins indicated that the N-terminal Ala and Ser residues are generally acetylated after the Met₁ cleavage.^{32,35} In *Drosophila melanogaster* NATs prefer the N-terminal Met, Ala, and Ser residues.³² A genomic survey of 45 different bacterial species suggested that N-terminal acetylation commonly occurred at the N-terminal Ser, Thr, or Ala residues of the protein after Met₁ cleavage.²⁹

The penultimate amino acid residues also have a profound influence on the N^α-acetylation of proteins. For example, in yeast, aspartate or glutamate in the penultimate position stimulated acetylation, whereas proline or lysine inhibited acetylation.⁵ Nevertheless, it is unclear whether the sequence specificity of N^α-acetylation is due to the selectivity of specific NATs.

1.5 Biological importance of N-terminal acetylation

Recently, many different roles of N-acetylation were reported in the literature (Figure 1.3).

1.5.1 Protein Stability

N^α-acetylation is involved in protecting proteins from degradation. In 1984, it was reported that the N^α-acetylated proteins are much more stable than non-acetylated proteins *in vivo*.⁵⁴ It later became evident that ubiquitination, the attachment of ubiquitin at the N-terminal amino acid residue and one of the initial events for protein degradation, requires the free N-terminal amino group. The cell-cycle regulator p21, the tumor suppressor p16^{INK4α} and the human papillomavirus oncoprotein-58 E7 showed ubiquitination at their N-terminal domain and subsequent degradation.^{55,56} Therefore, it was suggested that N^α-acetylation blocks the access of ubiquitin, thereby stabilizing the protein.⁵⁷

In contrast to the previous findings that N^α-acetylation protects proteins from degradation, a recent study showed that N^α-acetylation creates specific degradation signals for some yeast proteins.⁵³ The Doa10 ubiquitin ligase targets the N^α-acetylated Met, Ala, Val, Ser, Thr or Cys residues as degradation signals.

1.5.2 Protein folding and activity

The presence or absence of the N^α-acetyl group is critical for the activity of some biologically important proteins. N^α-acetylation of tropomyosin strengthens its interaction with actin filaments, which in turn enhances actin binding and its ability to regulate distinct classes of

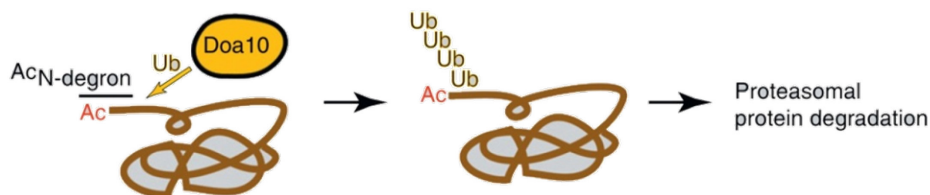
myosin.⁵⁸ The N^α-acetylated form of α-melanocyte stimulating hormone (α-MSH), a melanocortin 4 receptor, has higher potency in reducing energy intake than the unacetylated form.⁵⁹ Carboxypeptidase Y (CPY) inhibitor forms a tight complex with CPY only when N^α-acetylated.⁶⁰ Scott *et al.* reported that N^α-acetylation is necessary for the ubiquitin-like protein conjugations between E2 ubiquitin-conjugating enzyme (Ubc12) and Dcn1 E3ubiquitin ligase.⁶¹ A hemoglobin variant in which a valine residue is substituted with alanine at position 2 was N^α-acetylated and showed a decreased affinity for oxygen.^{62,63}

The N^α-acetylated form of N-TIMP-1, one of the mammalian tissue inhibitors of matrix metalloproteinases is inactive whereas the unacetylated form is active.⁶⁴ Contrary to the N^α-acetylated native protein that exhibited the sigmoidal kinetics, the unacetylated rat glycine N-methyltransferase (GNMT) showed hyperbolic kinetics at low pH.⁶⁵

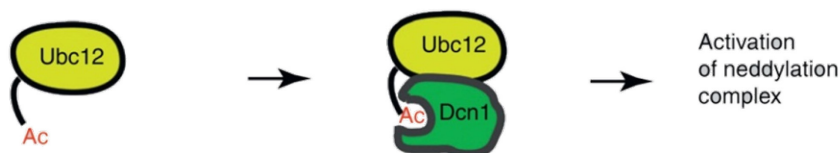
1.5.3 Protein Targeting

N^α-acetylation is important for the protein-membrane and protein-protein interactions that are involved in protein targeting. For example, N^α-acetylation catalyzed by NatC, targets yeast Arl3p and its mammalian orthologue, ARF-related protein (ARFRP1), to membranes by facilitating recognition of the receptor Sys1p/hSys1.⁶⁶⁻⁶⁸ Forte and colleagues demonstrated that N^α-acetylation leads to cytosolic retention of secretory proteins. The protein N^α-acetylation was also linked to the protein's ability to get translocated to the endoplasmic reticulum (ER). When a normally secreted protein was altered from an unacetylated to a N^α-acetylated form, it was inhibited from being translocated to the ER.⁶⁹

a) Protein stability

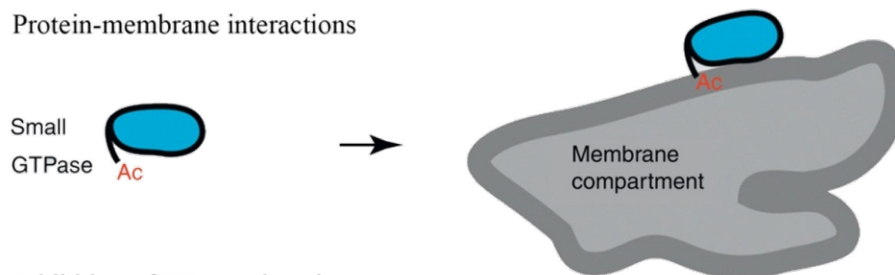


b) Protein folding and activity



c) Protein targeting

Protein-membrane interactions



Inhibition of ER-translocation

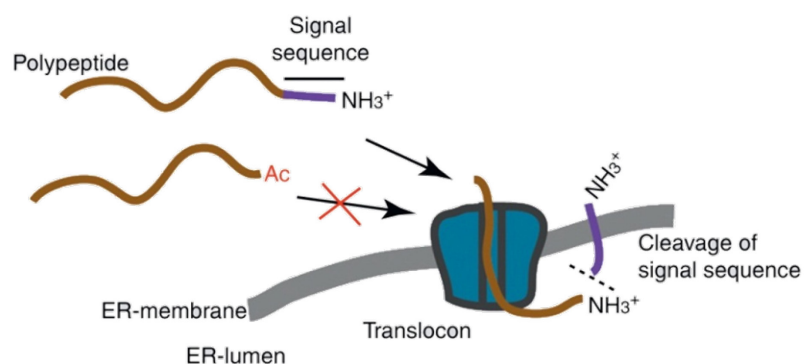


Figure 1.3 Examples of N^α-acetylation molecular functions: a) Protein stability: N-terminal acetylation creates specific degradation signals in yeast that destabilizes the protein.⁵³ b) Protein folding and activity: N^α-acetylation is required for the activation of the enzymatic complex, which involves the interaction between E3 ligase Ubc12 and the E2 ligase Dcn1 to subsequently activate neddylation of downstream substrates.⁶¹ c) Protein targeting: N^α-acetylation targets the membrane with the small GTPase Arl3 in yeast.⁶⁸ Also, N^α-acetylation is involved in cytosolic retention by inhibiting protein translocation to the endoplasmic reticulum.^{44,69}

1.6 N-terminal acetylation of recombinant proteins

Most eukaryotic proteins are not acetylated when ectopically expressed in *E. coli*. However, partial or complete N^α-acetylation were reported for several recombinant proteins, including the stathmin-like domains (SLDs), interferons, thymosin α (T α), and tissue inhibitors of metalloproteinases (TIMPs).⁷⁰⁻⁷⁴ Although the NATs for most other N^α-acetylated recombinant proteins remain unidentified, it was reported that RimJ catalyzes N^α-acetylation of recombinant T α fusion proteins in *E. coli*.⁷⁵ For most other proteins, however, the underlying mechanism of N^α-acetylation remains unknown.

1.7 Mass spectrometry (MS) analysis of N-terminal acetylation

Mass spectrometry (MS) is one of the most powerful and widely used tools in proteomics. This technique has made great contributions to the protein identification, quantification, subcellular localization, and posttranslational modifications.^{76,77} The molecular weights and fragmentation patterns of peptides derived from proteins can be obtained using MS.⁷⁸ Matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) are the two primary ionization methods that revolutionized the field of MS. Advances in protein separation technologies in recent years have made proteins more amenable for MS analysis. The most popular type of MS analysis is the bottom-up approach, in which peptide fragments are first obtained by proteolytic digest and subsequently analyzed by MS. In contrast, in the top-down approach, peptide fragments are generated by gas-phase proteolysis within the mass spectrometer and simultaneously identified by MS.⁷⁹ Tandem mass spectrometry (MS/MS) is a powerful

protein identification technique, providing protein sequence information by matching protein sequence databases.⁸⁰

1.8 Conclusions

It is evident that there is a significant lack of research on the N^α-acetylation mechanism in prokaryotes. Many studies used proteome databases to understand the mechanism of N^α-acetylation in different organisms. However, the function of N^α-acetylation is not fully understood and some results are even contradictory. In this dissertation, we used the Z-domain, a small three-helix bundle protein derived from *Staphylococcal* protein A, as a model protein to study the N^α-acetylation in *E. coli*. The NAT responsible for the N^α-acetylation of the Z-domain and the specific conditions to promote N^α-acetylation are described in Chapter 2. The sequence dependence on the N^α-acetylation of the Z-domain protein is discussed in Chapter 3. A simple and sensitive fluorescence-based method to distinguish the N^α-acetylation status of proteins is described in Chapter 4.

CHAPTER 2. CONTEXT-DEPENDENT N^α-ACETYLATION OF THE RECOMBINANT Z-DOMAIN PROTEIN IN *ESCHERICHIA COLI*

2.1 Introduction

In *E. coli*, three ribosomal proteins (S5, S18, and L7/L12), elongation factor EF-Tu, and chaperone SecB are acetylated at their N-termini. Several studies have previously demonstrated that three NATs, RimI, RimJ, and RimL modify the proteins S18, S5 and L7/L12, respectively.^{47,51} Although most eukaryotic proteins are not acetylated when expressed in *E. coli*, partial or complete N^α-acetylation has been reported for several ectopic recombinant proteins as described in Chapter 1. Recently, it has been demonstrated that RimJ is responsible for the N^α-acetylation of the thymosin α 1 (T α 1) fusion proteins in *E. coli*.⁷⁵ However, for most other proteins, the underlying mechanism of N^α-acetylation remains unknown.

In early studies, we observed that the Z-domain protein is also acetylated at its N-terminus when expressed in *E. coli* only under certain conditions. The Z-domain protein is a small three-helix bundle protein derived from *Staphylococcal* protein A, which has been extensively engineered as a non-immunoglobulin binding protein for a number of applications in biotechnology and therapeutics.^{81,82} Since extensive and site-specific, the N^α-acetylation of the Z-domain was possibly mediated by an NAT rather than a nonspecific chemical reaction.

In order to identify the NAT of the Z-domain, we initially employed a photochemical cross-linking experiment by using p-benzoyl-L-phenylalanine (Bpa), an unnatural amino acid which could form a covalent conjugate with its interacting molecules upon UV irradiation, and has proven useful for capturing transient protein-protein interactions in living cells.⁸³ In an

alternate approach, we examined the effects of different *E. coli* strains and expression plasmids on N^α-acetylation. We also investigated whether RimJ was involved in this N^α-acetylation.

2.2 Results and discussion

2.2.1 Identification of the putative NAT of the Z-domain by photocross-linking

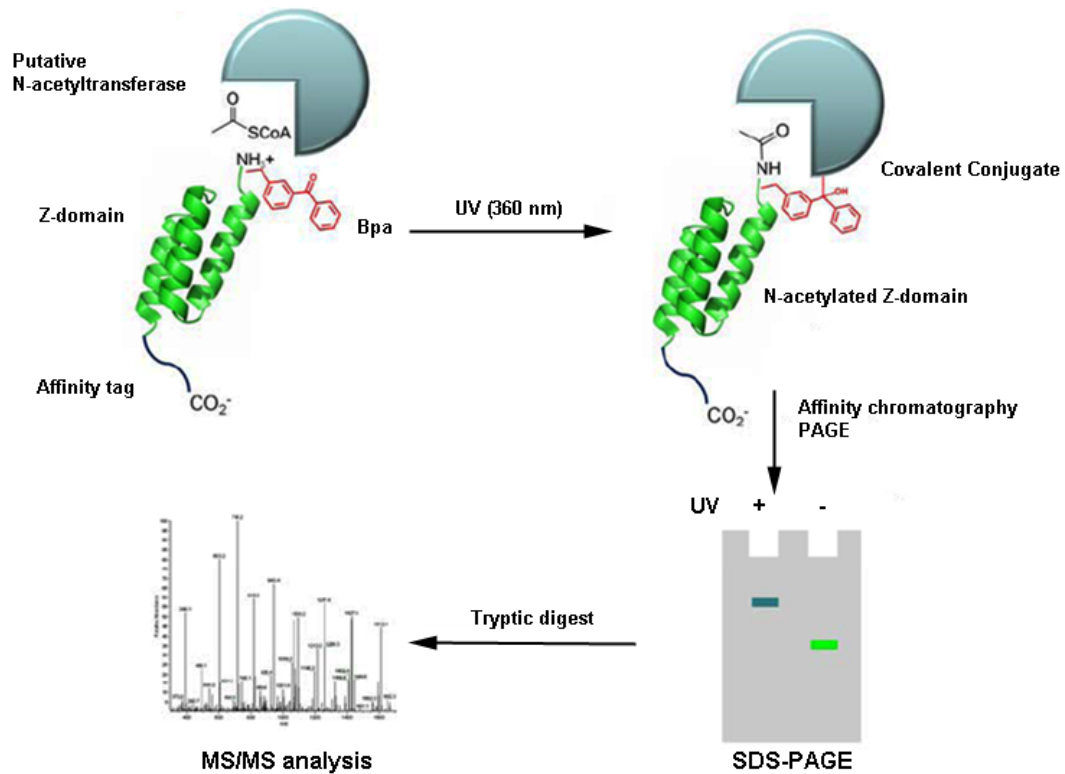
Initially, we attempted to identify the putative NAT of the Z-domain by photocross-linking in *E. coli*.⁸³ We constructed pET-Z, which encodes the wild type Z-domain protein with a C-terminal hexa-histidine tag under control of the T7 promoter. Three Z-domain mutants containing Bpa at positions 3, 4 and 5 were overexpressed in the BL21(DE3) *E. coli* cells co-transformed with pSup-BpaRS-6TRN and pET-Z mutants, in which the codons for Ser-3, Val-4, and Asp-5 of the Z-domain were changed to an amber (TAG) codon by site-directed mutagenesis. The purified Z-domain Bpa mutants were analyzed by ESI-MS and the results are summarized in Table 2.1. Bpa was successfully incorporated in all three mutants. Different N^α-acetylation patterns were observed for these Z-domain Bpa variants.

Table 2.1 ESI-MS analysis of the Z-domain mutant proteins.

Entry	Expression plasmid	Supplementary plasmid	Observed major mass peaks/Da	Major forms of the Z-domain
1	pET-Z (S3 TAG)	pSup-Bpa-6TRN	7968.3	Ac-Z ^a
2	pET-Z (V4 TAG)	pSup-Bpa-6TRN	7913.9	Z ^b
3	pET-Z (D5 TAG)	pSup-Bpa-6TRN	7898.2	Z

^a Acetylated Z-domain without Met₁ residue. ^b Unacetylated Z-domain without the Met₁.

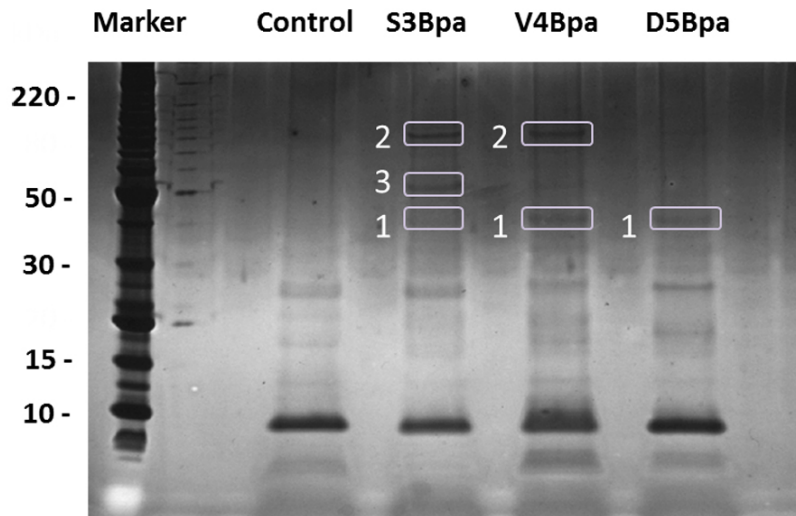
The overall strategy to identify the putative NAT of the Z-domain protein by the photochemical cross-linking is outlined in Scheme 2.1. The Bpa mutants of the Z-domain were expressed in frame with a six-histidine tag. Initially, the cell cultures were irradiated for 20, 30 or 45 minutes using a high power UV light (360nm), but no photocross-linking was detected. Upon UV irradiation for 1 hour, the Z-domain proteins covalently conjugated to the putative NAT were purified from the endogenous *E. coli* proteins by immobilized metal ion affinity chromatography (IMAC).



Scheme 2.1 Strategy to identify putative N-acetyltransferase of the Z-domain protein by photochemical cross-linking

The purified proteins were analyzed by silver-stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 2.1). Since unable to photocross-link, the Z-domain mutant containing tyrosine in position 3 was used as a negative control. A major

band around 8 kDa was visible, which corresponded to the Z-domain protein. A few protein bands were also visible on the gel as possible photocross-linked products.



Band	S3Bpa	V4Bpa	D5Bpa
1	Protein A	Protein A	Protein A EF-Ts
2	Protein A Hsp70 PBP	EF-Tu Protein F	
3	EF-Tu Protein A		

Figure 2.1 Silver stained SDS-PAGE analysis of the photo cross-linked Z-domain Bpa mutants in *E. coli*.

These discrete bands were excised and subsequently analyzed by a standard MS/MS proteomics technique. Unfortunately, the identified products were some abundant proteins such as EF-Tu and Hsp70 instead of any proteins relevant to protein acetylation. (Figure 2.1).

As shown in Figures 2.2, Z-domain Ser3Bpa mutant gave a single major peak at 7968.3 Da, which corresponds to the acetylated form of the Z-domain without the fMet₁ residue (the Ac-Z form, calculated mass 7968.7 Da). In contrast, the Val4Bpa and Asp5Bpa variants showed

their major peaks at 7913.9 Da and 7898.2 Da, respectively, both of which correspond to their unacetylated form of the Z-domain without the fMet₁ residue (the Z forms calculated masses 7914.5 Da and 7898.6 Da, respectively) (Figures 2.3 and 2.4).

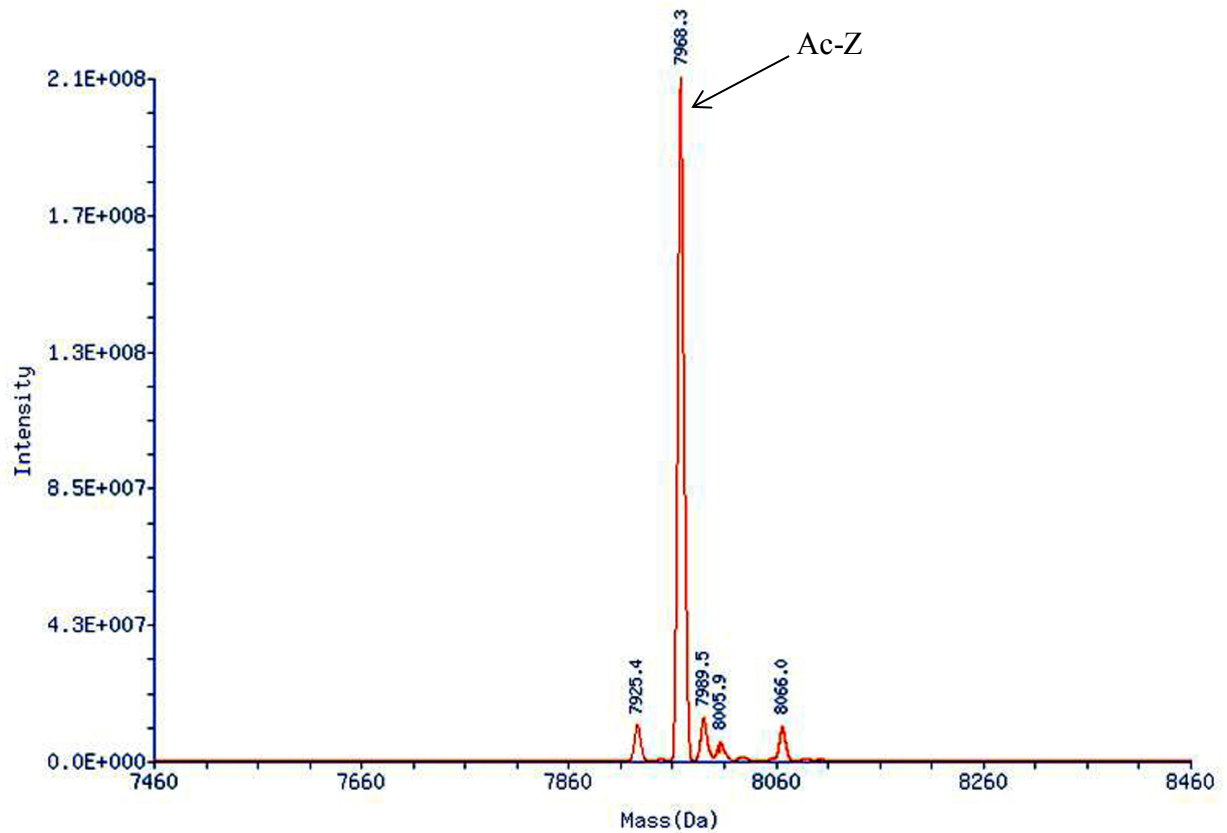


Figure 2.2 Mass-transformed ESI-MS spectrum of the Z-domain Bpa mutant that was expressed in BL21(DE3) cells co-transformed with pET-Z(S3TAG) and pSup-BpaRS-6TRN.

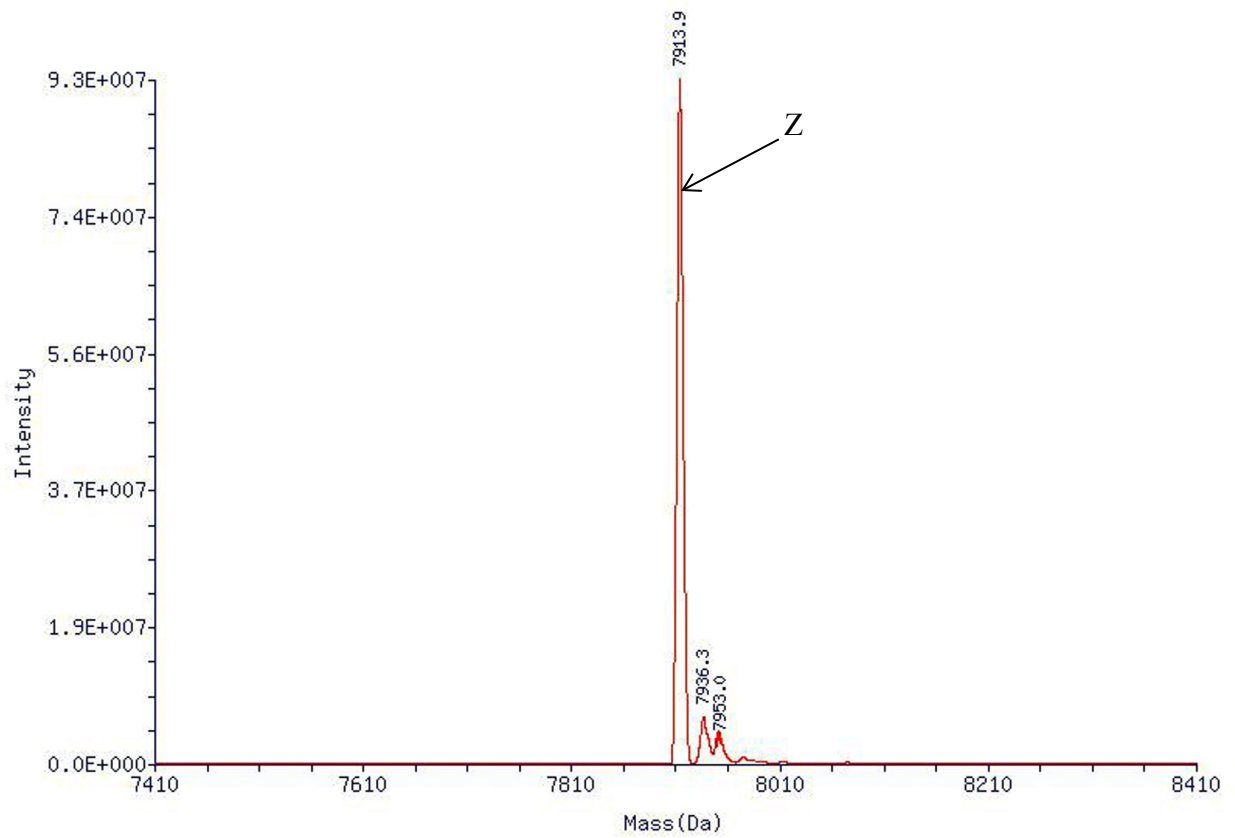


Figure 2.3 Mass-transformed ESI-MS spectrum of the Z-domain Bpa mutant that was expressed in BL21(DE3) cells co-transformed with pET-Z(V4TAG) and pSup-BpaRS-6TRN.

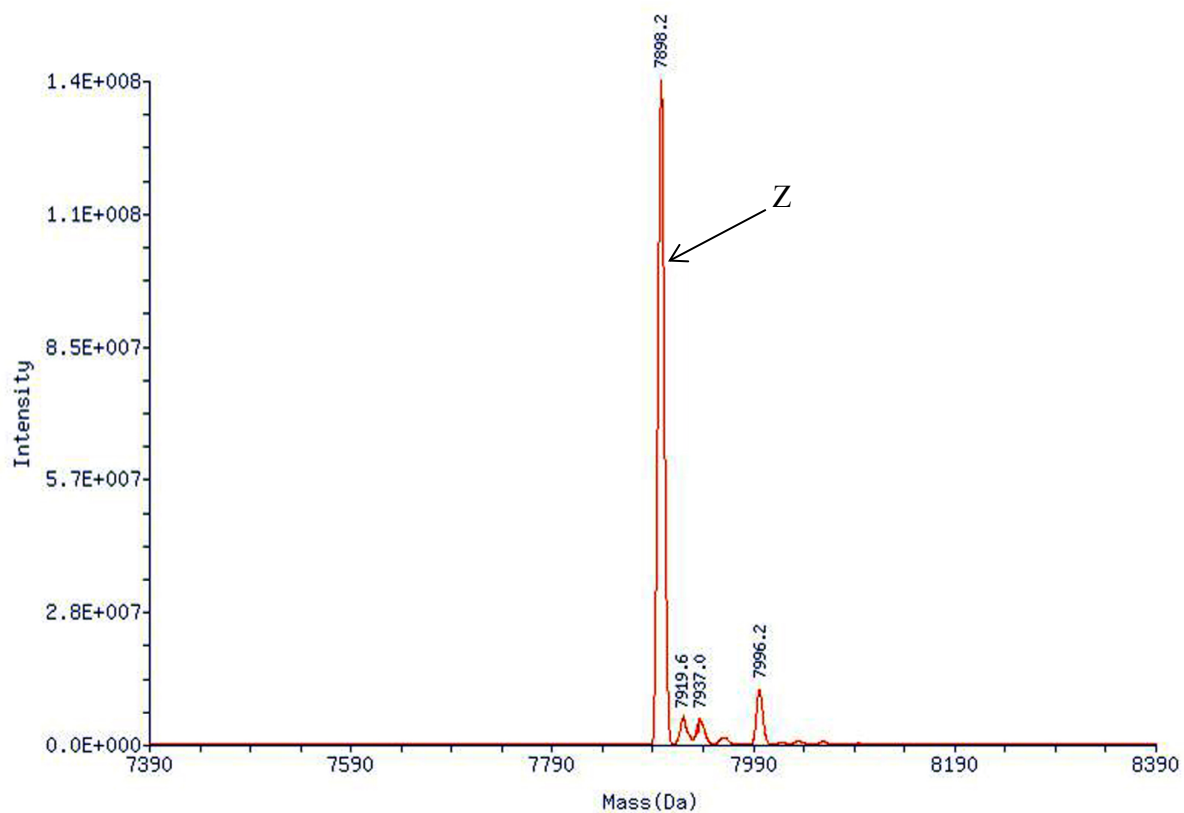


Figure 2.4 Mass-transformed ESI-MS spectrum of the Z-domain Bpa mutant that was expressed in BL21(DE3) cells co-transformed with pET-Z(D5TAG) and pSup-BpaRS-6TRN.

2.2.2 Unacetylated Z-domain using different *E.coli* strains and plasmids

In order to investigate the influence of differences in *E. coli* strains on N^α-acetylation of the Z-domain, we initially used pBAD-Z (Biotin), a plasmid which encodes the Z domain with C-terminal biotinylation and hexa-histidine tags under control of the arabinose-inducible araBAD promoter. First, we expressed the Z-domain protein from pBAD-Z (Biotin) in a specific *E. coli* strain, in which the *rimI*, *rimJ*, or *rimL* gene is deleted (JW4335, JW1053 and JW1423 respectively) to examine whether any of these ribosomal N-acetyltransferases were involved in the N^α-acetylation of the Z-domain. Mass spectrometry analysis of the purified proteins gave single peaks at around 10028 Da, which correspond to the Z-form (calculated average mass 10029.0 Da) (Table 2.2, entries 2-4; Figure 2.5).

Table 2.2 ESI-MS analysis of unacetylated Z-domain proteins expressed from different expression plasmids and *E. coli* strains.

Entry	<i>E. coli</i> strain	Expression plasmid	Supplementary Plasmid	Gene deleted	Observed major mass peaks/Da	Major forms of the Z-domain
1	AVB100	pBAD-Z		----	10028.2	Z ^a
2	JW4335	pBAD-Z		<i>rimI</i>	10028.3	Z
3	JW1053	pBAD-Z		<i>rimJ</i>	10028.2	Z
4	JW1423	pBAD-Z		<i>rimL</i>	10028.2	Z
5	JW2293	pBAD-Z		<i>Ack</i>	10028.3	Z
6	JW2294	pBAD-Z		<i>Pta</i>	10028.2	Z
7	JW4030	pBAD-Z		<i>Acs</i>	10028.3	Z
8	LCB90	pBAD-Z		<i>Ack</i>	10028.2	Z
9	XL1-BLUE	pBAD-Z			7761.9	Z
10	XL1-BLUE	pBAD-Z	pREP2-YC		7761.9	Z
11	JM109	pBAD-Z			7762.2	Z
12	JM109	pBAD-Z	pREP2-YC		7762.2	Z
13	JM109(DE3)	pBAD-Z			7764.5	Z
14	JM109(DE3)	pET-Z			7764.5	Z
15	AD494(DE3)	pET-Z			7761.8	Z
16	K12(WT)	pBAD-Z			7764.1	Z
17	B(WT)	pBAD-Z			7763.8	Z
18	MG1655	pBAD-Z			7763.7	Z
19	BL21(DE3)	pBAD-Z			7763.1	Z

^a Unacetylated Z-domain without the initiating methionine (Met₁) residue.

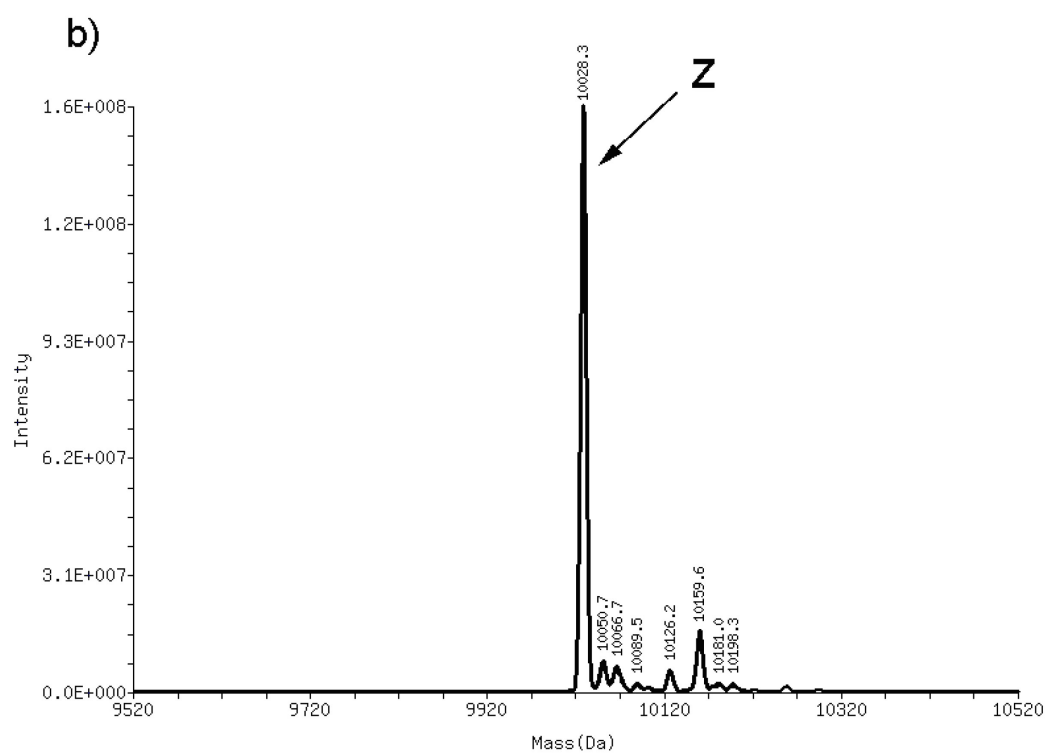
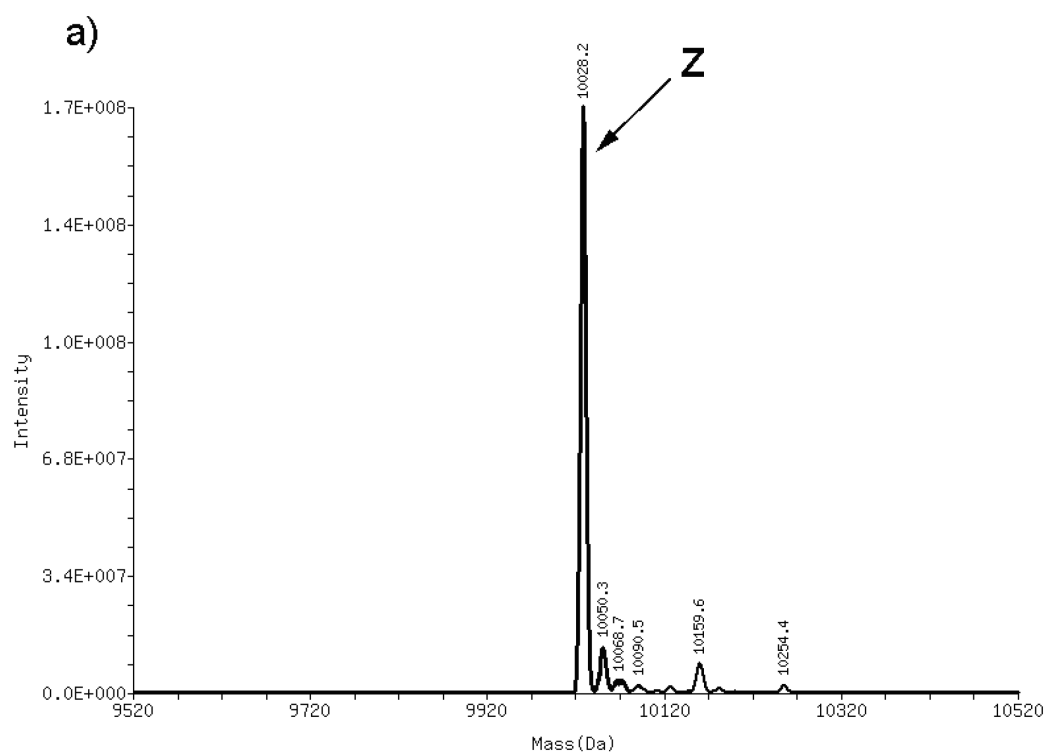


Figure 2.5 Mass-transformed ESI-MS spectra of the Z-domain proteins transformed with pBAD-Z(Biotin) in different *E. coli* strains: a) AVB100 and b) JW 4335.

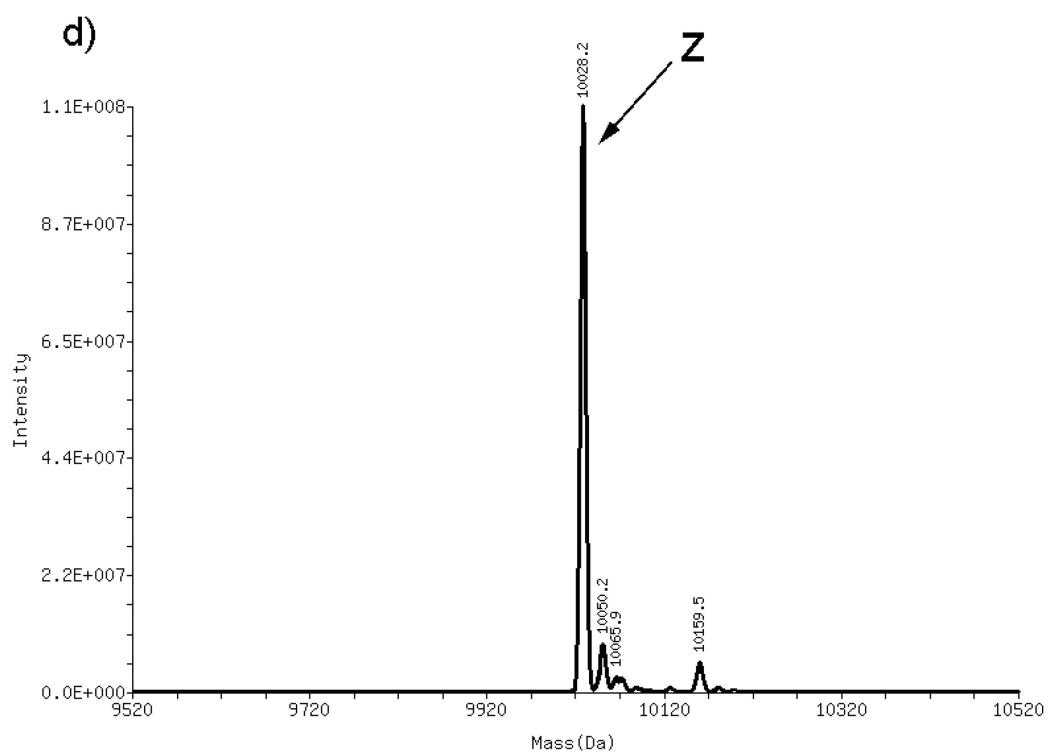
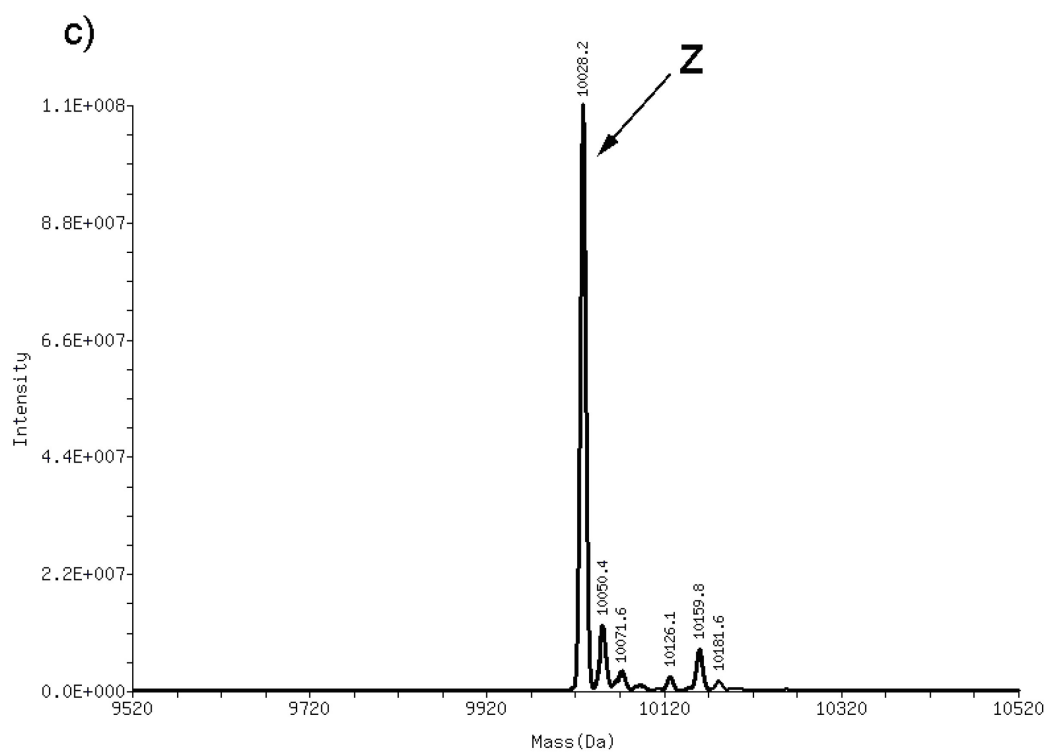


Figure 2.5 Cont. Mass-transformed ESI-MS spectra of the Z-domain proteins transformed with pBAD-Z(Biotin) in different *E. coli* strains: c) JW 1053 and d) JW 1423.

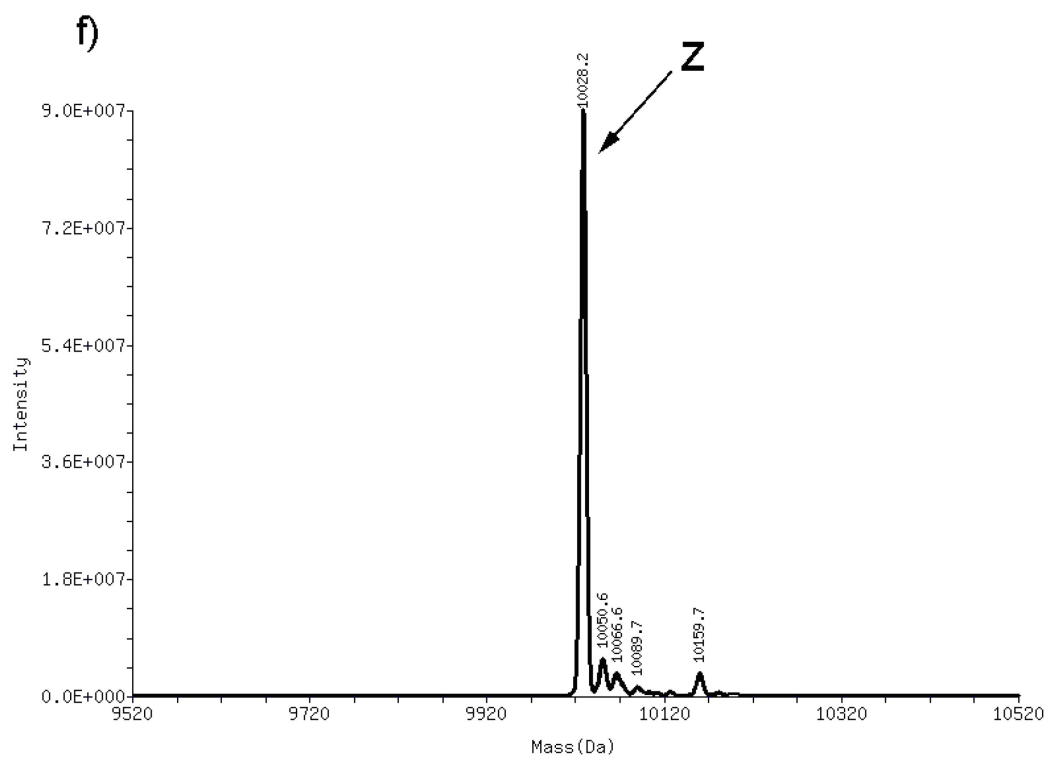
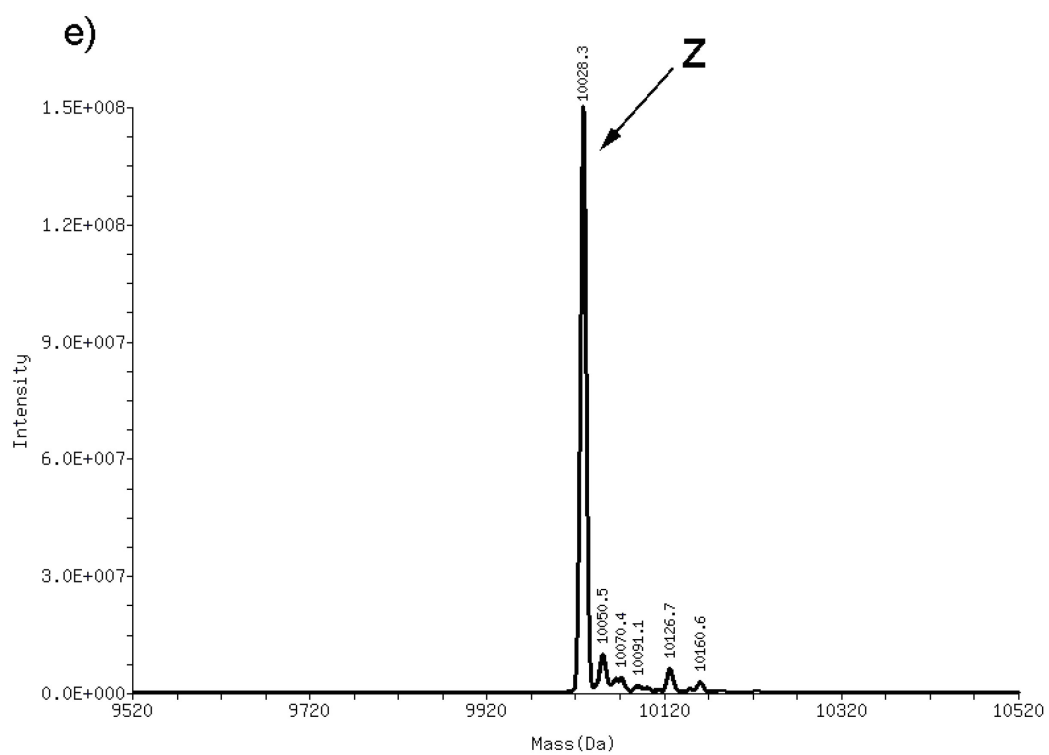


Figure 2.5 Cont. Mass-transformed ESI-MS spectra of the Z-domain proteins transformed with pBAD-Z(Biotin) in different *E. coli* strains: e) JW 2293 and f) JW 2294.

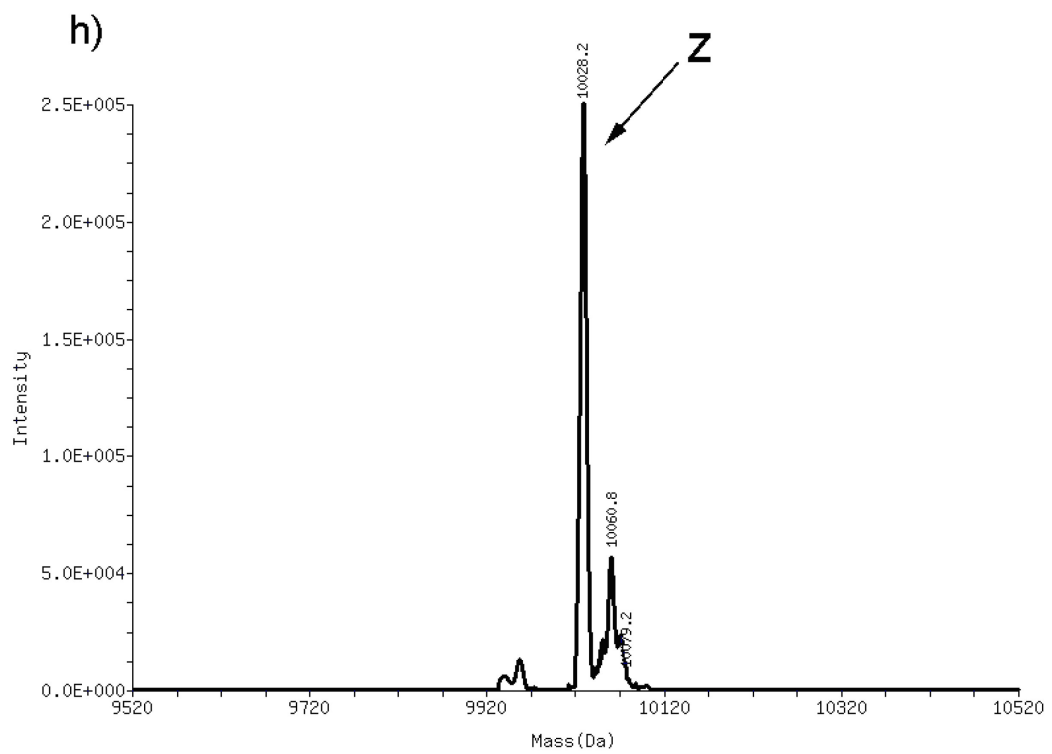
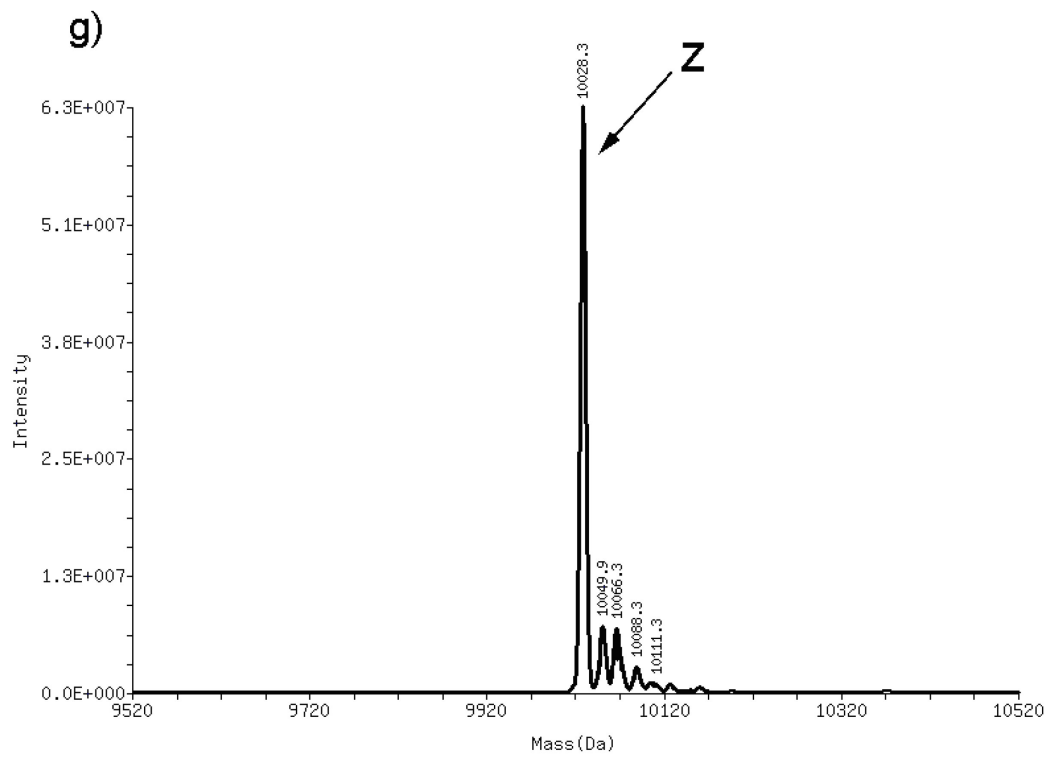


Figure 2.5 Cont. Mass-transformed ESI-MS spectra of the Z-domain proteins transformed with pBAD-Z(Biotin) in different *E. coli* strains: g) JW 4030 and h) LCB90.

Also, we investigated three enzymes involved in ϵ -N-acetylation of CheY, a flagella protein which plays a major role in *E. coli* chemotaxis. These enzymes are, phosphotransacetylase (*Pta*), acetate kinase (*Ack*), and acetyl-CoA synthetase (*Acs*).⁸⁴ The Z-domain proteins were expressed in the *E. coli* strain, in which each enzyme was not present (JW2294, JW4030 and LCB90, respectively). ESI-MS of the isolated proteins gave single peaks at around 10028 Da, which correspond to the Z form (Table 2.2, entries 6-8; Figure 2.5).

We next examined whether there was any difference in the acetylation pattern when the Z-domain is expressed in various B and K12 strains of *E. coli* deficient in proteases Lon and OmpT. B strains are more suitable for high expression yields than the K12 strains. The *E. coli* B strains produce less amounts of acetate than the K12 strains.⁸⁵ In order to test whether all these differences have impact on the Z-domain N ^{α} -acetylation, we expressed the Z-domain in *E. coli* B wild type strains and K12 strains. The Z-form was mostly obtained in these strains (Table 2.2, entries 9, 11, 13-19; Figures 2.6-2.14).

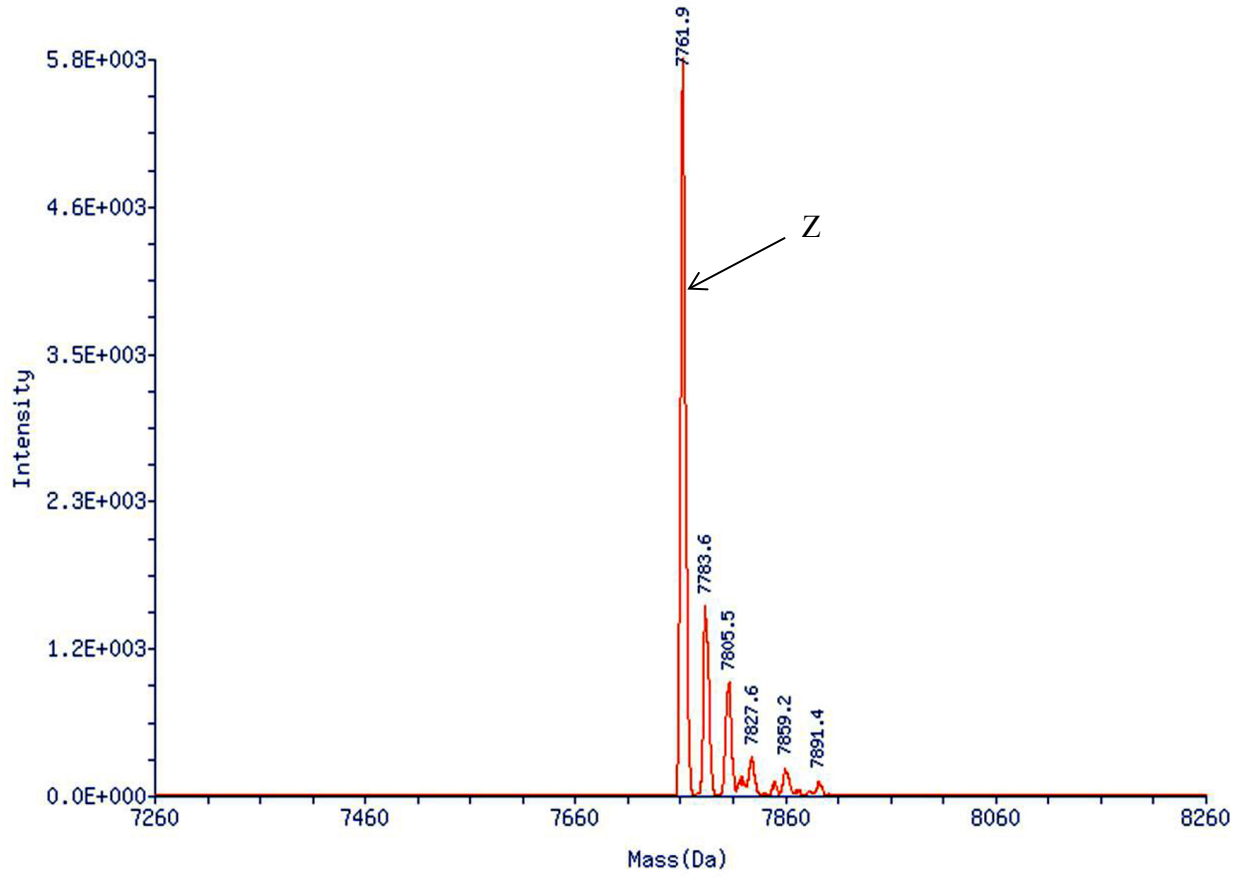


Figure 2.6 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in XL1-BLUE cells transformed with pBAD-Z.

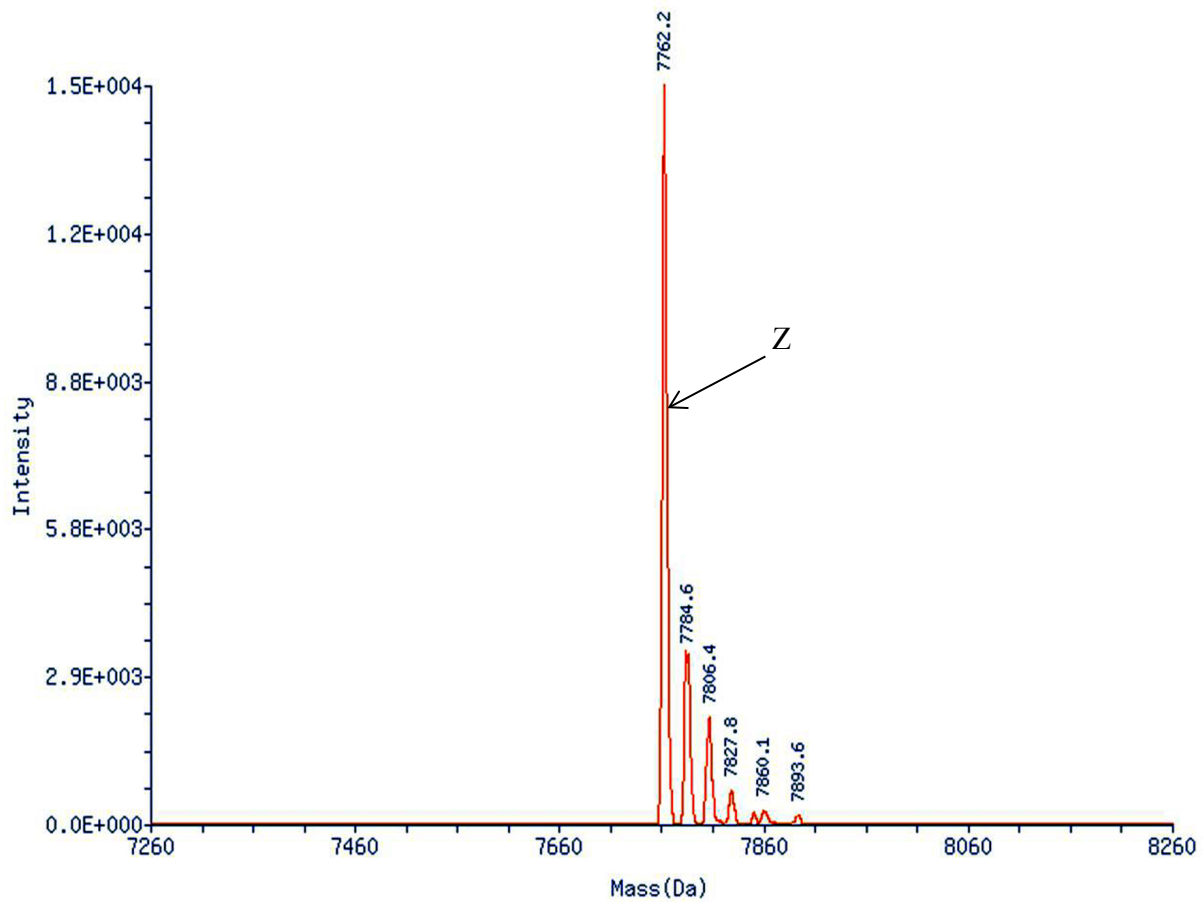


Figure 2.7 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109 cells transformed with pBADZ.

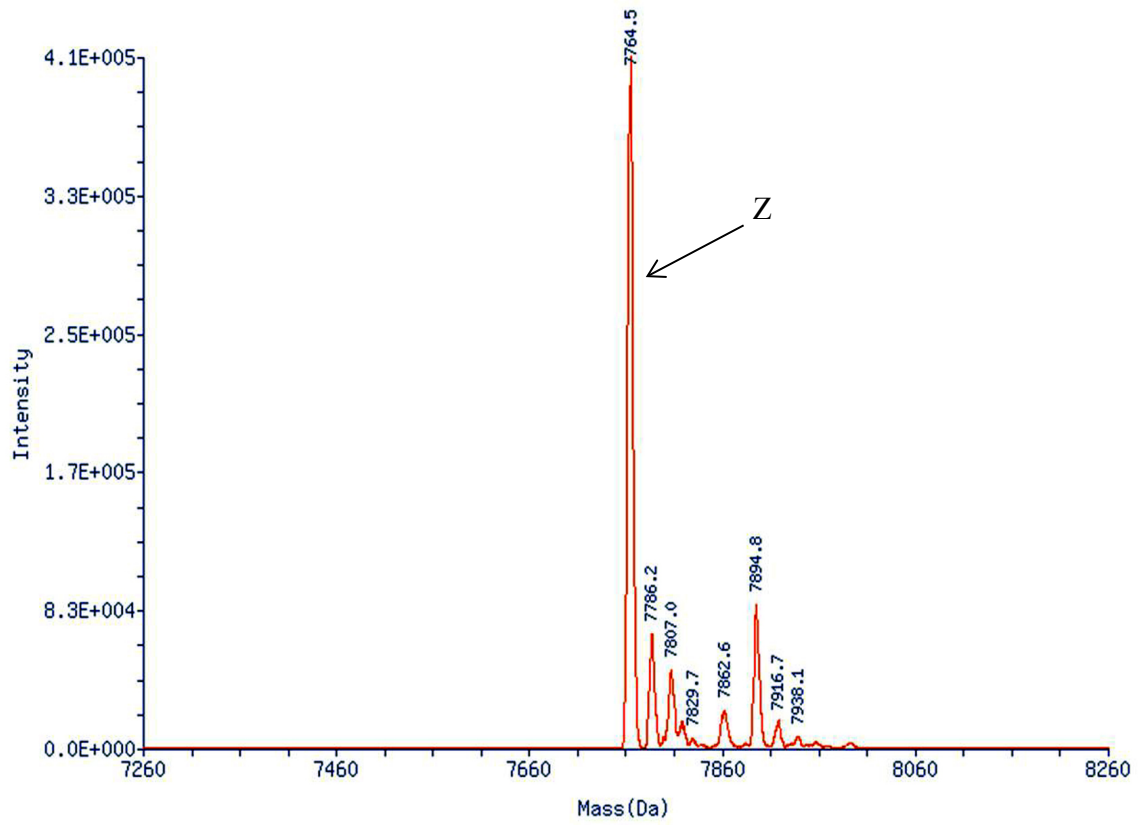


Figure 2.8 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells transformed with pBAD-Z.

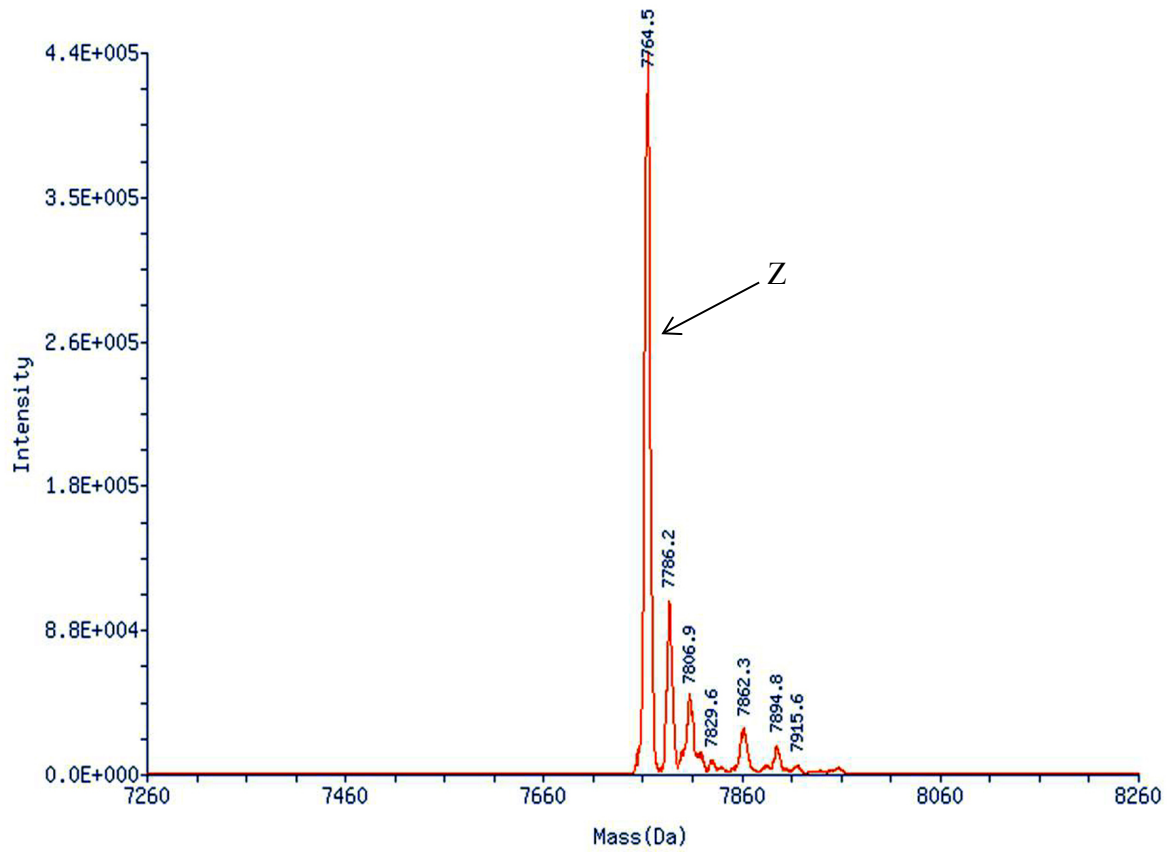


Figure 2.9 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells transformed with pET-Z.

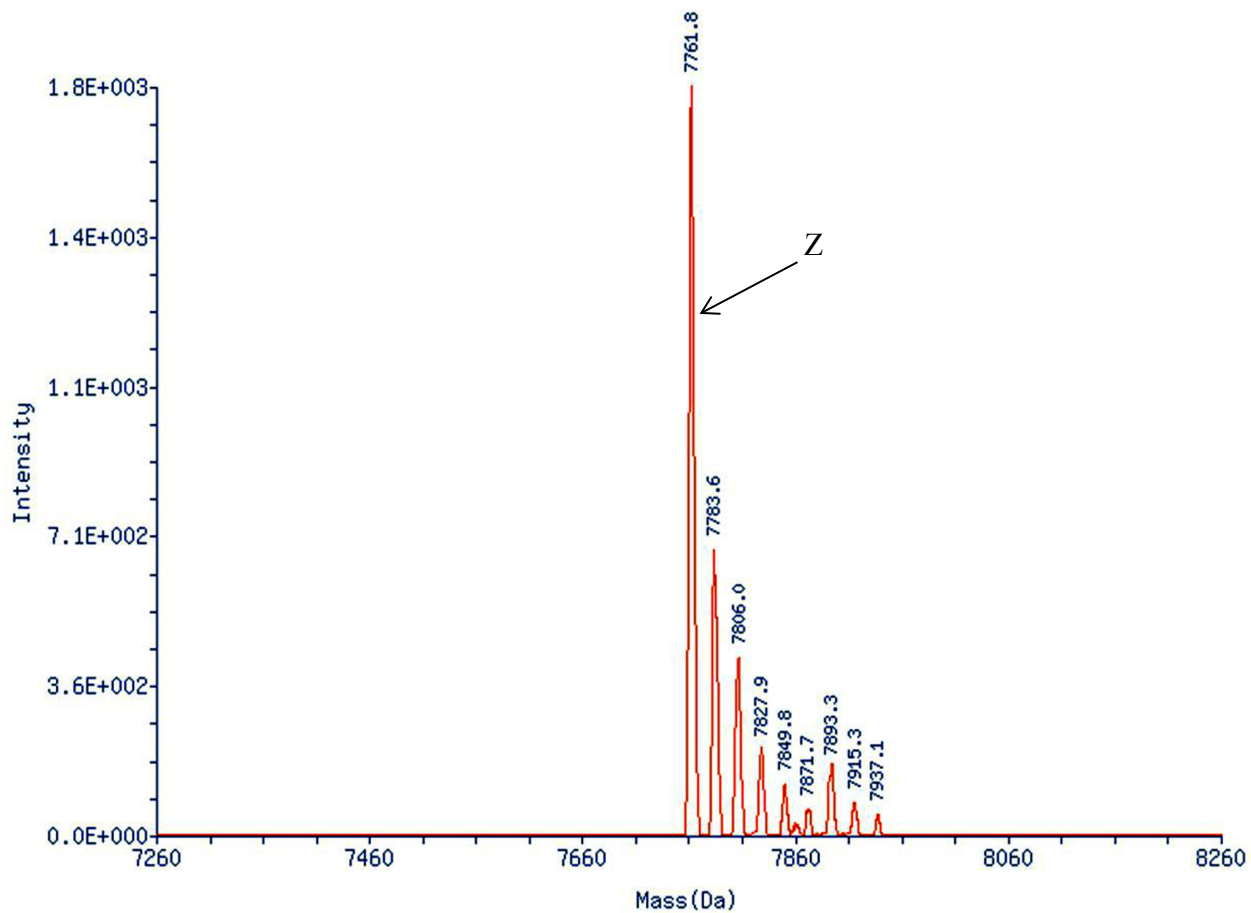


Figure 2.10 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in AD494(DE3) cells transformed with pET-Z.

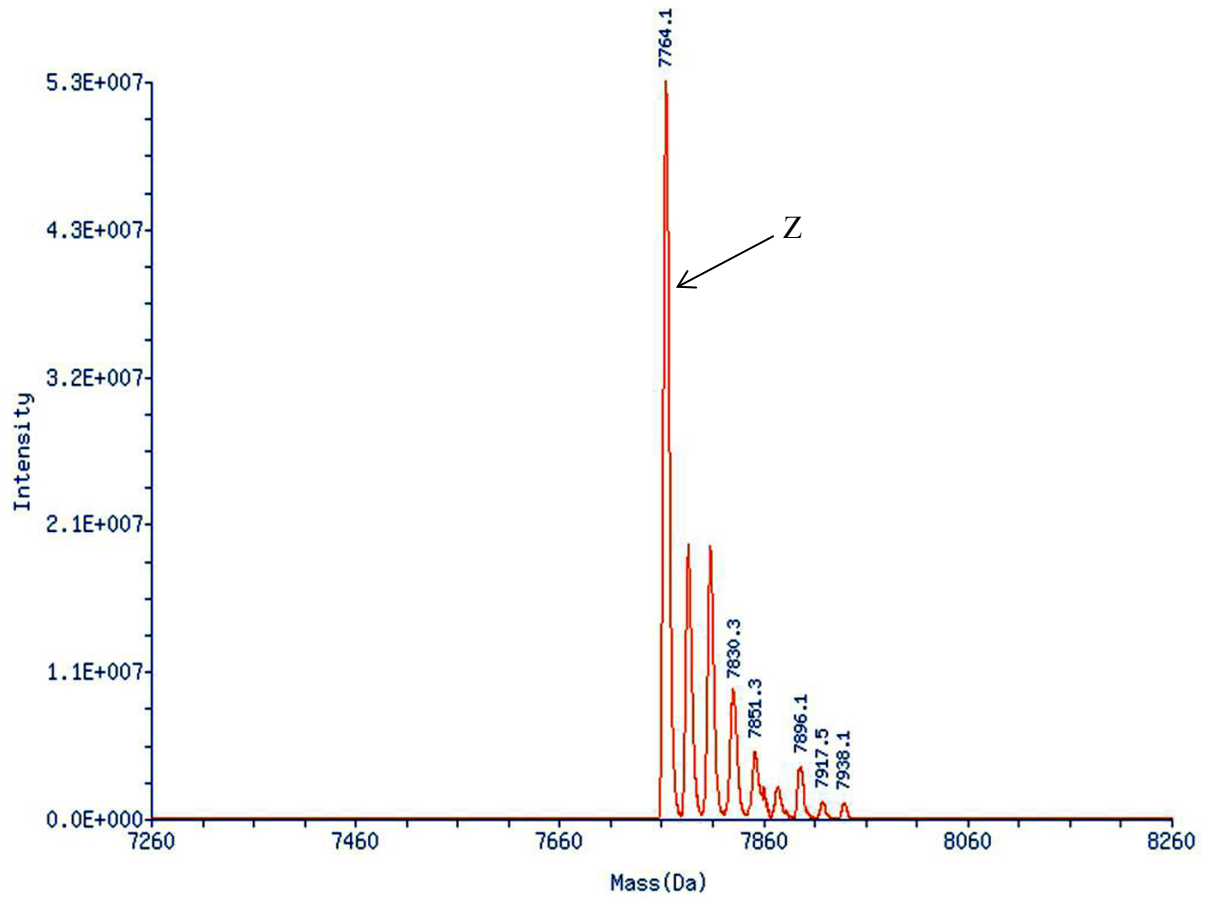


Figure 2.11 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in K12(WT) cells transformed with pBAD-Z.

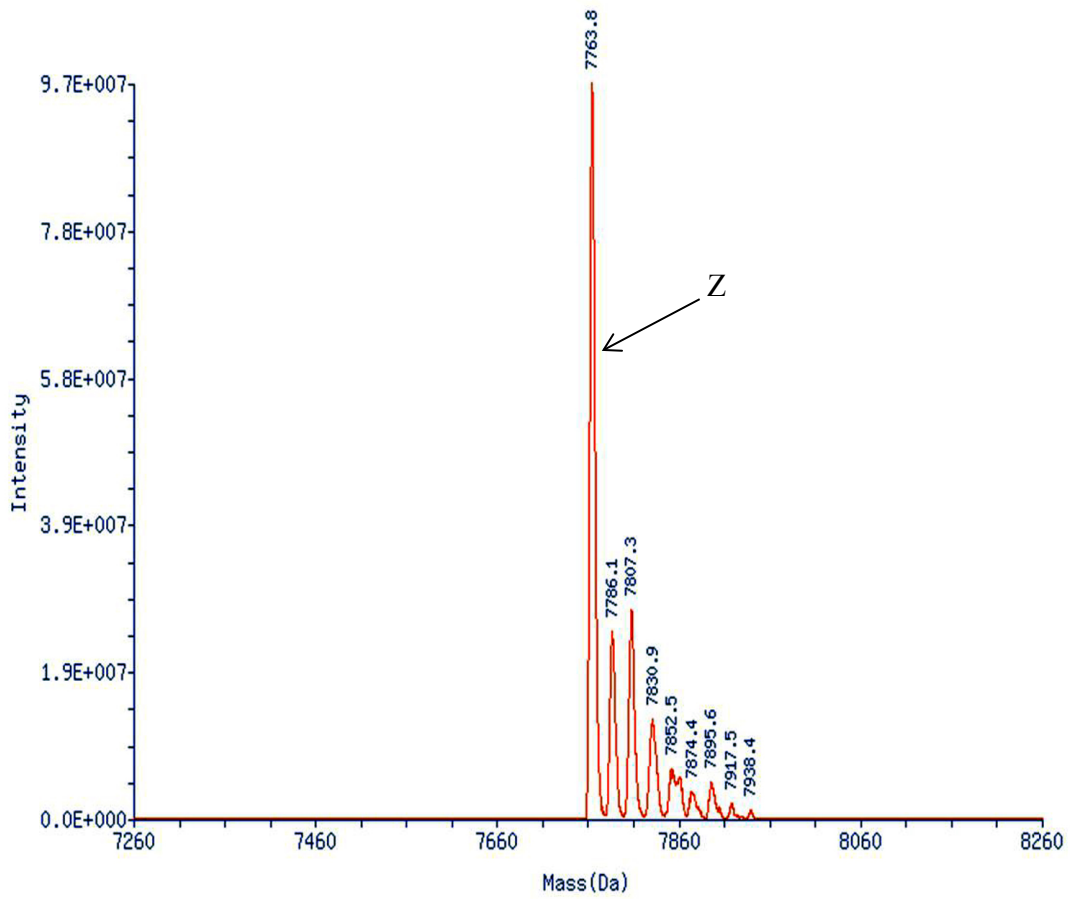


Figure 2.12 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in B(WT) cells transformed with pBAD-Z.

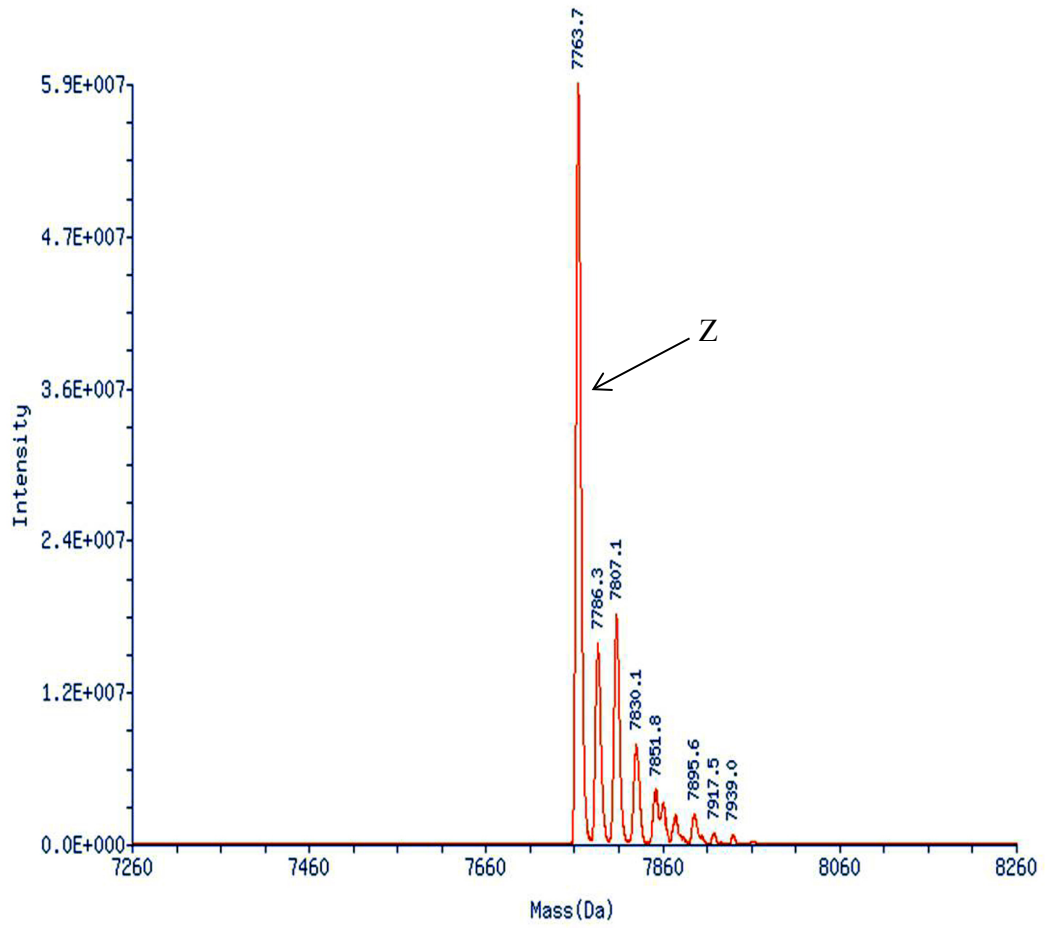


Figure 2.13 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in MG1655(Seq) cells transformed with pBAD-Z.

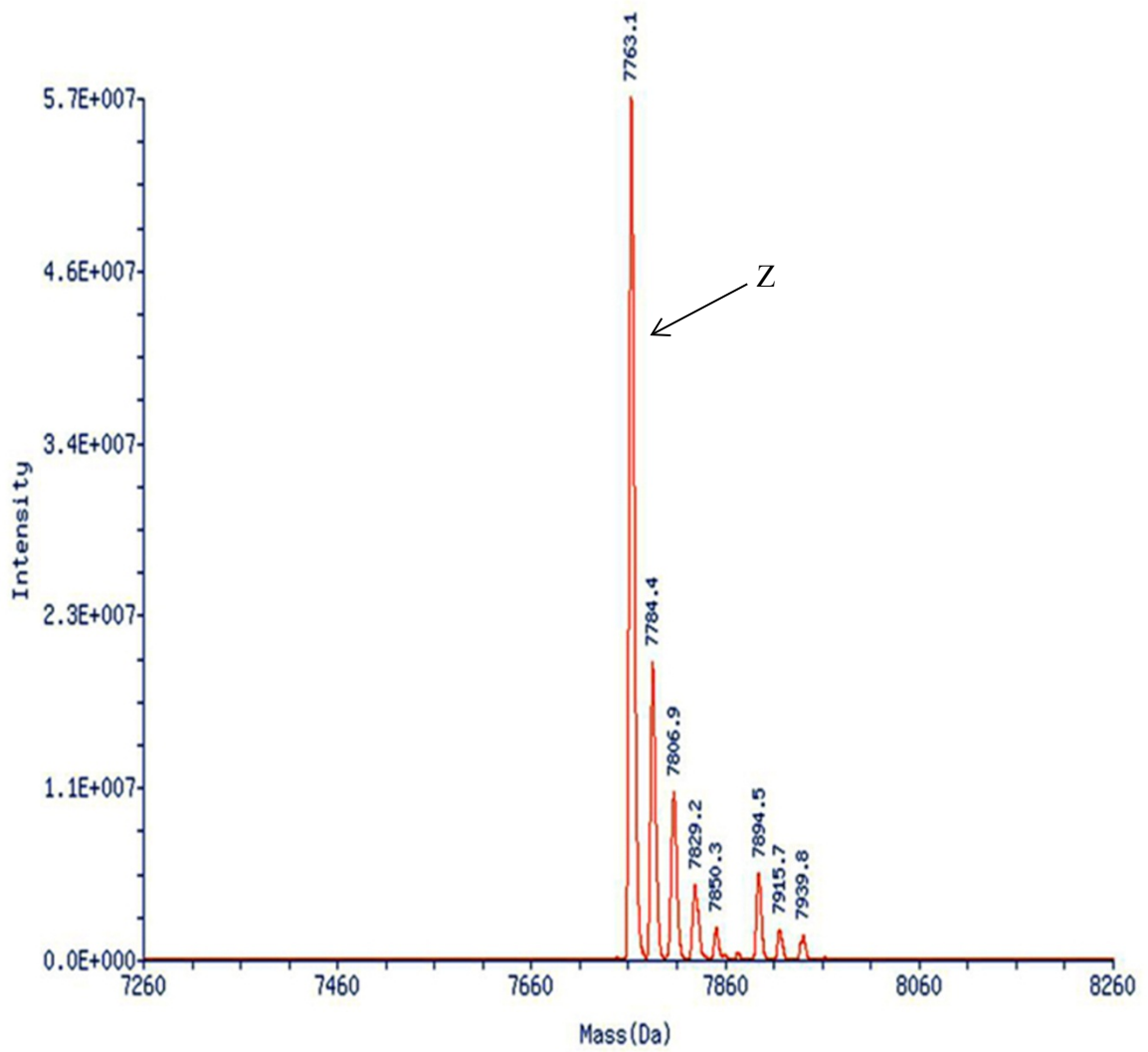


Figure 2.14 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells transformed with pBAD-Z.

2.2.3 T7 RNA Polymerase does not contribute to acetylation of the Z-domain

In order to examine whether T7 RNA polymerase had impact on the Z-domain N^α-acetylation, we expressed the protein in the *E. coli* cells harboring pREP-2YC, a plasmid which encodes the T7 RNA polymerase gene under control of the araBAD promoter. When expressed from pBAD-Z in XL1-BLUE or JM109 cells in the presence of pRep2-YC, the Z-form was mainly obtained with an observed mass of 7762.2 Da (calculated mass of 7762.2 Da). It should be noted that no acetylation was observed in JM19(DE3) or BL21(DE3) cells, which harbor the chromosomal copy of the T7 RNA polymerase gene. Therefore, T7 RNA polymerase did not have any direct effect on the N^α-acetylation (Table 2.2, entries 10 and 12; Figures 2.15 and 2.16).

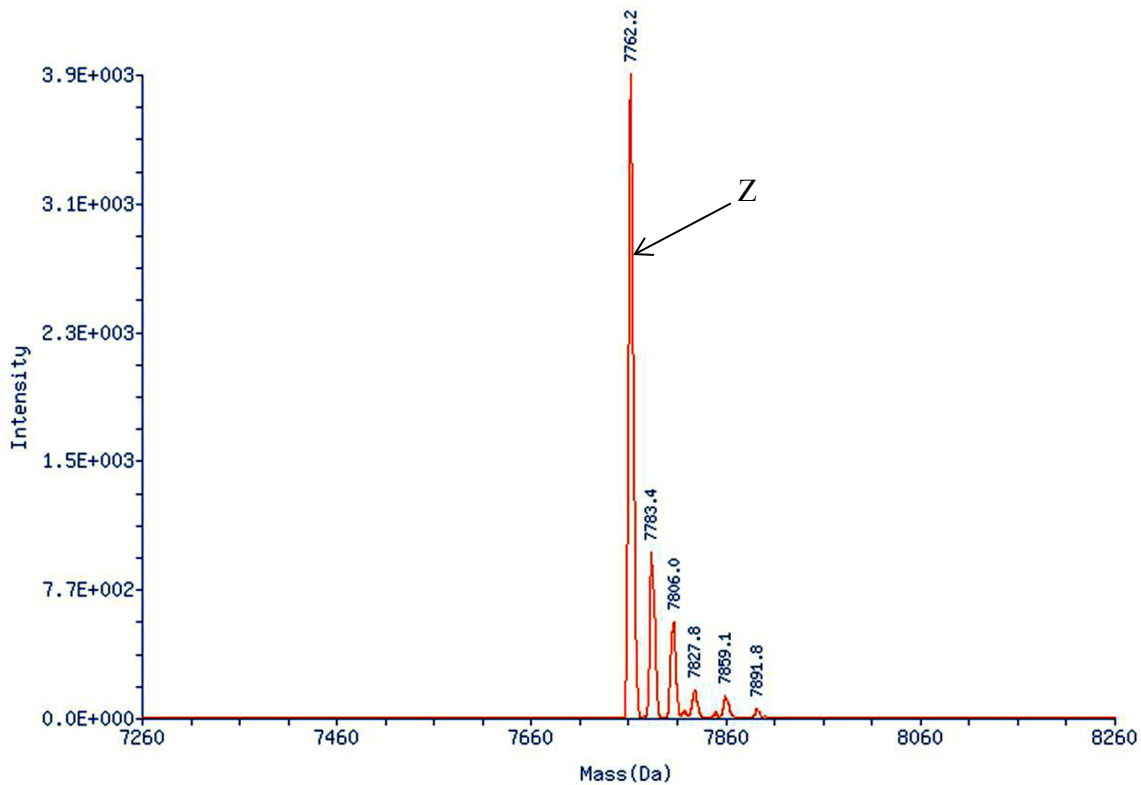


Figure 2.15 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in XL1-BLUE cells co-transformed with pBAD-Z and pRep2-YC.

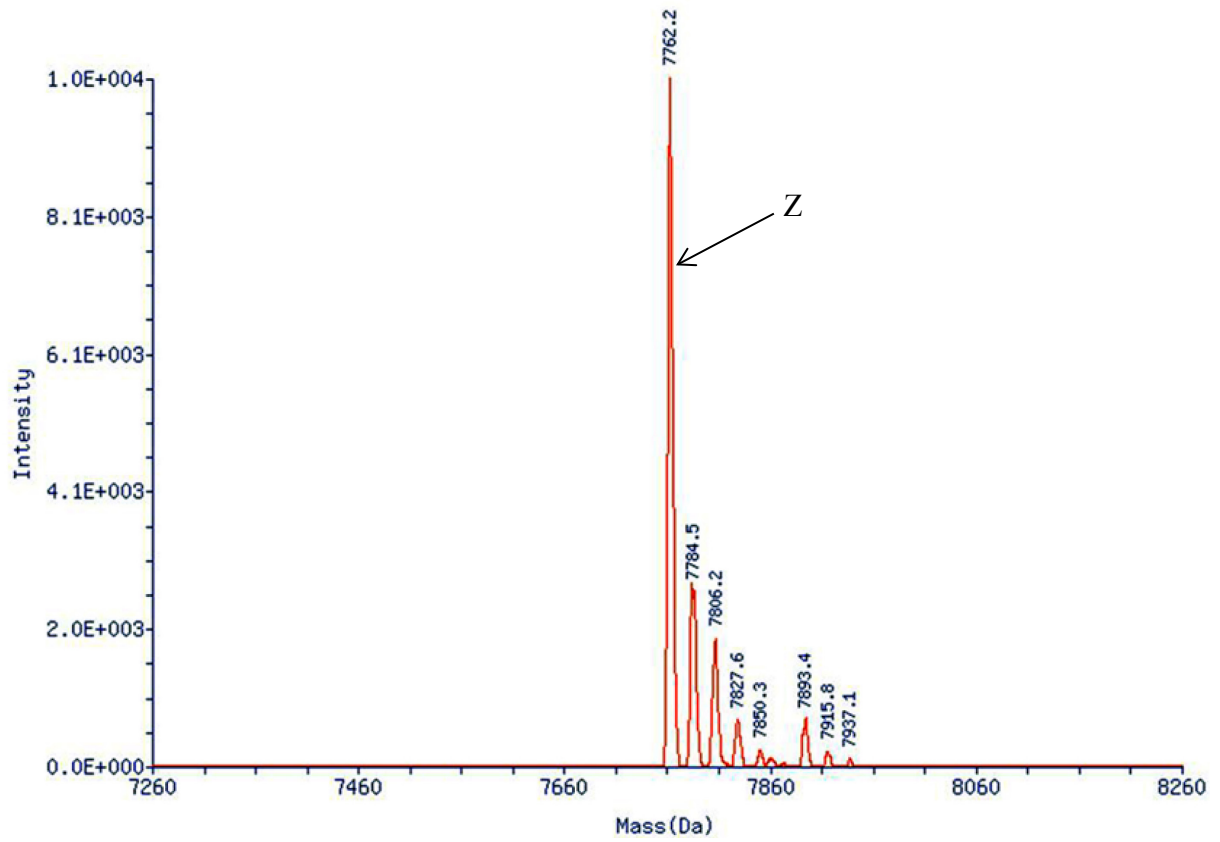


Figure 2.16 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109 cells co-transformed with pBAD-Z and pRep2-YC.

2.2.4 N^α-Acetylation of the Z-domain is catalyzed by RimJ

As initially observed for the Z-domain Ser3Bpa mutant, N^α-acetylation was also detected for the Z-domain expressed in BL21(DE3) *E. coli* cells co-transformed with pET-Z and pSup-JYRS-6TRN. The latter plasmid encodes six copies of amber suppressor tRNA^{Tyr} and tyrosyl tRNA synthetase, both of which were derived from *Methanococcus jannaschii*. Although required only for the incorporation of tyrosine into the Z-domain in response to amber stop codons, pSup-JYRS-6TRN was purposely included in the wild type Z-domain expression to directly compare the expression levels of the wild type and mutant proteins in the presence of the plasmid. The transformed cells were grown at 37°C in PA-5052, a chemically defined auto-inducing rich medium in which the protein expression is induced by lactose as soon as glucose is deprived. The overexpressed Z-domain protein was purified from the cell lysate by IMAC in an overall yield of 100 mg L⁻¹. SDS-PAGE analysis revealed that the Z-domain was highly expressed in *E. coli* and purified to homogeneity (Figure 2.17).

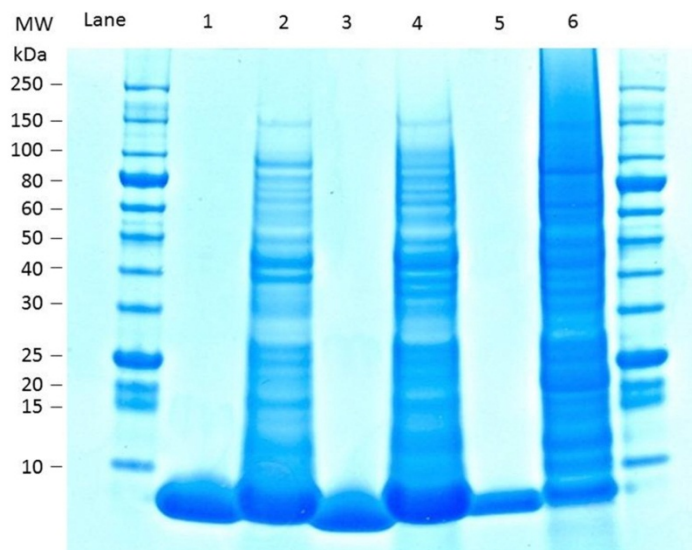


Figure 2.17 SDS-PAGE analysis of the Z-domain in *E. coli* cell lysates and the protein purified by immobilized Ni ion-affinity chromatography. Lanes 1 and 2: the Z-domain purified from BL21(DE3) *E. coli* cells co-transformed with pET-Z and pSup-JYRS-6TRN, and its corresponding cell lysate, respectively; Lanes 3 and 4: the Z-domain purified from BL21(DE3) *E. coli* cells transformed with pET-Z alone and its corresponding cell lysate, respectively; Lanes 5 and 6: the Z-domain purified from BL21(DE3) *E. coli* cells co-transformed with pBAD-Z and pACYCDuet-RimJ, and its corresponding cell lysate, respectively.

ESI-MS of the purified protein gave upon mass-transformation (“deconvolution”) two major mass peaks at 7762.9 Da and 7807.1 Da, which correspond to the the Z form (calculated average mass 7762.4 Da) and the Ac-Z form (calculated average mass 7804.5 Da), respectively (Table 2.3, entry 2; Figure 2.18). Also detected were the mass peaks at 7784.7 Da and 7829.4 Da, which correspond to the Na⁺ adducts of the Z and Ac-Z forms, respectively. On the other hand, when the protein was expressed from pET-Z in BL21(DE3) *E. coli* cells without pSup-JYRS-6TRN, ESI-MS of the isolated protein only showed a single major peak upon mass-transformation at 7763.5 Da, which corresponds to the Z form (Table 2.3, entry 1; Figure 2.19). The presence of pSup-JYRS-6TRN, therefore, had a significant impact on the N^α-acetylation of the Z-domain in BL21(DE3) cells.

Table 2.3 ESI-MS analysis of the Z-domain proteins expressed from different expression plasmids and wild type *E. coli* strains.

Entry	<i>E. coli</i> strain	Expression plasmid	Supplementary plasmid	Observed major mass peaks/Da	Major forms of the Z-domain
1	BL21(DE3)	pET-Z		7763.5	Z ^a
2	BL21(DE3)	pET-Z	pSup-JYRS-6TRN	7762.9, 7807.1	Z, Ac-Z ^b
3	BL21(DE3) Δ <i>rimJ::Kan</i>	pET-Z	pSup-JYRS-6TRN	7763.5	Z
4	BL21(DE3)	pET-Z	pACYCDuet-RimJ	7804.6	Ac-Z
5	BL21(DE3) Δ <i>rimJ::Kan</i>	pET-Z	pACYCDuet-RimJ	7763.2	Z
6	BL21(DE3)	pET-Z	pACYCDuet	7762.4	Z
7	BL21(DE3)	pBAD-Z	pACYCDuet-RimJ	7805.0	Ac-Z
8	BL21(DE3) Δ <i>rimJ::Kan</i>	pBAD-Z	pACYCDuet-RimJ	7763.8	Z
9	JM109(DE3)	pET-Z		7764.5	Z
10	JM109(DE3)	pET-Z	pACYCDuet	7763.6	Z
11	JM109(DE3)	pET-Z	pSup-JYRS-6TRN	7762.1	Z
12	JM109(DE3)	pET-Z	pACYCDuet-RimJ	7764.5, 7806.7	Z, Ac-Z
13	DH10B	pBAD-Z		7763.5	Z
14	DH10B	pBAD-Z	pSup-JYRS-6TRN	7762.2	Z
15	DH10B	pBAD-Z	pACYCDuet-RimJ	7761.5	Z
16	DH10B	pBAD-Z	pACYCDuet	7762.2	Z

^a Unacetylated Z-domain without Met₁ residue. ^b Acetylated Z-domain without the Met₁

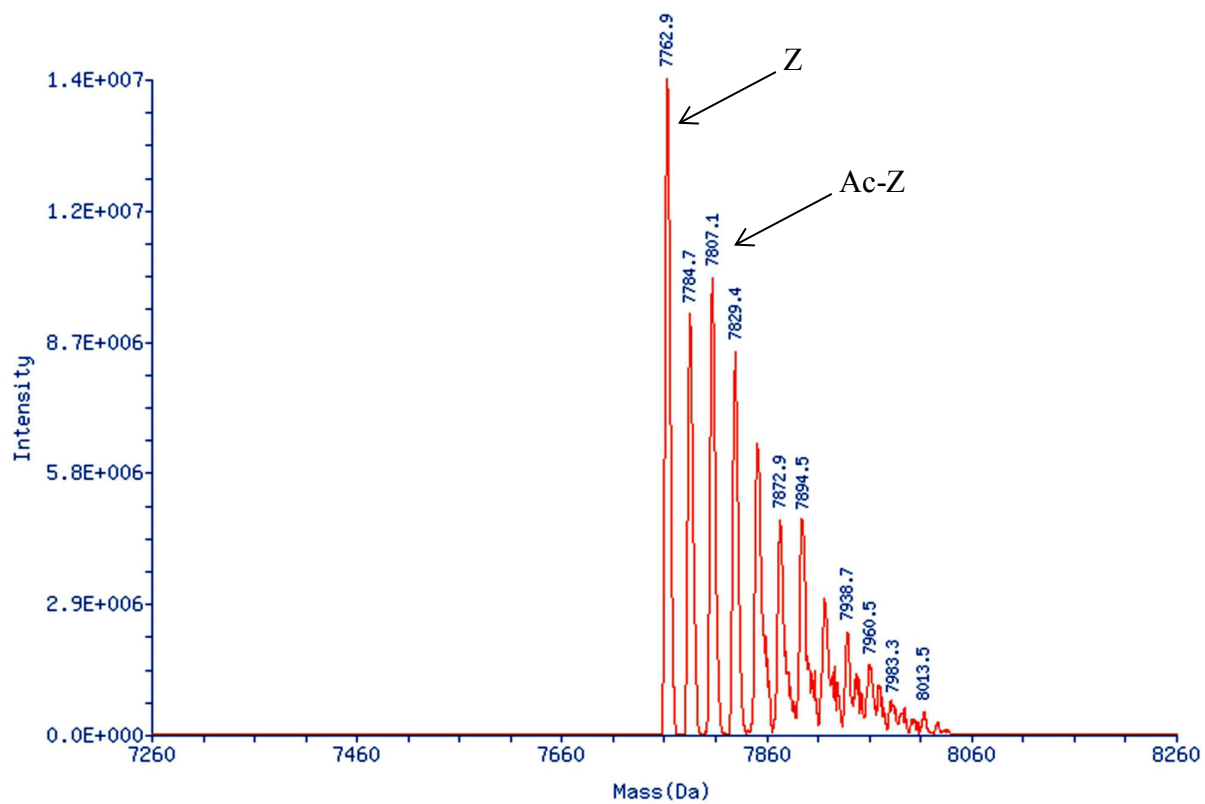


Figure 2.18 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pET-Z and pSup-JYRS-6TRN.

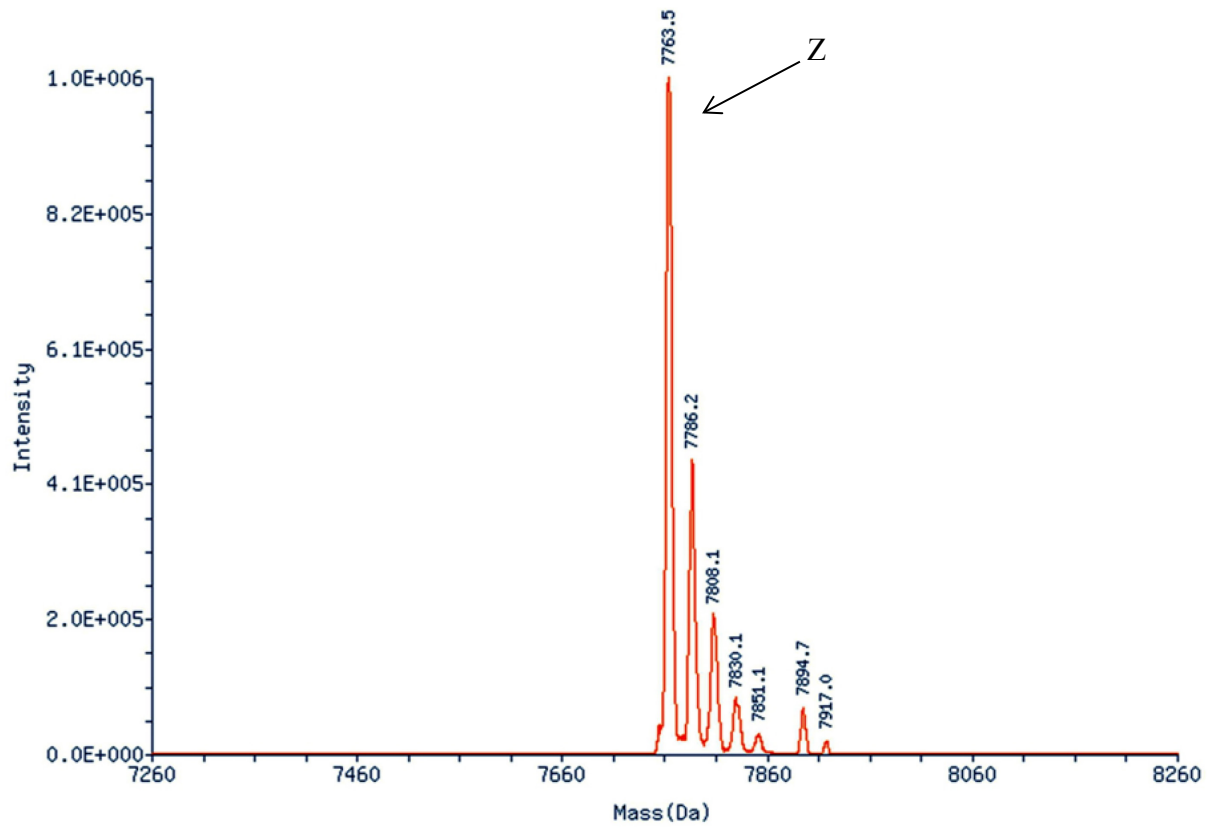


Figure 2.19 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells transformed with pET-Z.

RimJ was identified as the NAT responsible for the N^α-acetylation of the Tα1 fusion protein in *E. coli*. Therefore, in order to investigate whether this NAT is also responsible for the N^α-acetylation of the Z-domain, we constructed pACYCDuet-RimJ, a p15A-based plasmid that encodes RimJ under control of the T7 promoter. Indeed, this plasmid, when co-expressed with pET-Z in BL21(DE3) *E. coli*, provided the Ac-Z form as the major product, although a minor Z-form was also detected (Table 2.3, entry 4; Figure 2.20). In a control experiment, Z-domain was expressed in BL21(DE3) co-transformed with pET-Z and pACYCDuet in which the Z-form was obtained as the major product (Table 2.3, entry 6; Figure 2.21). The latter plasmid does not encode the *rimJ* gene.

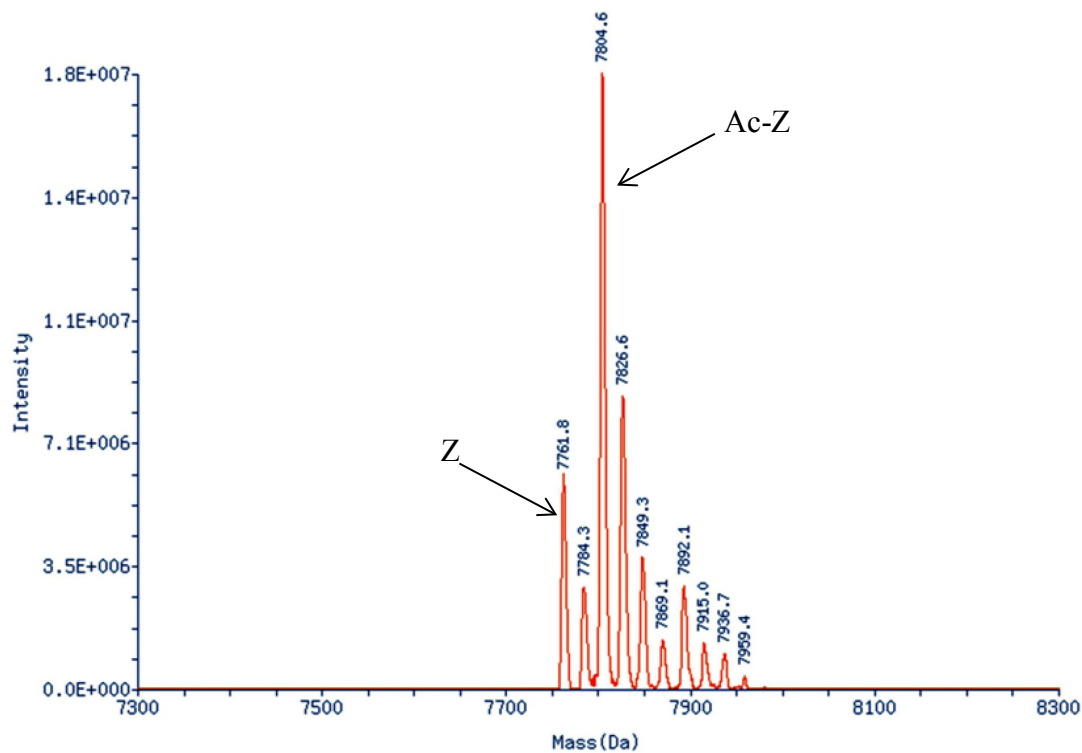


Figure 2.20 Mass-transformed ESI-MS spectra of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pET-Z and pACYCDuet-RimJ.

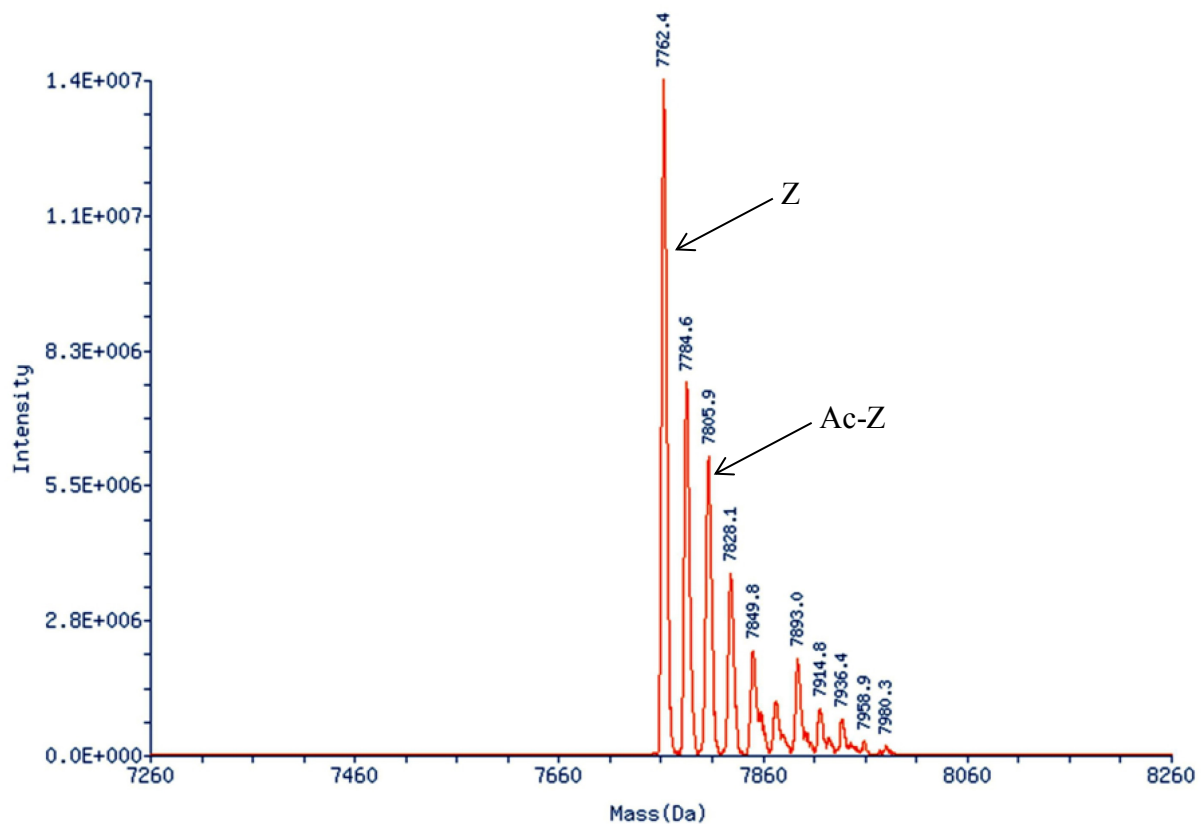


Figure 2.21 Mass-transformed ESI-MS spectra of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pET-Z and pACYCDuet.

When the protein was expressed in JM109(DE3) *E. coli* cells co-transformed with pACYCDuet, ESI-MS of the isolated protein showed a major peak at 7763.6 Da, corresponding to the Z form (Table 2.3, entry 10; Figure 2.22). When the Z-domain was expressed in JM109(DE3) co-transformed with pET-Z and pSup-JYRS-6TRN, a slight increase in N^α-acetylation of the Z-domain was observed (Table 2.3, entry 11; Figure 2.23). Consistent with the observation that the Z-domain was acetylated in BL21(DE3) cells in the presence of pSup-JYRS-6TRN, the Z-domain N^α-acetylation seems more efficient in BL21(DE3) cells than in JM109(DE3) cells. The presence of pACYCDuet-RimJ significantly increased the N^α-acetylation of the Z-domain in JM109(DE3) cells (Table 2.3, entry 12; Figure 2.24).

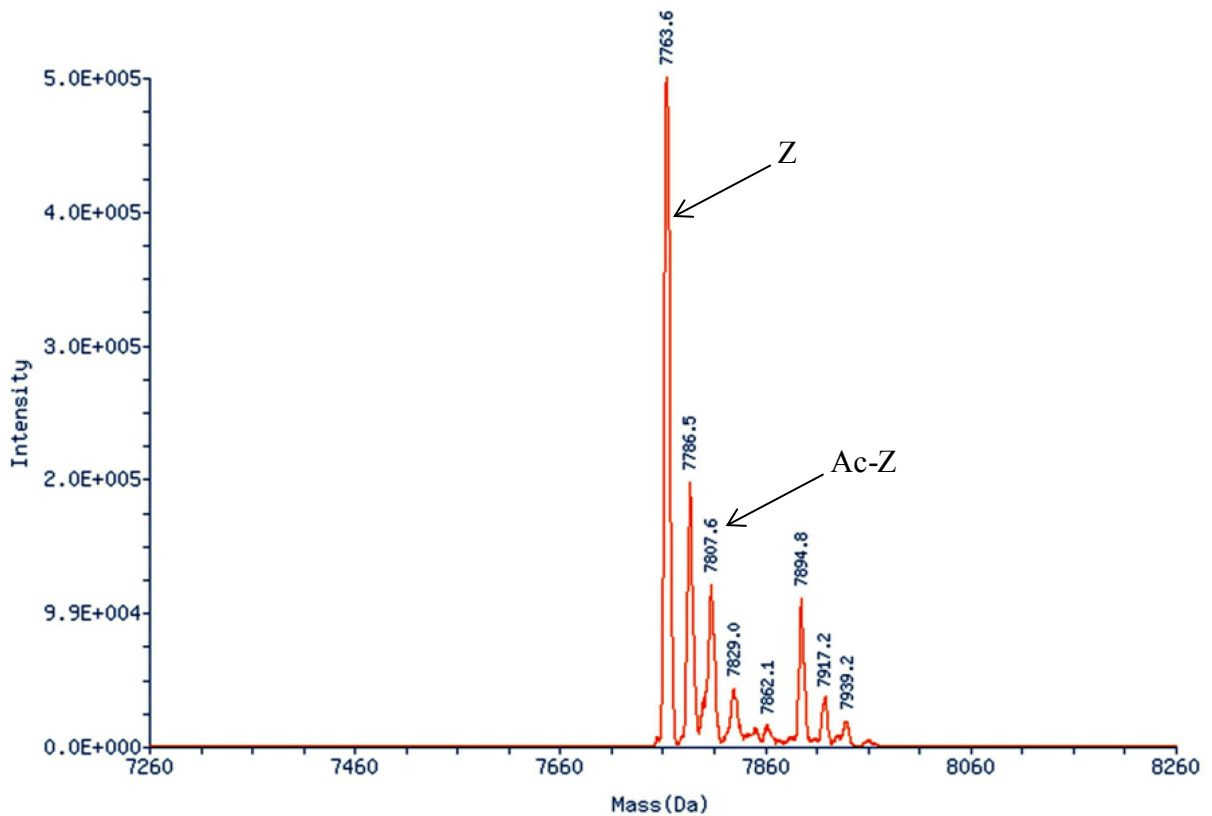


Figure 2.22 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells co-transformed with pET-Z and pACYCDuet.

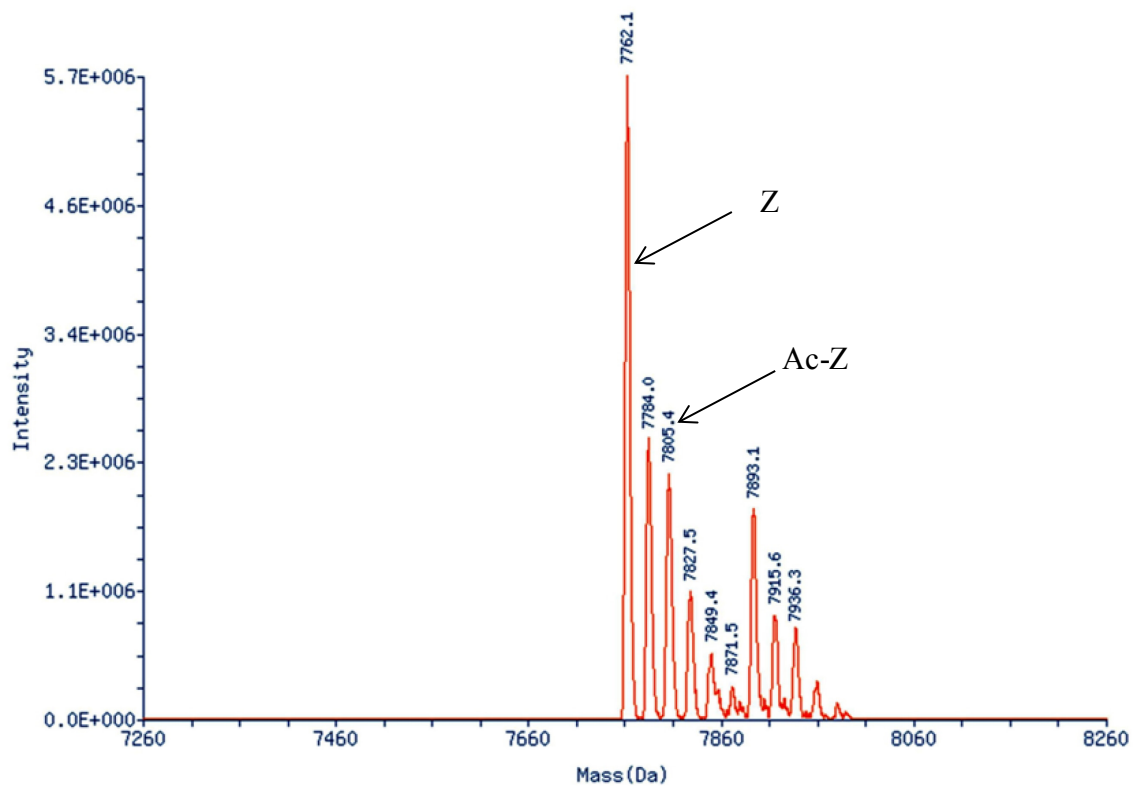


Figure 2.23 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells co-transformed with pET-Z and pSup-JYRS-6TRN.

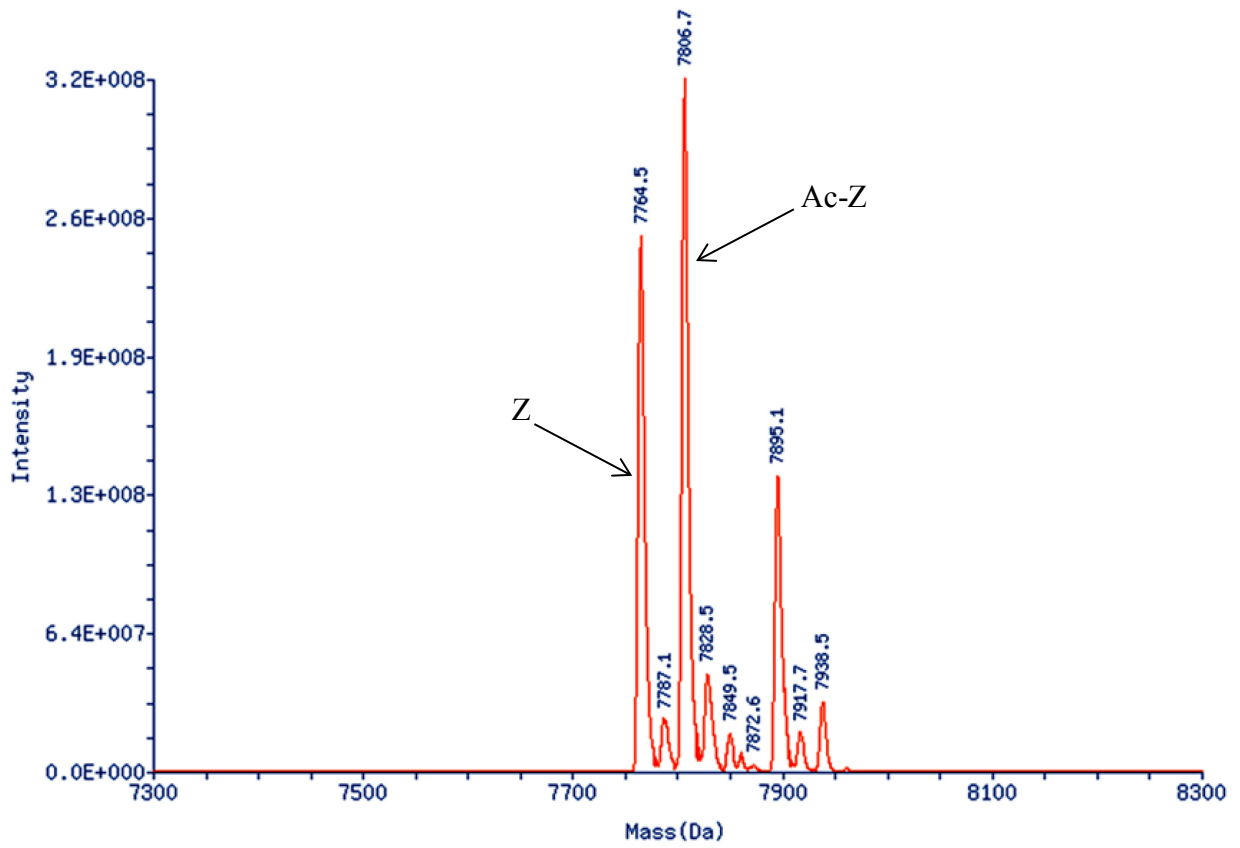


Figure 2.24 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in JM109(DE3) cells co-transformed with pET-Z and pACYCDuet-RimJ.

When grown in the PA-5052 medium containing 0.2% arabinose, DH10B *E. coli* cells transformed with pBAD-Z mainly produced the Z-form (Table 2.3, entry 13; Figure 2.25). Likewise, when the protein was expressed in DH10B cells co-transformed with pBAD-Z and pACYCDuet, the Z-form was mainly obtained (Table 2.3, entry 16; Figure 2.26). Minor N^α-acetylation was observed for the protein expressed from pBAD-Z when pSup-JYRS-6TRN was added in the DH10B cells (Table 2.3, entry 14; Figure 2.27). When pBAD-Z and pACYCDuet-RimJ were co-expressed in DH10B cells, the Z-form was obtained as the major product along with the Ac-Z as a minor product (Table 2.3, entry 15; Figure 2.28).

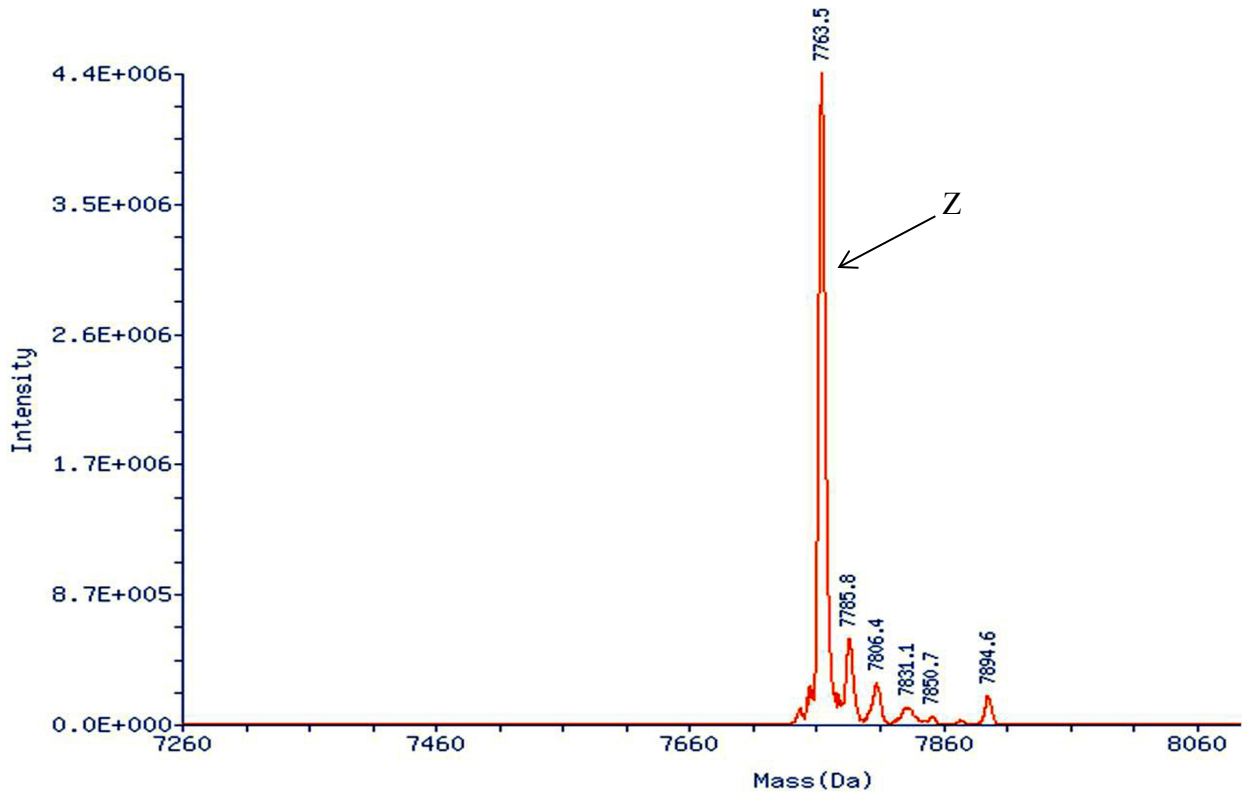


Figure 2.25 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells transformed with pBAD-Z.

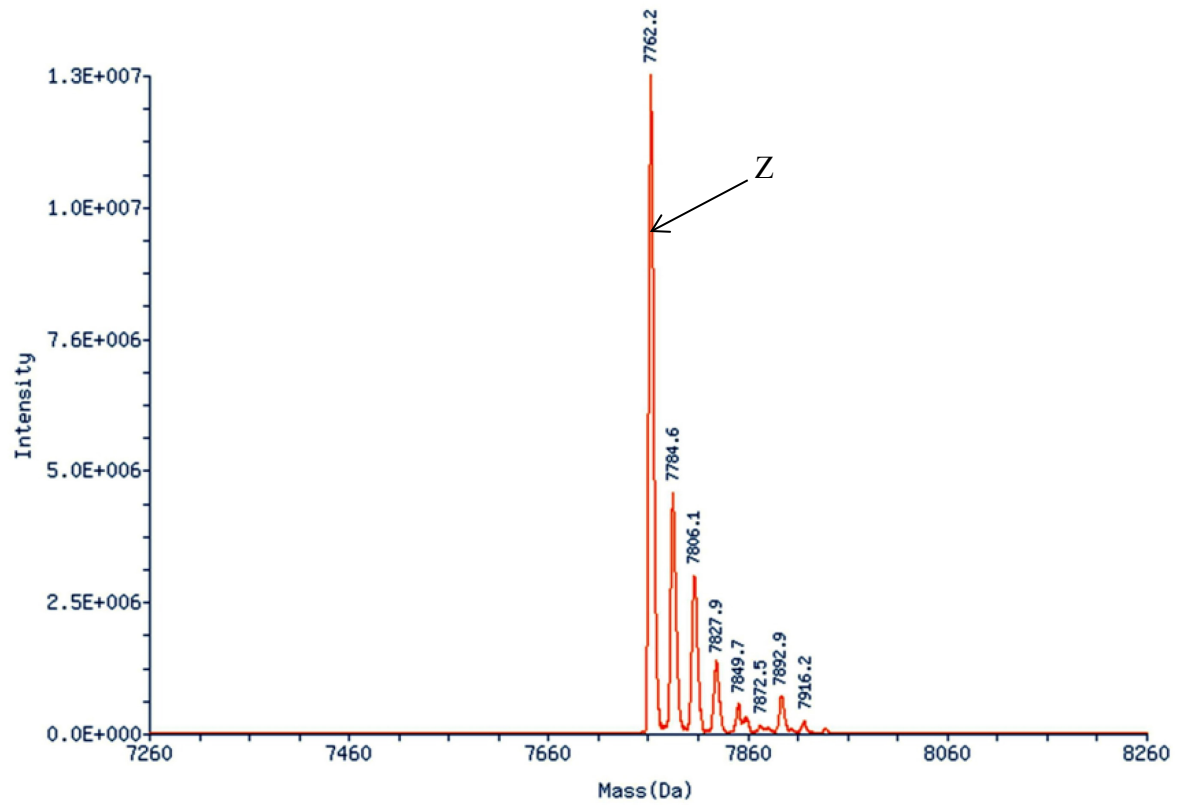


Figure 2.26 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells co-transformed with pBAD-Z and pACYCDuet.

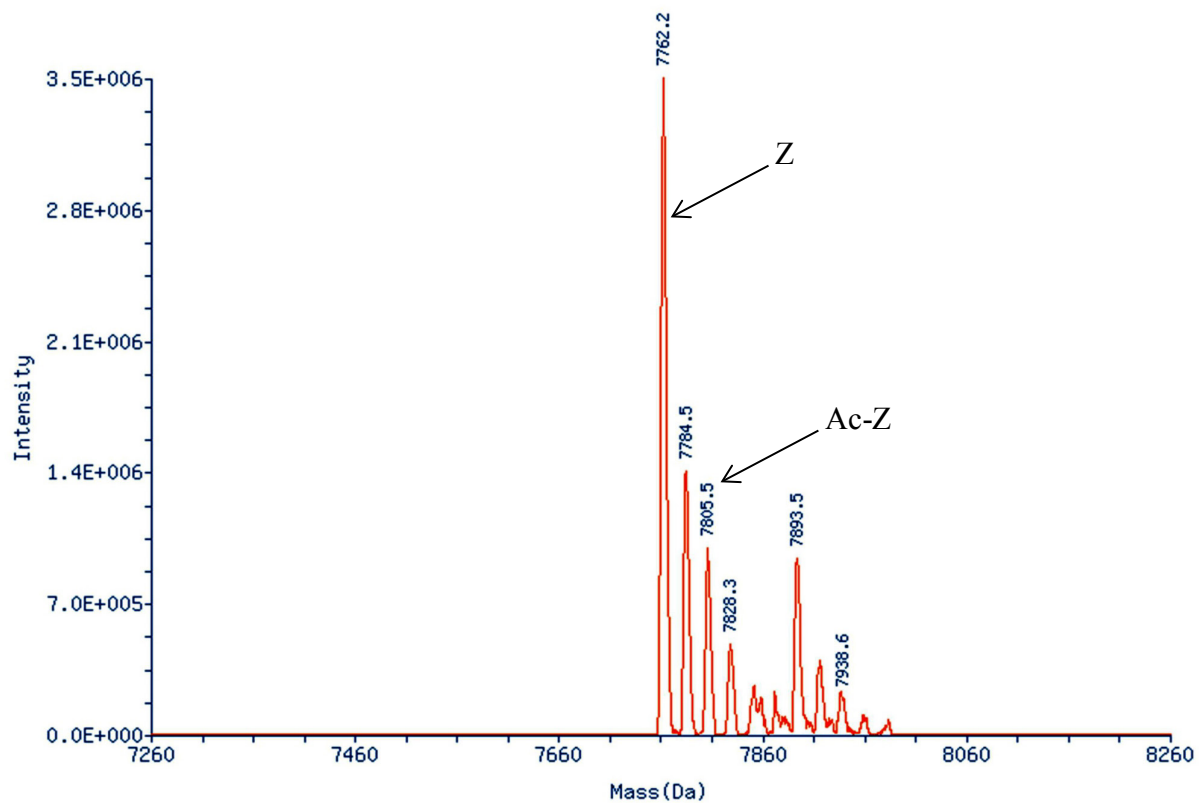


Figure 2.27 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells co-transformed with pBAD-Z and pSup-JYRS-6TRN.

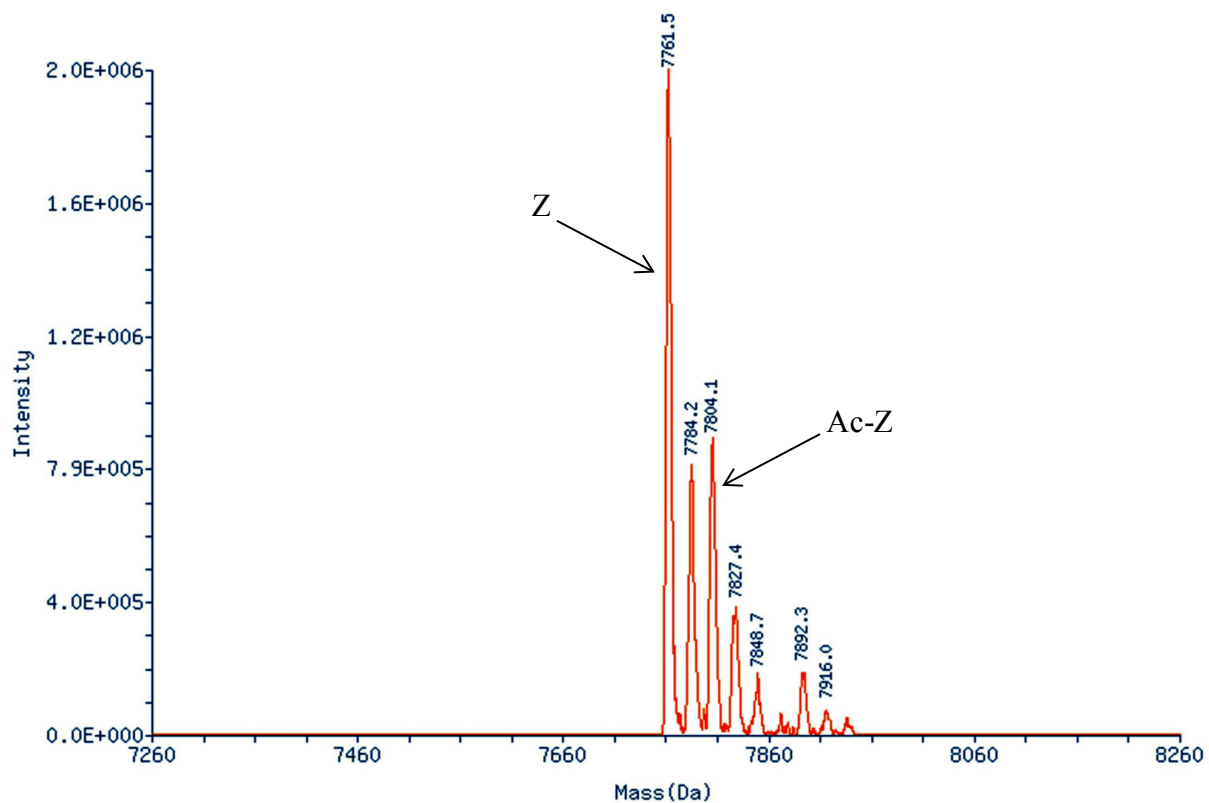


Figure 2.28 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in DH10B cells co-transformed with pBAD-Z and pACYCDuet-RimJ.

When the Z-domain protein was expressed from pBAD-Z and pACYCDuet-RimJ in BL21(DE3) cells, the Ac-Z form was mostly obtained (Table 2.3, entry 7; Figure 2.29). In order to investigate whether RimJ was directly involved in the N^α-acetylation of the protein, the chromosomal *rimJ* gene was deleted from the BL21(DE3) strain by the λ Red recombinase-based gene replacement method to generate the BL21(DE3) Δ *rimJ*::*kan* strain. In the *rimJ* knockout strain, a substantial decrease in the Ac-Z form was evident when the Z-domain was expressed in the *rimJ* knockout strain transformed with pBAD-Z and pACYCDuet-RimJ (Table 2.3, entry 8; Figure 2.30). A similar result was obtained when pET-Z and pACYCDuet-RimJ were used in BL21(DE3) Δ *rimJ*::*kan* strain (Table 2.3, entry 5; Figure 2.31).

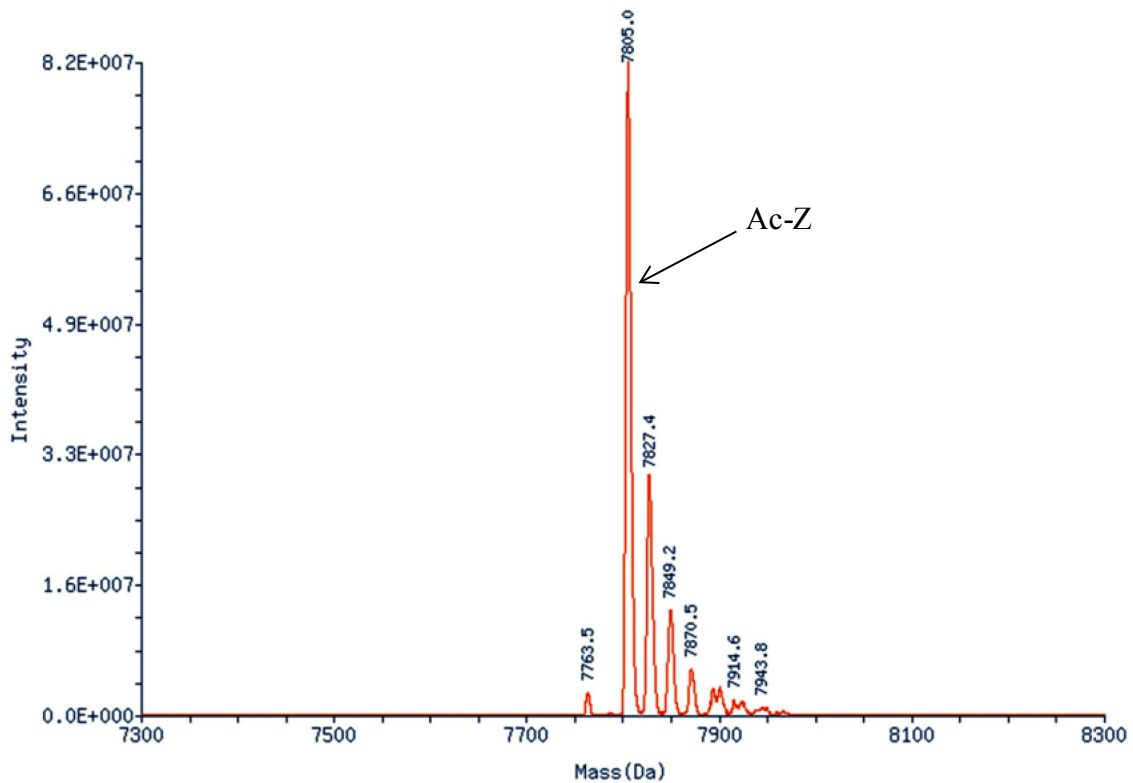


Figure 2.29 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) cells co-transformed with pBAD-Z and pACYCDuet-RimJ.

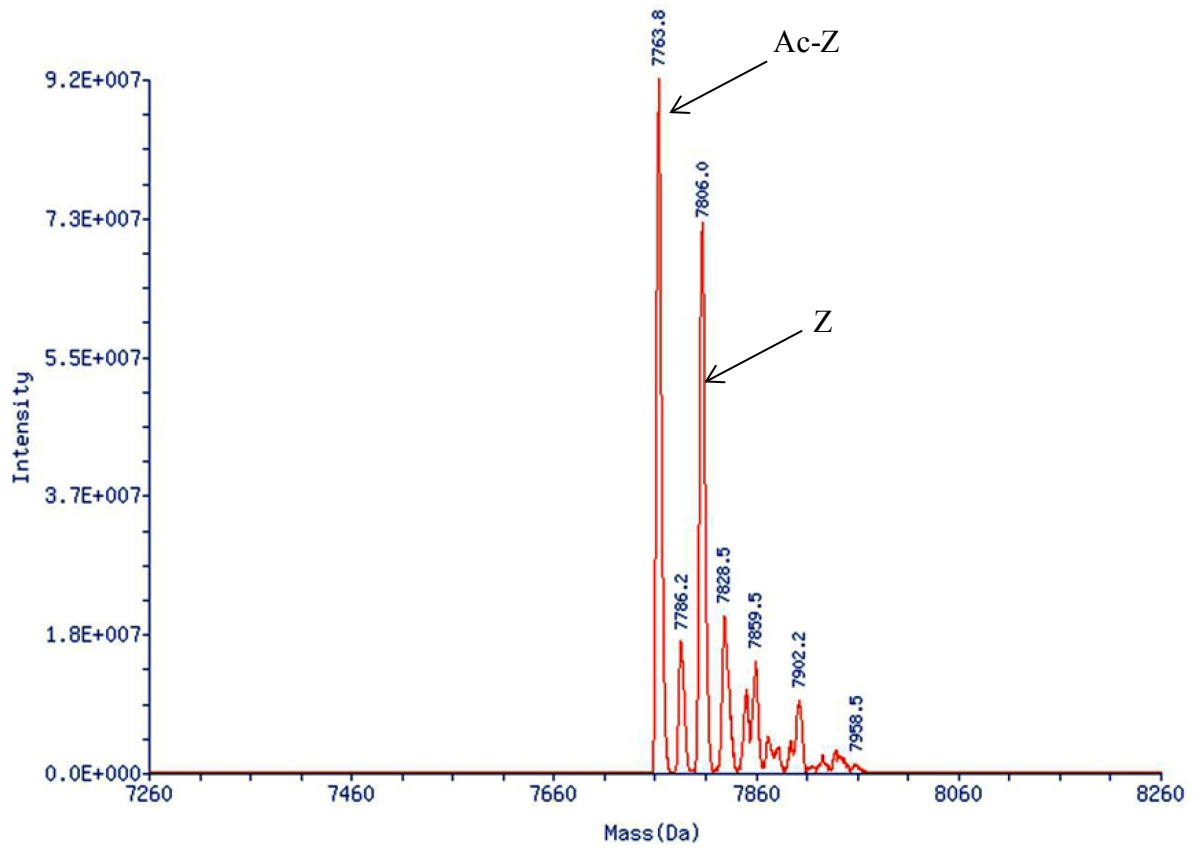


Figure 2.30 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) $\Delta rimJ::kan$ cells co-transformed with pBAD-Z and pACYCDuet-RimJ.

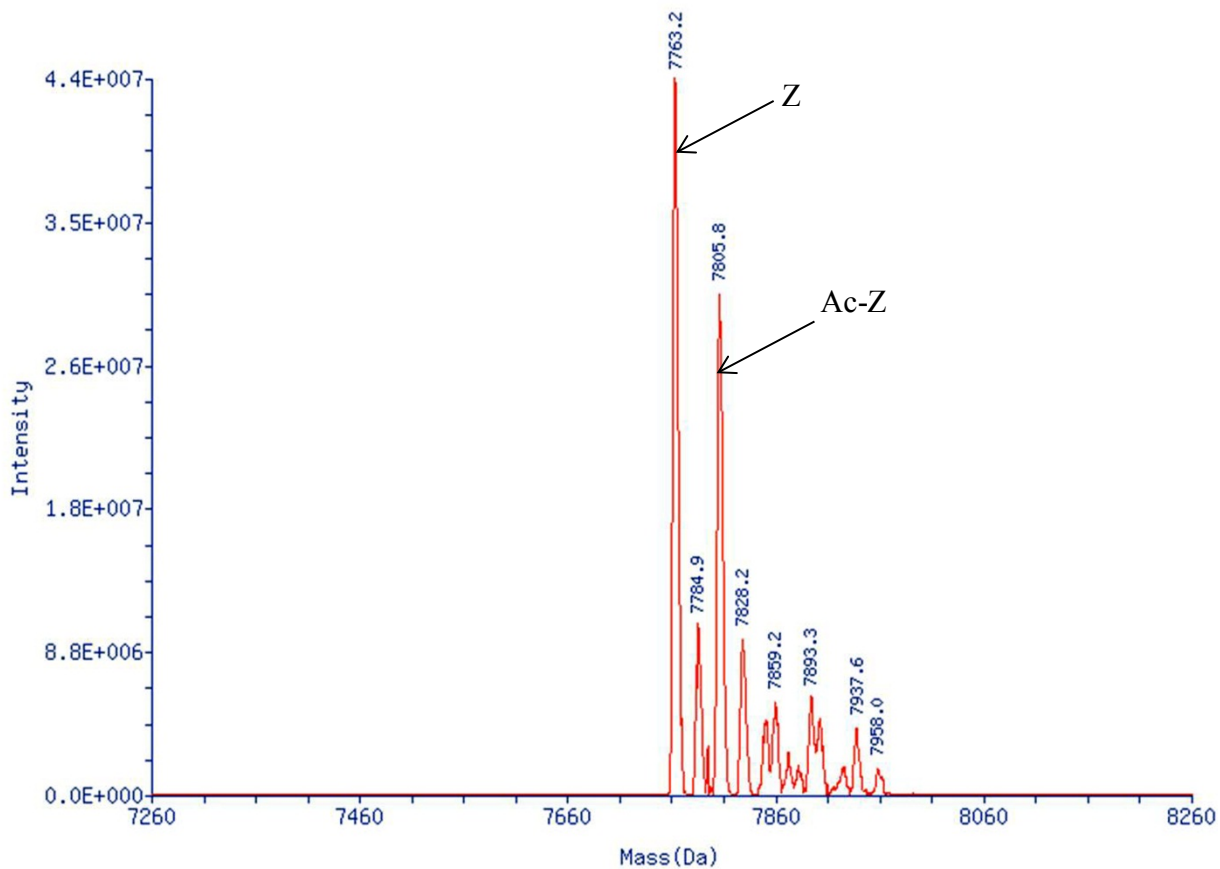


Figure 2.31 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) $\Delta rimJ::Kan$ cells co-transformed with pET-Z and pACYCDuet-RimJ.

Our results clearly indicate that a high cellular level of RimJ enhances N^α-acetylation of the Z-domain. Although all *E. coli* strains described thus far have an intact chromosomal copy of the *rimJ* gene, a constitutively expressed housekeeping gene that catalyzes N^α-acetylation of the S5 ribosomal protein, little acetylation of the Z-domain was observed from any single expression plasmid in these strains themselves. Possibly, the basal levels of RimJ are not enough for the efficient acetylation of the Z-domain, although the Tα1 fusion protein was at least partially acetylated by endogenous levels of RimJ in JM109(DE3) cells.

We also investigated whether RimJ is involved in the pSup-mediated N^α-acetylation of the Z-domain. When Z-domain was expressed in BL21(DE3) Δ *rimJ::kan* co-transformed with pET-Z and pSup-JYRS-6TRN, the Z form was mainly observed (Table 2.3, entry 3; Figure 2.32), whereas its precursor strain with the intact *rimJ* gene provided a substantial amount of N^α-acetylated Z-domain under the same conditions. These results suggest that the presence of pSup-JYRS-6TRN may induce a high expression of RimJ, which in turn enhances N^α-acetylation of the Z-domain. However, the exact mechanism is subject to a further study.

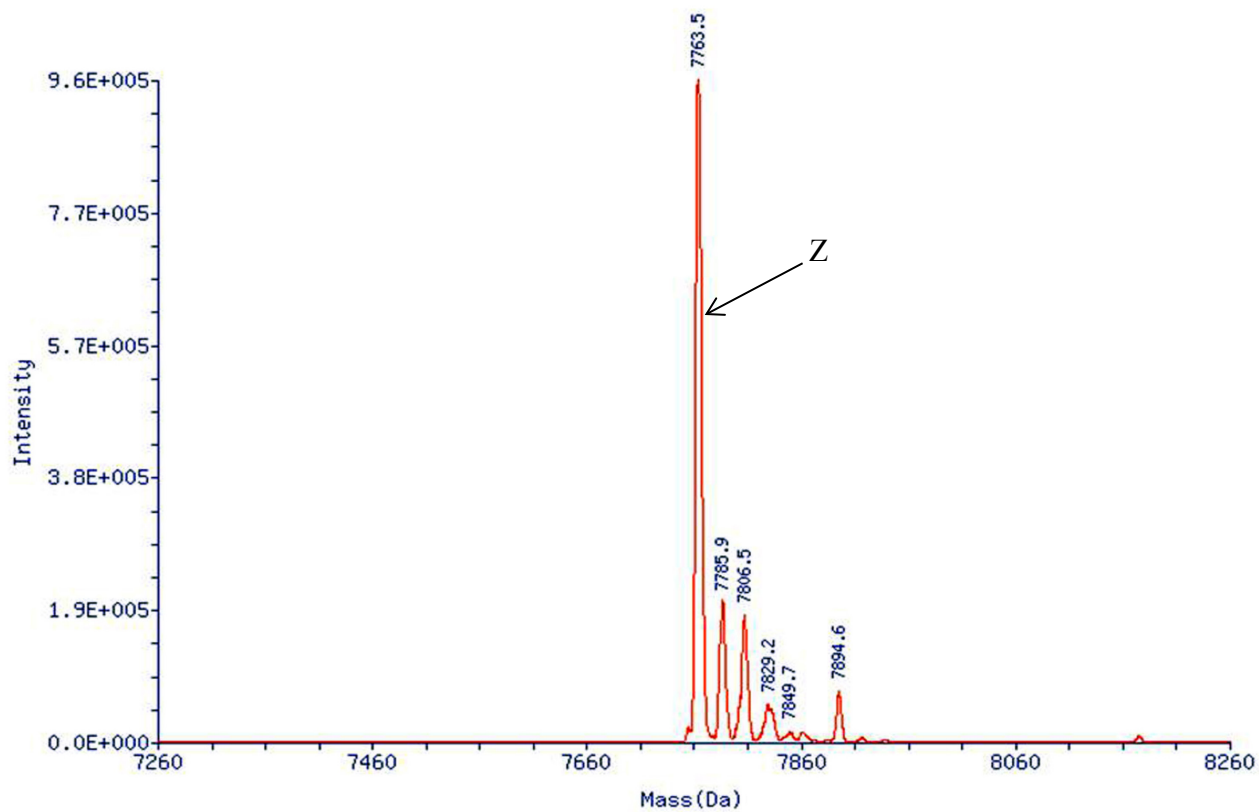


Figure 2.32 Mass-transformed ESI-MS spectrum of the Z-domain protein that was expressed in BL21(DE3) $\Delta rimJ::Kan$ cells co-transformed with pET-Z and pSup-JYRS-6TRN.

We unambiguously determined that the Z-domain is acetylated at its N-terminus by proteolysis and subsequent ESI-MS analysis of the resulting peptide fragments. Since resistant to trypsin digestion under standard aqueous conditions, the Z-domain was incubated with trypsin in 50 mM ammonium bicarbonate buffer containing 5% 2,2,2-trifluoroethanol for 20 hours. High resolution ESI-MS analysis of the Z-domain expressed in BL21(DE3) *E. coli* cells co-transformed with pET-Z and pSup-JYRS-6TRN clearly showed the mass peaks at 663.3302 Da and 705.3402 Da, which correspond to the mono protonated forms of the unacetylated N-terminal fragment (TSVDNK, calculated monoisotopic mass 663.3308 Da) and the acetylated N-terminal fragment (Ac-TSVDNK, calculated monoisotopic mass 705.3414 Da), respectively (Figure 2.33). The tryptic digest of the protein expressed in BL21(DE3) *E. coli* cells co-transformed with pET-Z only showed the unacetylated N-terminal peptide fragment in the ESI-MS (Figure 2.34). The ESI-MS analysis of the tryptic fragments of Z-domain expressed in BL21(DE3) *E. coli* cells co-transformed pBAD-Z and pACYCDuet-RimJ mainly showed the acetylated N-terminal peptide fragment along with a small amount of the unacetylated (Figure 2.35).

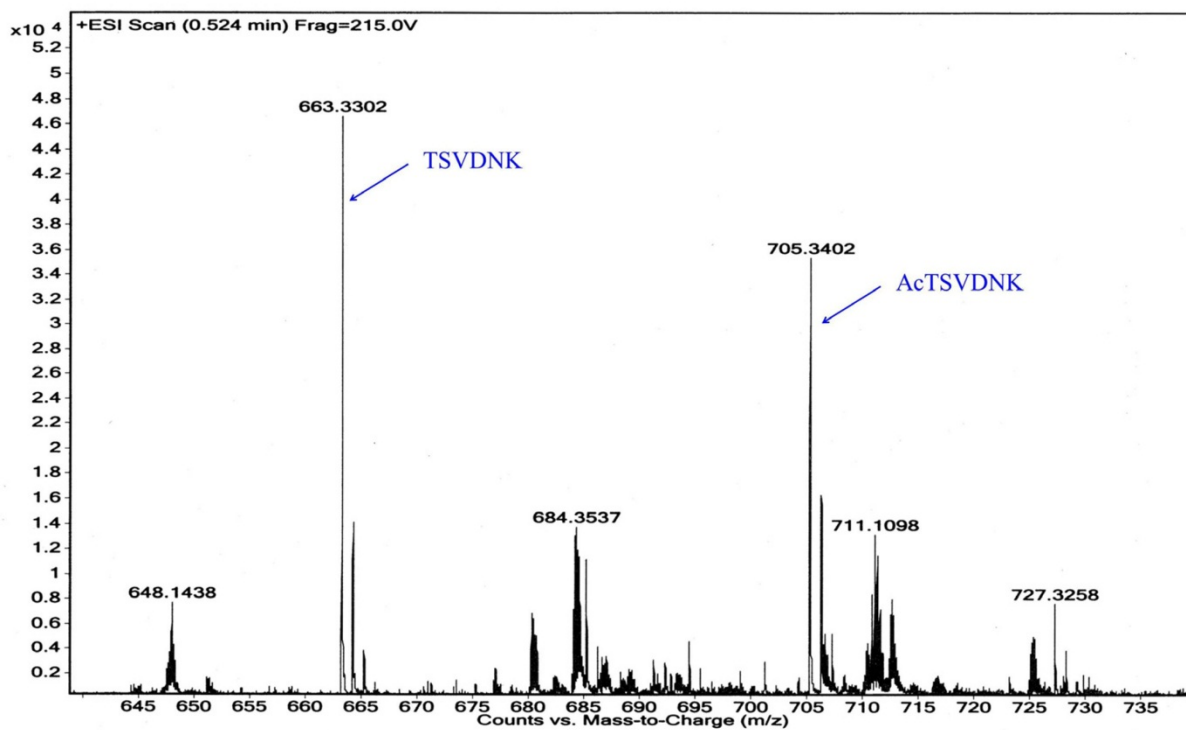


Figure 2.33 High Resolution ESI-MS analysis of the N-terminal peptide fragments obtained by the trypsin proteolysis of the Z-domain, which was expressed in BL21(DE3) *E. coli* cells in the presence of pET-Z and pSup-JYRS-6TRN.

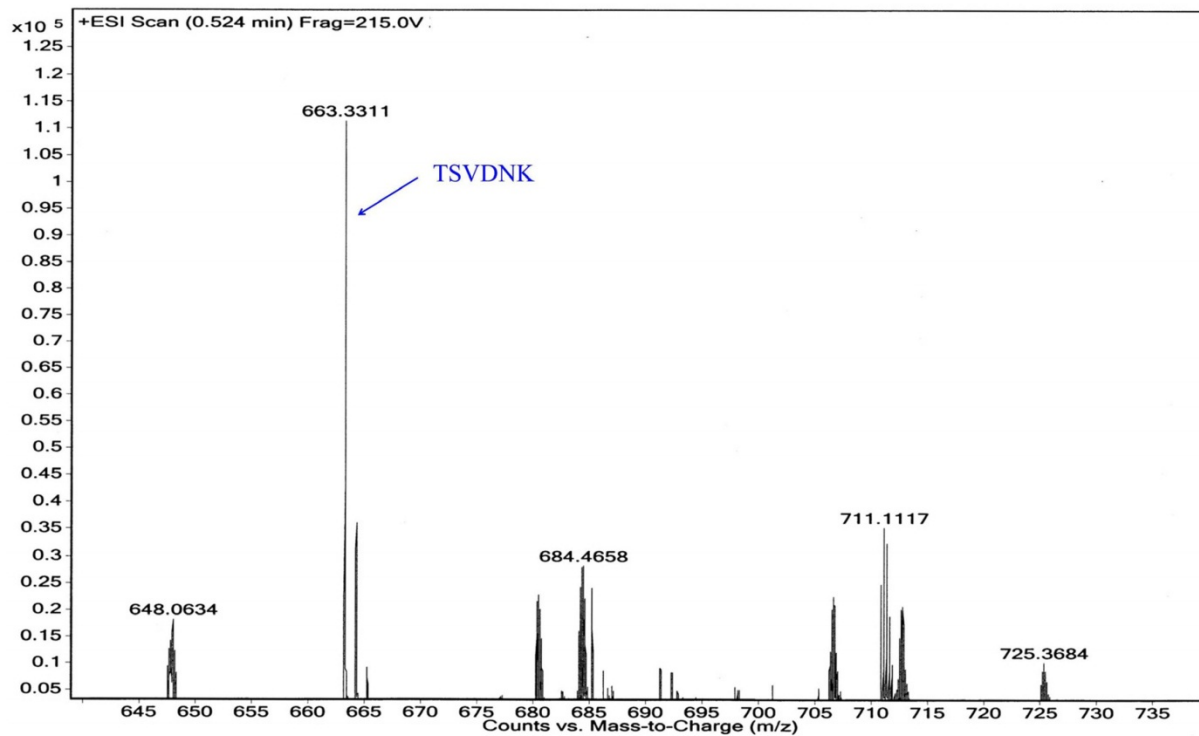


Figure 2.34 High Resolution ESI-MS analysis of the N-terminal peptide fragments obtained by the trypsin proteolysis of the Z-domain, which was expressed in BL21(DE3) *E. coli* cells in the presence of pET-Z.

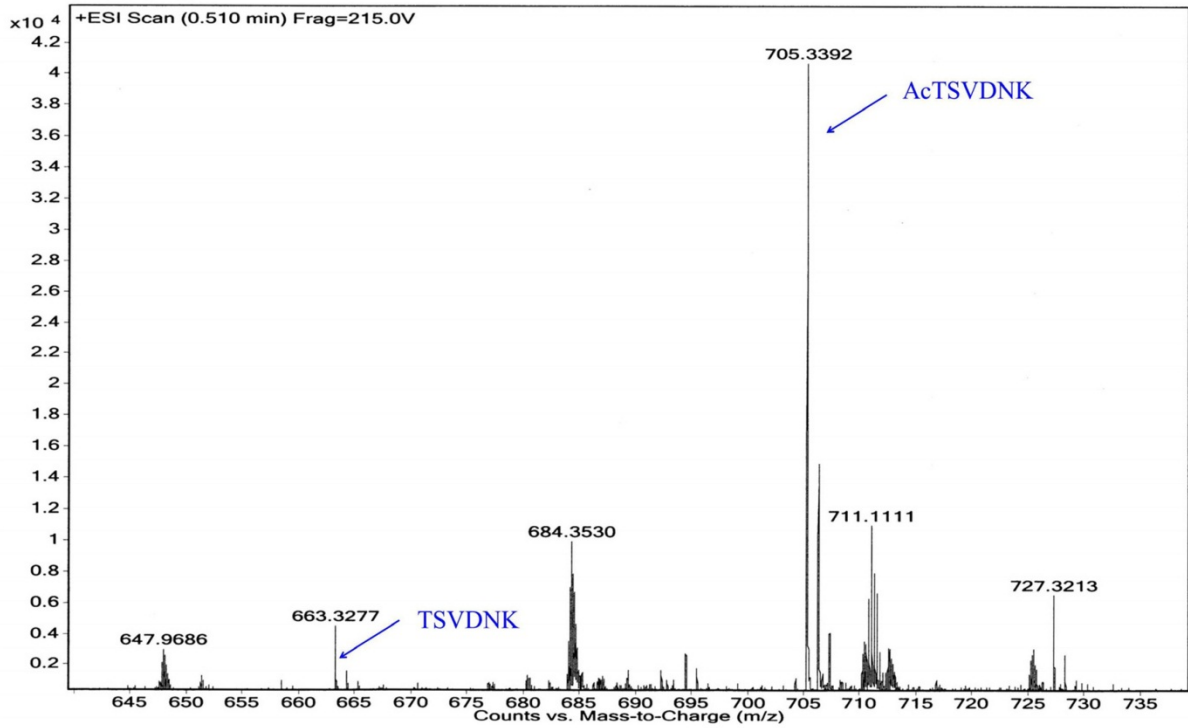


Figure 2.35 High Resolution ESI-MS analysis of the N-terminal peptide fragments obtained by the trypsin proteolysis of the Z-domain, which was expressed in BL21(DE3) *E. coli* cells in the presence of pBAD-Z and pACYCDuet-RimJ.

2.3 Conclusions

We observed that the recombinant Z-domain protein, which is not normally acetylated in standard expression conditions, can be fully N^α-acetylated by highly expressed RimJ in *E. coli*. The N^α-acetylation process also depends on the *E. coli* strains, expression plasmids and amino acid residues near the N-terminus. Although this study was focused on the Z-domain protein, it is certainly worth trying to investigate other recombinant proteins that could possibly be N-terminal acetylated under the same conditions as the Z-domain. In this area we have already studied myoglobin since it was readily available in the lab. When myoglobin was expressed along RimJ, the protein obtained was unacetylated with the initiating methionine intact. This outcome might be possibly due to the influence of the N-terminal sequence favoring unacetylation. Some future studies might include the introduction of different mutations at the N-terminal site of this protein to analyze if the sequence and the overexpression of RimJ would favor the N-terminal acetylation of this protein.

Some other future studies that the Ryu laboratory is interested in include active site mapping experiments with the Bpa mutants *in vitro* to analyze the binding specificity of RimJ, since no structural information of this NAT is available. Also, classical experiments introducing different C-terminal deletions of the Z-domain by site-specific mutagenesis to analyze the RimJ selectivity differences caused by the different mutations in the substrate.

2.4 Experimental section

2.4.1 General.

XL1-Blue *E. coli* cells (Stratagene) were used for cloning and maintaining plasmids. XL1-BLUE, BL21(DE3), JM109(DE3), JM109, AD494(DE3), K12(WT), B(WT), MG1655 (Seq), AVB100, JW4335, JW1053, JW1423, JW2293, JW2294, JW4030, LCB90 and DH10B *E. coli* cells were used for protein expression. *PfuUltra* High-Fidelity DNA polymerase (Stratagene) was used for polymerase chain reaction (PCR). 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. DNA and protein concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

2.4.2 Construction of plasmids

pET-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). pBAD-Z, a plasmid that encodes the Z-domain C-terminal hexa-histidine fusion gene under control of *araBAD* promoter, was constructed by cloning the fusion gene between the NcoI and HindIII sites of pBAD/*myc*-His (Invitrogen). pBAD-Z (Biotin) was similarly constructed by including the biotinylated tag downstream of the Z-domain gene. The *rimJ* gene was amplified from MG1655 *E. coli* genomic DNA by PCR. pACYCDuet-RimJ, a p15A plasmid that encodes the *rimJ* gene under control of T7 promoter, was generated by inserting the PCR-amplified *rimJ* gene between the NdeI and XhoI sites of pACYCDuet-I (Novagen). All plasmids constructed were verified by sequencing.

2.4.3 Protein expression and purification

Different *E. coli* strains were transformed with the plasmids shown in Table 2.4. The transformed cells were grown overnight at 37°C in the PA-5052 auto-induction medium containing 100 µg/mL carbenicillin (and 50 g/mL chloramphenicol for the double transformants). Cells were harvested by centrifugation and lysed by sonication in lysis buffer (50 mM 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 native conditions according to the manufacturer's protocol. Purified protein was concentrated by ultrafiltration and analyzed by SDS-PAGE using Coomassie® G-250 stain (Invitrogen, Inc) or BioRad Silver Stain Plus kit for protein visualization.

Table 2.4 *E. coli* strains and plasmids used for the expression of the Z-domain proteins.

E. coli strains	Plasmids
AVB100 JW4335 JW1053 JW1423 JW2293 JW2294 JW4030 LCB90	pBAD-Z(Biotin)
XL1-BLUE JM109 K12(WT) B(WT) MG1655 BL21(DE3) JM109(DE3) DH10B	pBAD-Z
AD494(DE3) BL21(DE3) JM109(DE3)	pET-Z
XL1-BLUE JM109	pBAD-Z+pREP2-YC
BL21(DE3) JM109(DE3) DH10B	pET-Z+pSup-JYRS-6TRN or pBAD-Z+pSup-JYRS-6TRN
BL21(DE3) JM109(DE3) DH10B	pET-Z+pACYCDuet-RimJ or pBAD-Z+ pACYCDuet-RimJ
BL21(DE3) JM109(DE3) DH10B	pET-Z+pACYCDuet or pBAD-Z+pACYCDuet

2.4.4 Expression of the Z-domain Bpa mutants

The Z-domain Bpa mutants were produced as previously reported.⁸⁶ The codons for Ser-3, Val-4, and Asp-5 of the Z-domain in pET-Z were mutated to an amber (TAG) codon by site directed mutagenesis to generate pET-Z(S3TAG), pET-Z(V4TAG) and pET-Z(D5TAG), respectively. These plasmids were co-transformed with pSup-BpaRS-6TRN or pSup-JYRS-6TRN (Negative control) into BL21(DE3) cells. The transformed cells were grown in the presence of 1mM Bpa overnight at 37°C in the PA-5052 medium containing 100 µg/mL carbenicillin and 50 g/mL

chloramphenicol. The mutant proteins were isolated and purified as described above for the wild type protein.

2.4.5 Photochemical cross-linking experiment

The *E.coli* cells containing the Z-domain Bpa/Tyr mutants were irradiated after 1 hour of protein induction using a high power UV light (365nm, 450 W) for 20, 30, 45 or 60 minutes. The purified proteins were run on a polyacrylamide gel (Thermo Scientific). The gel was stained using BioRad Silver Stain Plus kit. The cross-linked products were analyzed by the MS/MS.

2.4.6 Deletion of the rimJ gene from the BL21(DE3) E. coli strain

The *rimJ* gene was deleted from the chromosome of BL21(DE3) *E. coli* cells according to the method of Datsenko and Wanner.⁸⁷ The kanamycin resistance cassette of pKD13 was amplified by PCR using two oligonucleotides: 5'-ATG TTT GGC TAT CGC AGT AAC GTG CCA AAA GTG CGC TTA ACC ACA GAC CGG TGT AGG CTG GAG CTG CTT C-3' and 5'-TTA GCG GCC GGG CGT CCA GTC TGG GGT AGT TAA TGC CGT CAG TAC ATG GGA ATT AGC CAT GG TCC-3'. The PCR product was transformed into electrocompetent BL21(DE3) *E. coli* cells containing pKD46, a plasmid encoding red recombinase. The transformants were incubated at 30°C for 12 h and plated on a large Luria Bertani (LB) agar plate containing kanamycin. The plate was incubated at 43°C for 24 h. Single colonies were picked and amplified in LB media. The genomic DNA was isolated and amplified by PCR using two oligonucleotides: 5'-ATG TTT GGC TAT CGC AGT AA CG-3' and 5'- TTA GCG GCC GGG CGT CCA GTC-3'. The successful integration of the knockout cassette was confirmed by examining the size of the PCR product.

2.4.7 Mass spectrometry

Electrospray ionization mass spectrometry was performed on quadrupole (LCQ) or linear ion trap (LTQ) mass spectrometers (both from Thermo Scientific, San Jose, CA) equipped with an electrospray ionization (ESI) source and operated with the Xcalibur data acquisition software. ESI spray voltage and capillary temperature were maintained at 2.5-3.0 kV and 200 °C, respectively. Protein solutions of 10 to 50 ng/ μ L concentration were sprayed at an infusion rate of 5 μ L/min. At least ten scans covering mass-to-charge (m/z) ratios of 200 to 2000 Th were averaged to obtain ESI mass spectra from which multiple-charged protein ions were deconvoluted by using the ProMass for XCalibur program (Novatia, Monmouth Junction, NJ). Peak heights in the resultant graphs were considered proportional to the abundance of the proteins with the corresponding molecular mass (m) present in the samples. High resolution ESI mass spectra, obtained by using the Fourier-transform ion cyclotron resonance (FTICR) operating in tandem with the LTQ linear ion trap (LTQ-FT, Thermo), were used to confirm deconvolution by ProMass. FTICR full-scan mass spectra were acquired at 100,000 nominal mass resolving power ($M/\Delta M$ at m/z 400 and taking the full width at half maximum intensity, FWHM, as ΔM) from m/z 400 to 2000 Th using the automatic gain control mode of ion trapping.

2.4.8 Theoretical mass calculations

The theoretical mass of the Z-domain was calculated based on its amino acid sequence:

MTSVDNKINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAKKLNDAQA
PKGSHHHHH. The calculated average mass of the Z-domain with the cleavage of the N-terminal methionine (the Z form) is 7762.4 Da. The calculated average mass of the N ^{α} -acetylated Z-domain with the removal of the N-terminal methionine (the Ac-Z form) is 7804.5 Da. The theoretical mass of the Z-domain including the biotin tag was calculated based on its amino acid

sequence: MTSVDNKINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAK KLNDAQAPKGSKLGLNDIFEAQKIEWHEVDHHHHHH. The calculated average mass of the Z-domain (Biotin) with the cleavage of the N-terminal methionine is 10029.0 Da. The calculated average mass of the N^α-acetylated Z-domain (Biotin) with the removal of the N-terminal methionine is 10071.0 Da. The molecular weight of Bpa is 269.3. The theoretical mass of the Z-domain Bpa mutants were calculated based on the molecular weight difference between the wild type amino acid residue and Bpa at each position as shown in Table 2.5.

Table 2.5 Molecular weight difference between the wild type amino acid residue and Bpa.

The Z-domain mutant	Molecular weight difference	Calculated average mass of the Z-form (Da)	Calculated average mass of the Ac-Z form (Da)
Ser3Bpa	+164.2	7926.6	7968.7
Val4Bpa	+152.1	7914.5	7956.6
Asp5Bpa	+136.2	7898.6	7940.7

2.4.9 Trypsin proteolysis and high resolution ESI-MS analysis

A 120 μ L solution of the isolated protein (20 μ g) in 50 mM ammonium bicarbonate buffer (pH 8.5) containing 50% 2,2,2-trifluoroethanol was heated at 60°C for 1 h, cooled to room temperature, and diluted 10-fold with 50 mM ammonium bicarbonate buffer. Digestion was performed at 37°C for 20 h in the presence of 1 μ g trypsin. The reaction mixture was evaporated to dryness and re-dissolved in 700 μ L 0.1% aqueous acetic acid. A 3 μ L aliquot of the digested sample was directly injected to Agilent 6224 Accurate-Mass TOF LC/MS system and analyzed by MassHunter Workstation software.

CHAPTER 3. SEQUENCE REQUIREMENTS ON THE CLEAVAGE OF INITIATING METHIONINE & THE N^α-ACETYLATION OF THE Z DOMAIN IN ESCHERICHIA COLI.

3.1 Introduction

As described in Chapter 2, the Z-domain protein is N^α-acetylated in the presence of highly expressed Rim J.⁸⁸ The N^α-acetylation was observed for the Ser3Bpa mutant but not for the Val4Bpa or the Asp5Bpa mutants. These results indicated that the amino acid residues near the N-terminus influenced N^α-acetylation of the Z-domain as reported in the literature.^{5,32,35,53} In eukaryotes, no simple consensus motifs were found for protein N^α-acetylation, although proteins with small amino acid residues in the second position of the protein are most frequently acetylated after the removal of the initiating methionine residue.⁵ A similar pattern was observed for a few N^α-acetylated endogeneous and ectopic proteins in *E. coli* as described in Chapter 1. However, the amino acid requirements of the protein N^α-acetylation had never been systematically investigated. Towards a better understanding of the sequence requirements for the N^α- acetylation of proteins, we decided to study the Z-domain acetylation pattern when each of the amino acid residues at the first few positions was substituted to the other 19 natural amino acids. Establishing the sequence requirements of the Z-domain would provide a molecular basis to predict or control N^α-acetylation of recombinant proteins in *E. coli*.

3.2 Results and discussion

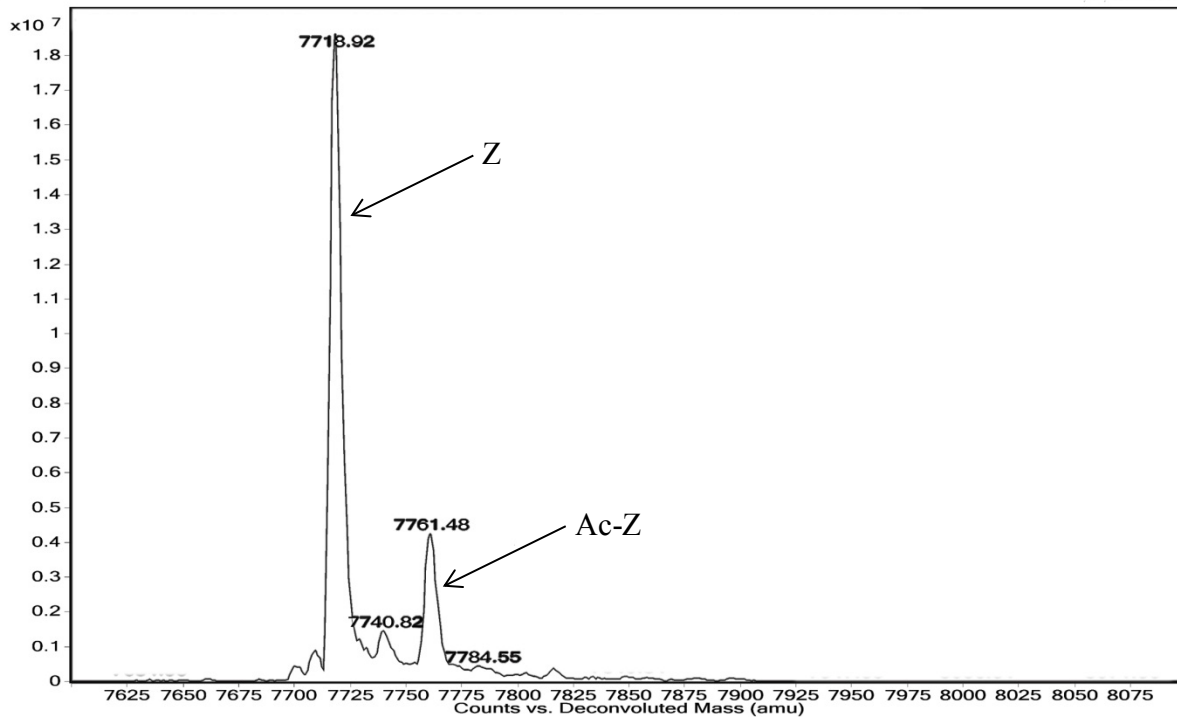
We used the Z-domain as a model protein to establish the sequence dependence of the Met₁ cleavage and RimJ-mediated N^α-acetylation in *E. coli*. The Z-domain variants differing by the second or third amino acid residue were expressed and analyzed by mass spectrometry, by which a 131-Da mass decrease and a 42-Da mass increment indicated Met₁ cleavage and N^α-acetylation, respectively.

In order to construct the genes for the Z-domain variants containing 20 different amino acids in positions 2 or 3, the corresponding codon in the pET-Z plasmid was replaced with the nucleotide sequence NNK where N is all four nucleotides and K is G or T. According to the standard genetic code, NNK covers all 20 amino acids. About a half of the possible 20 variants for each position were obtained by this initial site-directed mutagenesis. The remaining Z-domain gene variants containing less frequent genetic codes were obtained by replacing the corresponding codon with more targeted random nucleotide sequences (NAK and HTT for the second codon and VAA for the third codon where H is A, C or T; V is A, C, or G) or with the codon for the specific amino acid (i.e. TGG for Trp; ATG for Met).

Each Z-domain variant was individually expressed as a C-terminal His-tag fusion in BL21(DE3) *E. coli* cells, co-transformed with its corresponding pET-Z mutant plasmid and pACYCDuet-RimJ. In the presence of the latter plasmid from which RimJ is overexpressed, the Z-domain could be N^α-acetylated.⁸⁸ The transformed cells were grown in instant TB, an auto-induction medium in which the protein expression is induced by lactose as soon as glucose is deprived. The overexpressed Z-domain variants were purified from the cell lysate by IMAC. Each of the purified Z-domain variants differing by the second or third amino acid residue was analyzed by ESI-MS and subsequent deconvolution of multiple-charged protein ions as shown in

Figures. 3.1 and 3.2, respectively. For all variants, the observed masses were matched with the calculated values within less than one Da (Tables 3.1 and 3.2).

a)



b)

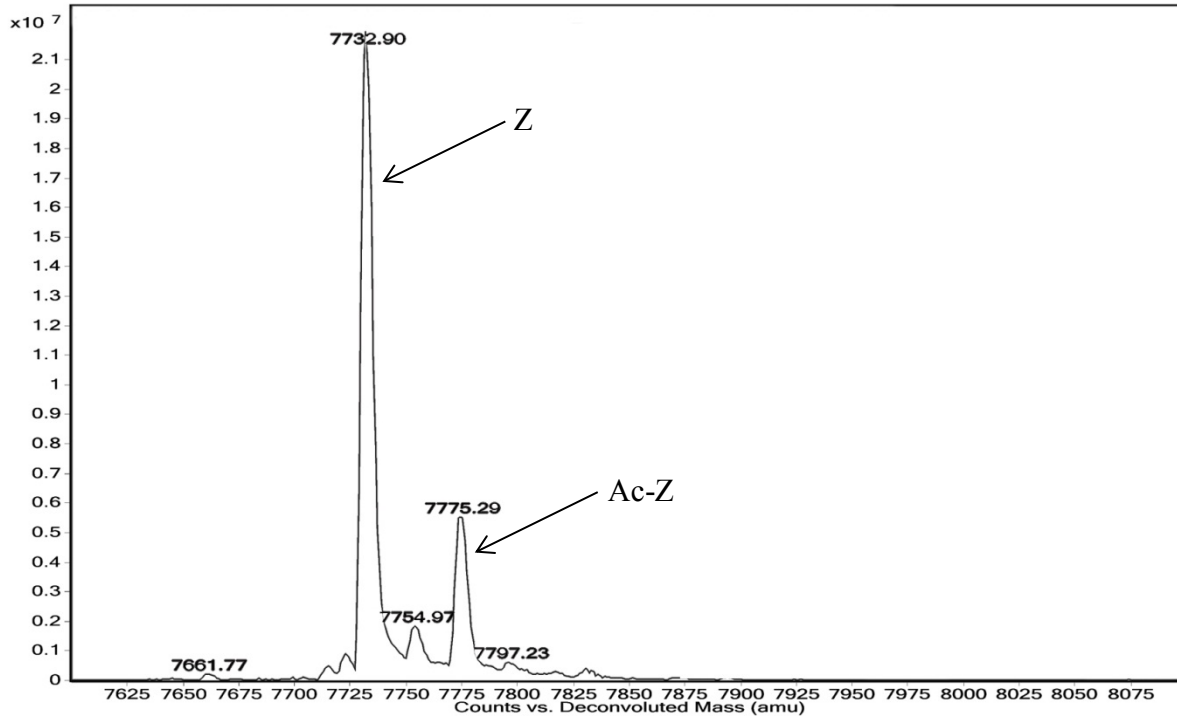
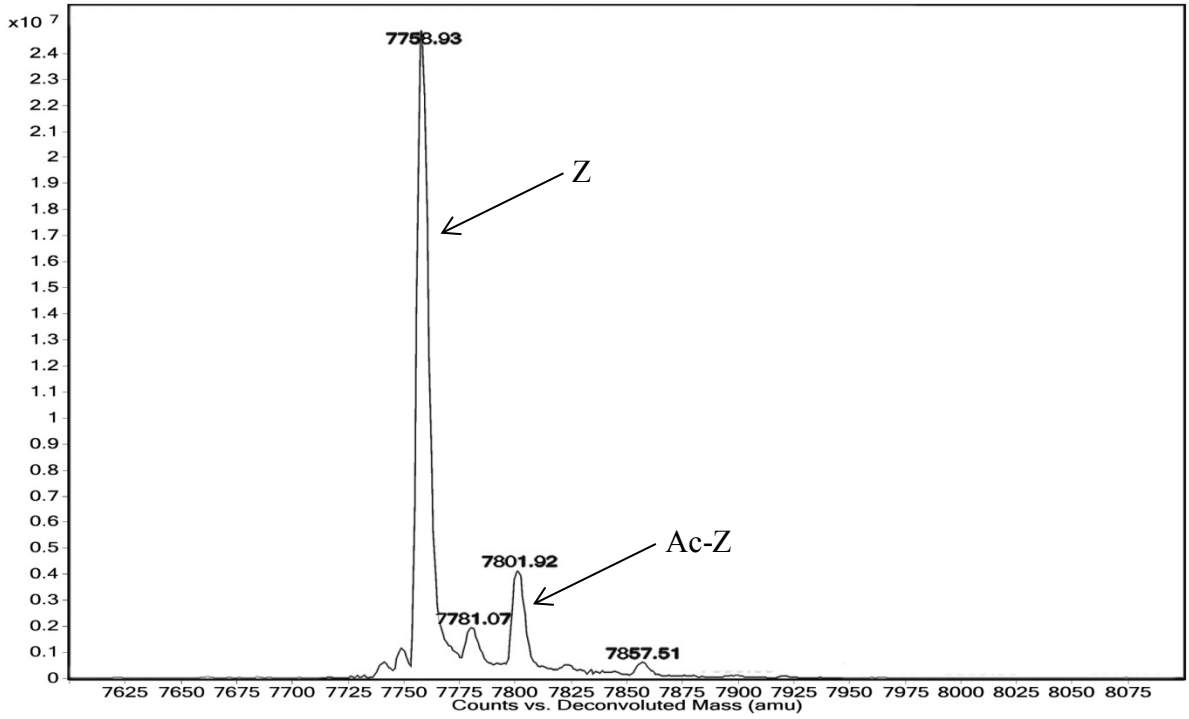


Figure 3.1 Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: a) glycine and b) alanine.

c)



d)

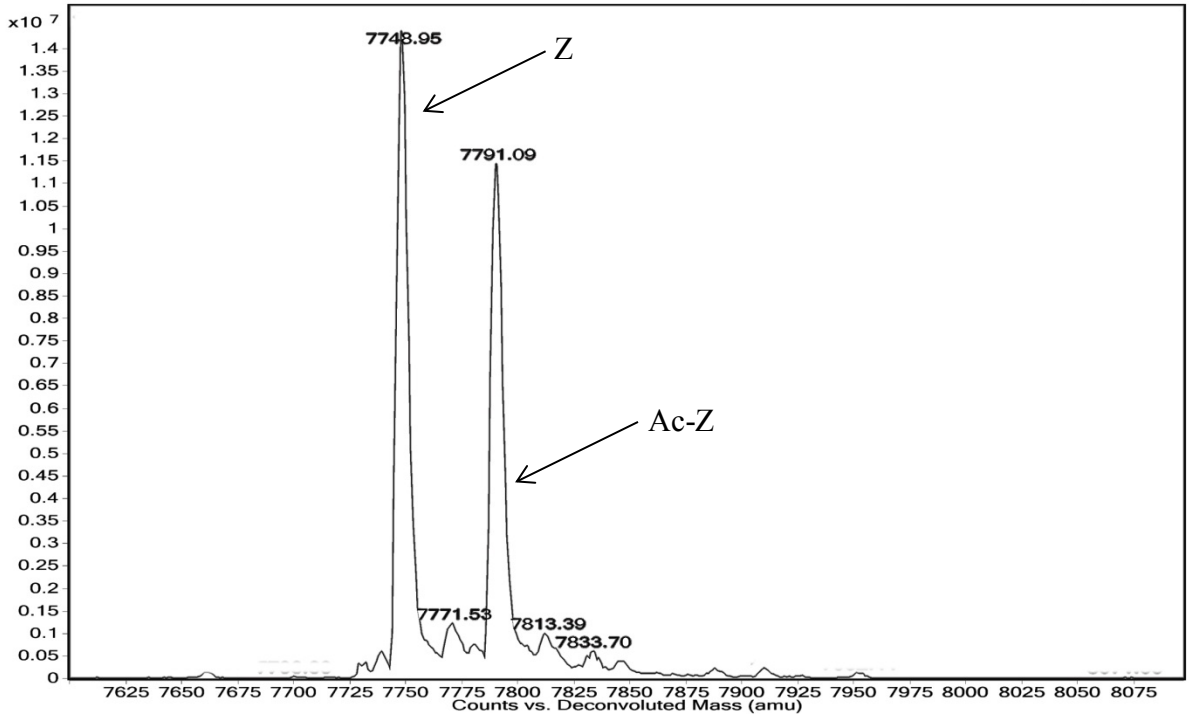
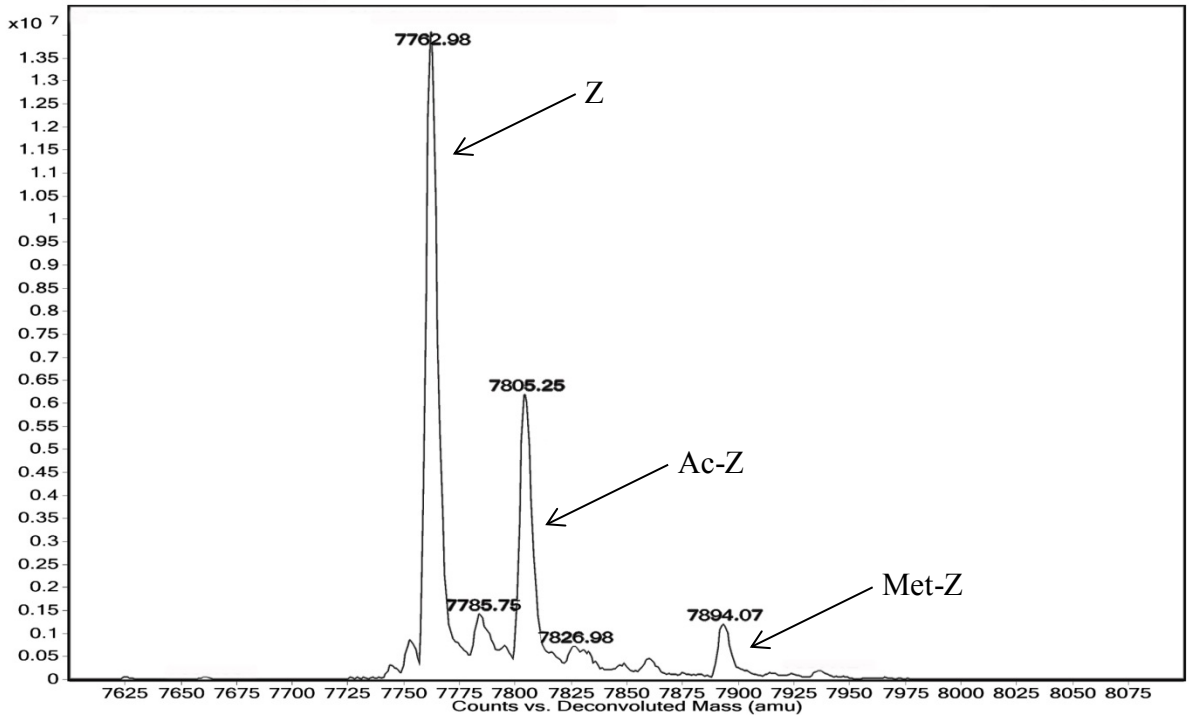


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: c) proline and d) serine.

e)



f)

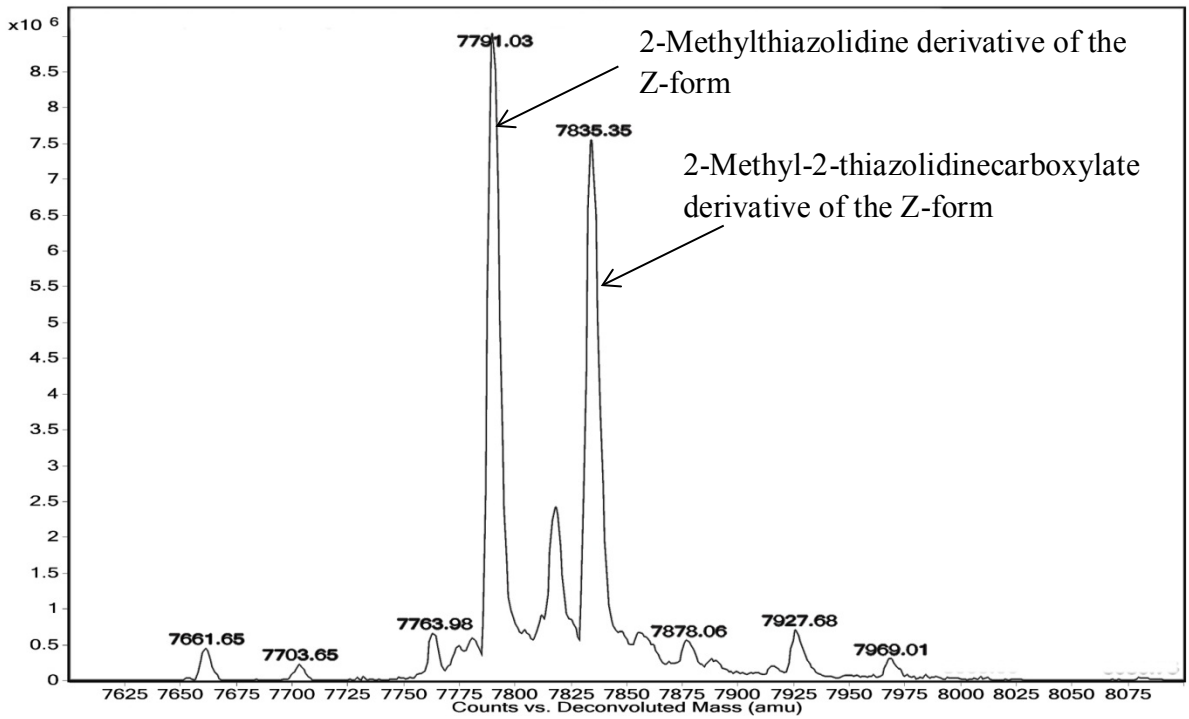
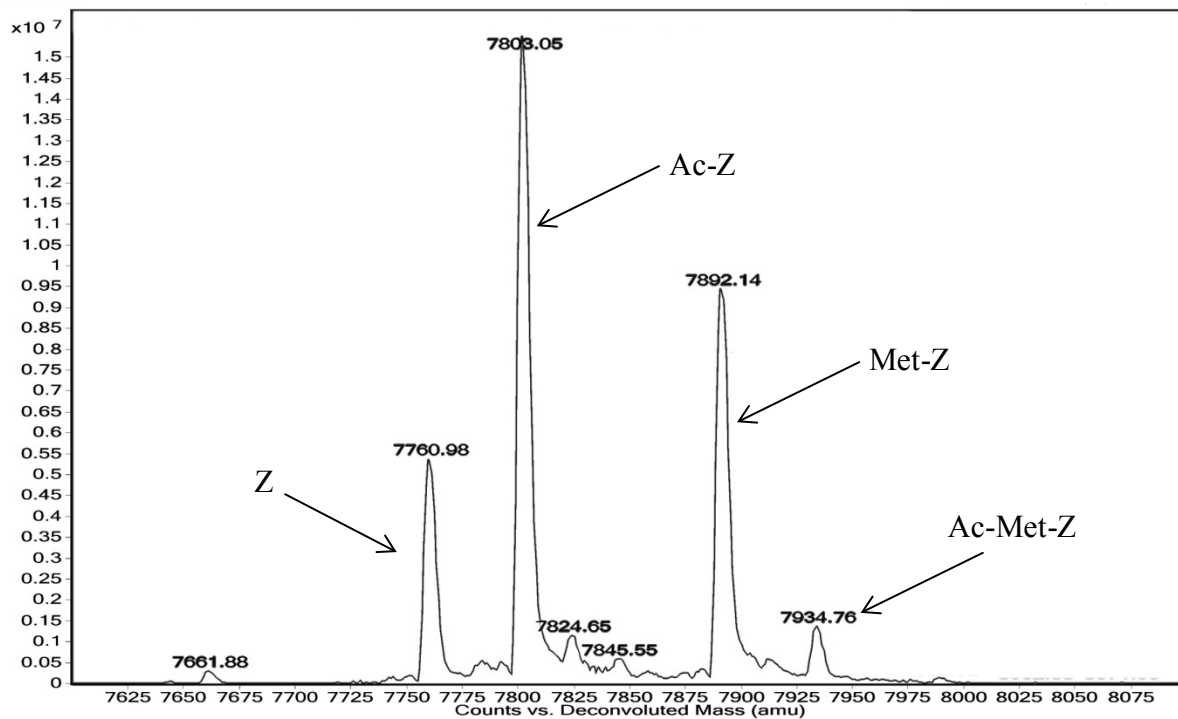


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: e) threonine and f) cysteine.

g)



h)

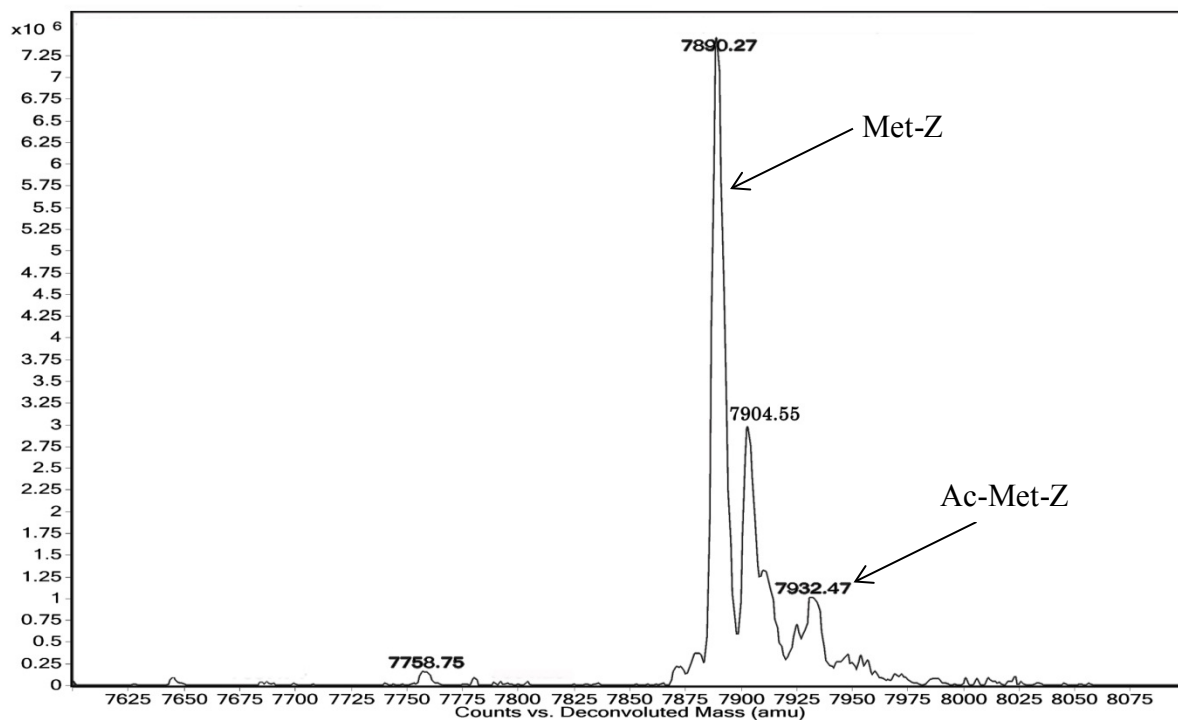
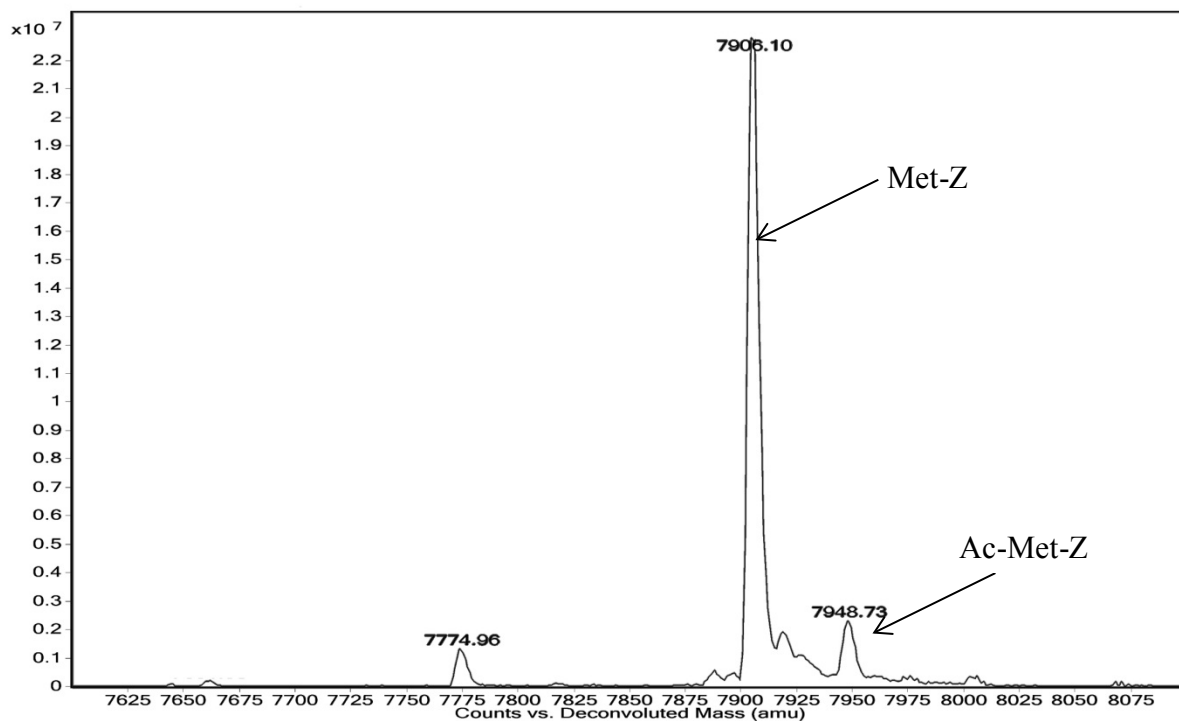


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: g) valine and h) leucine.

i)



j)

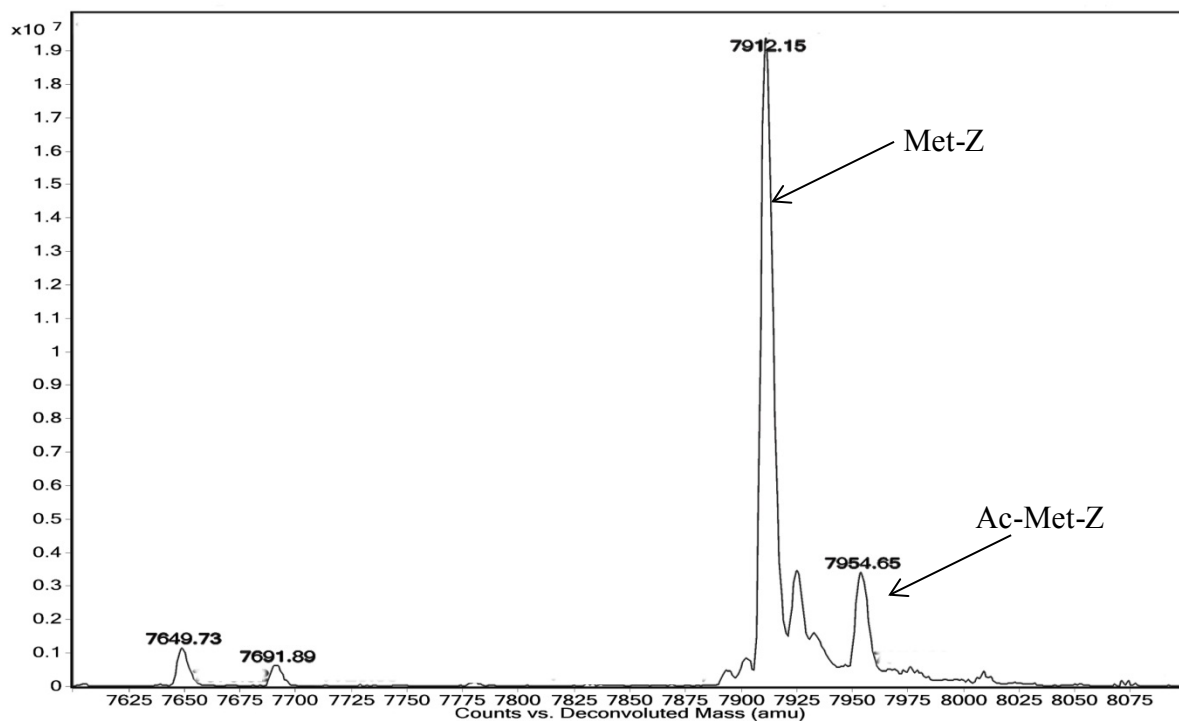
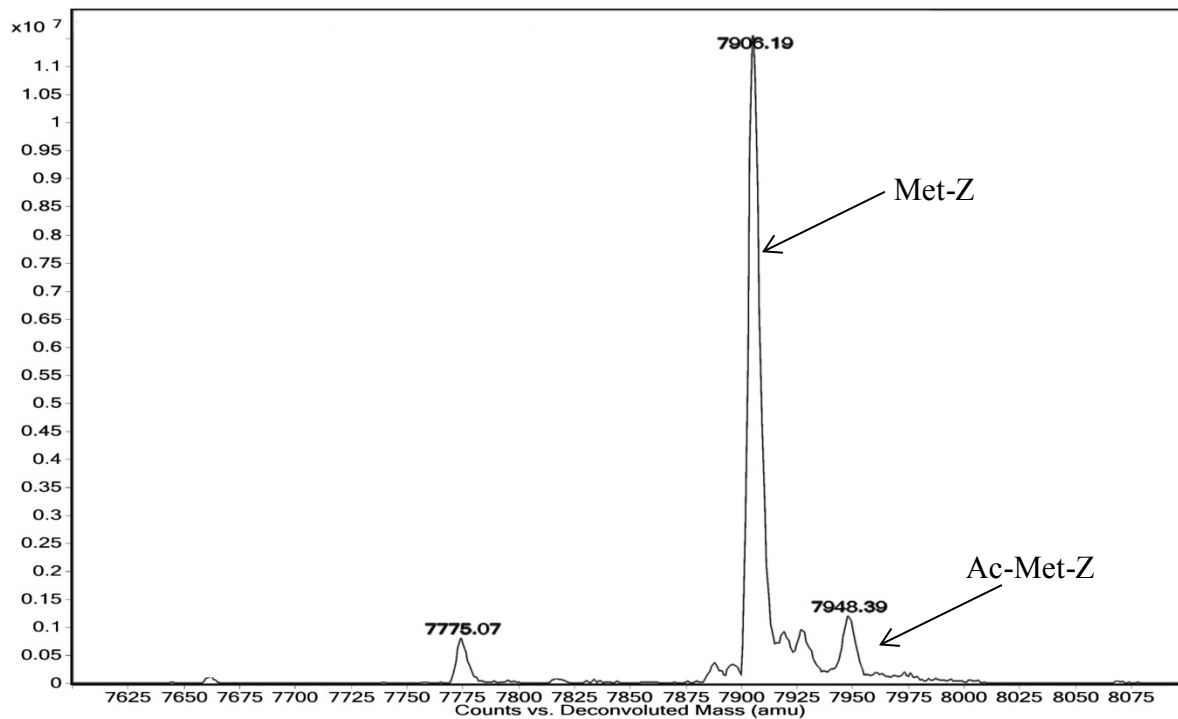


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: i) isoleucine and j) methionine.

k)



l)

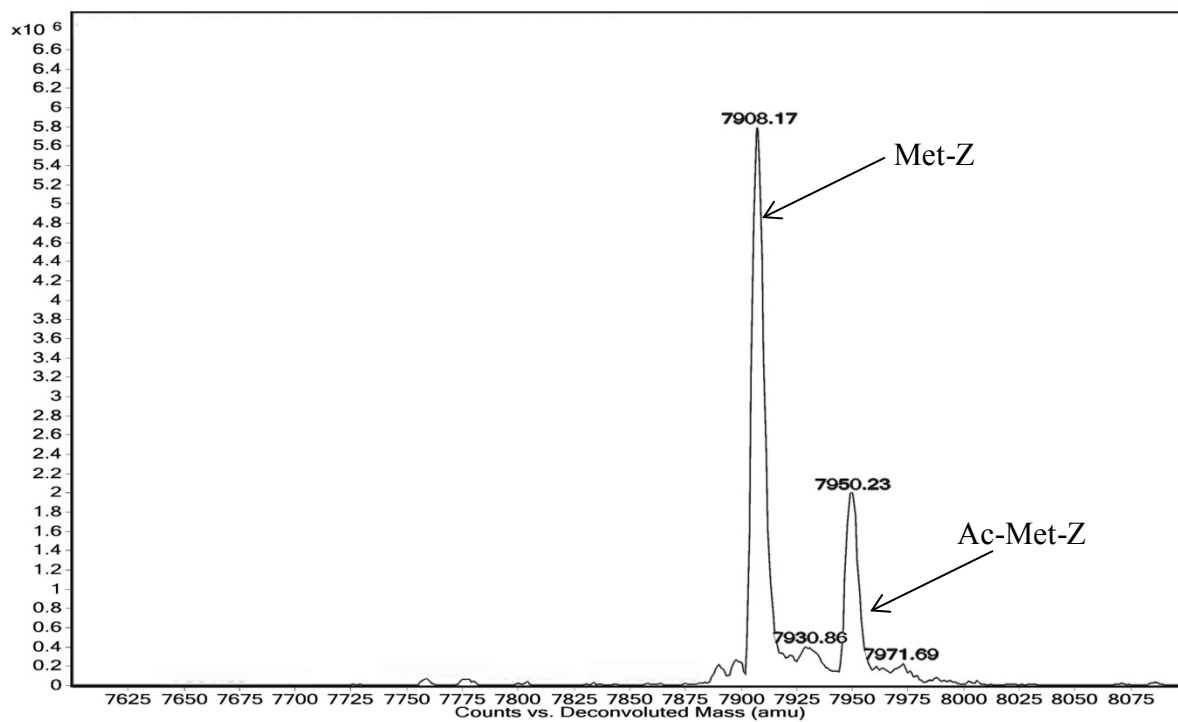
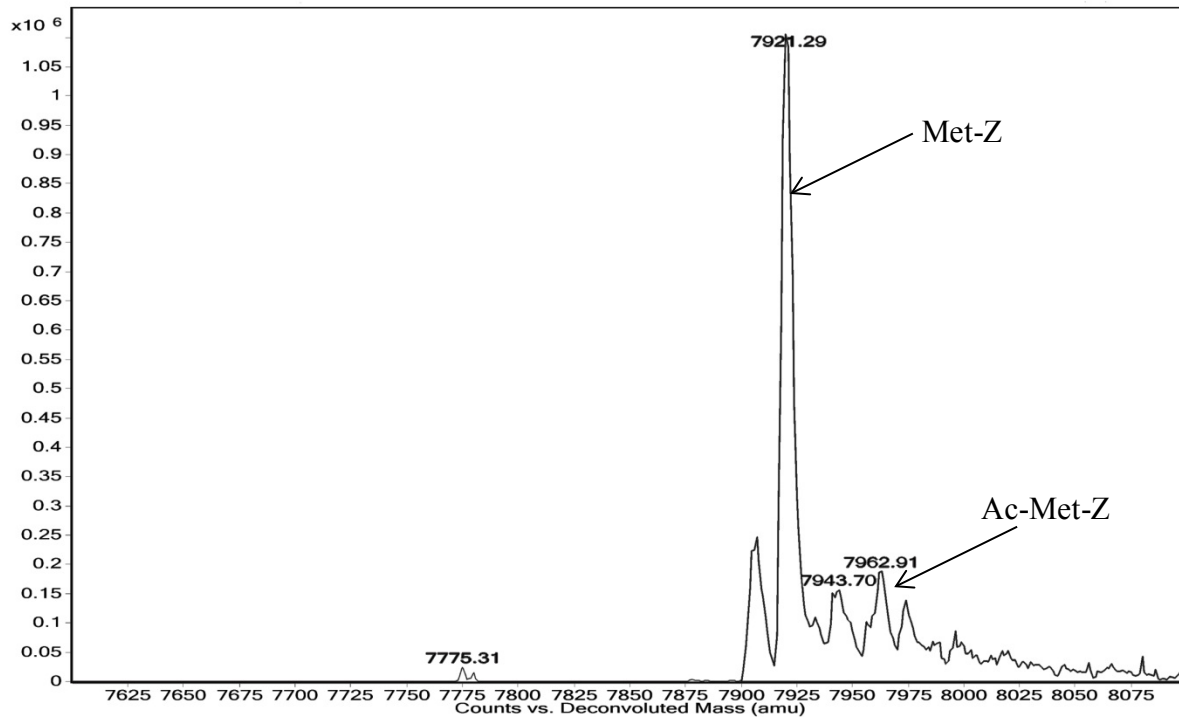


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: k) asparagine and l) aspartic acid.

m)



n)

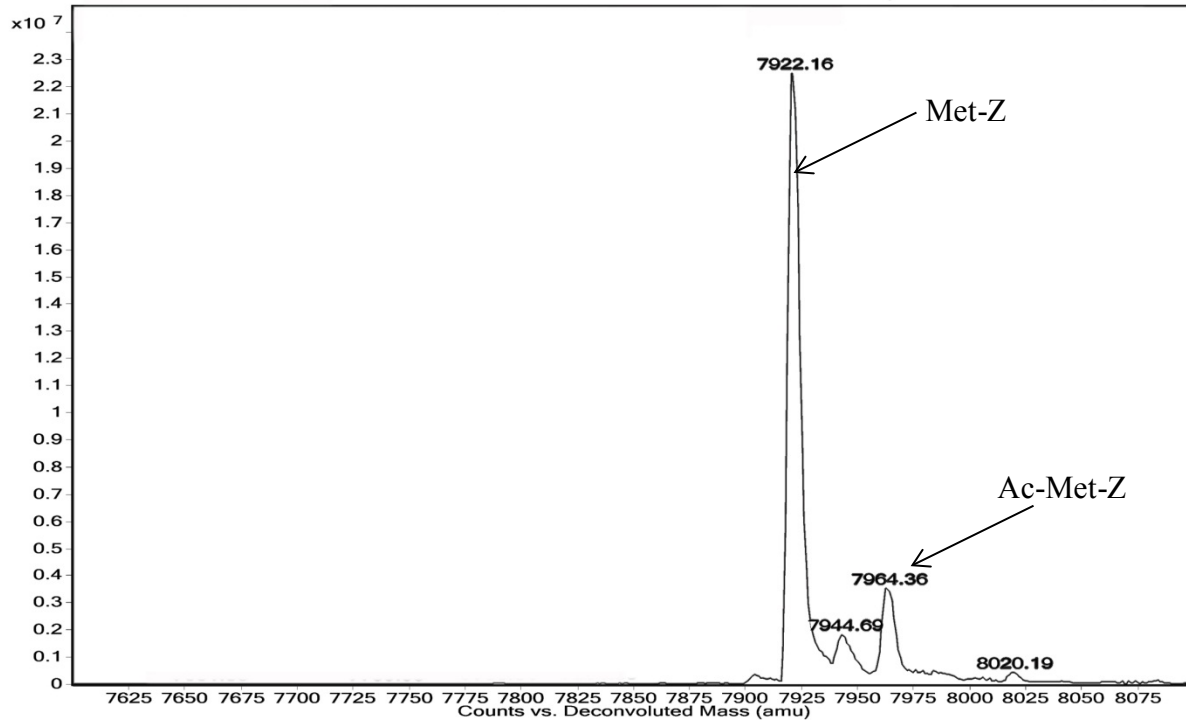
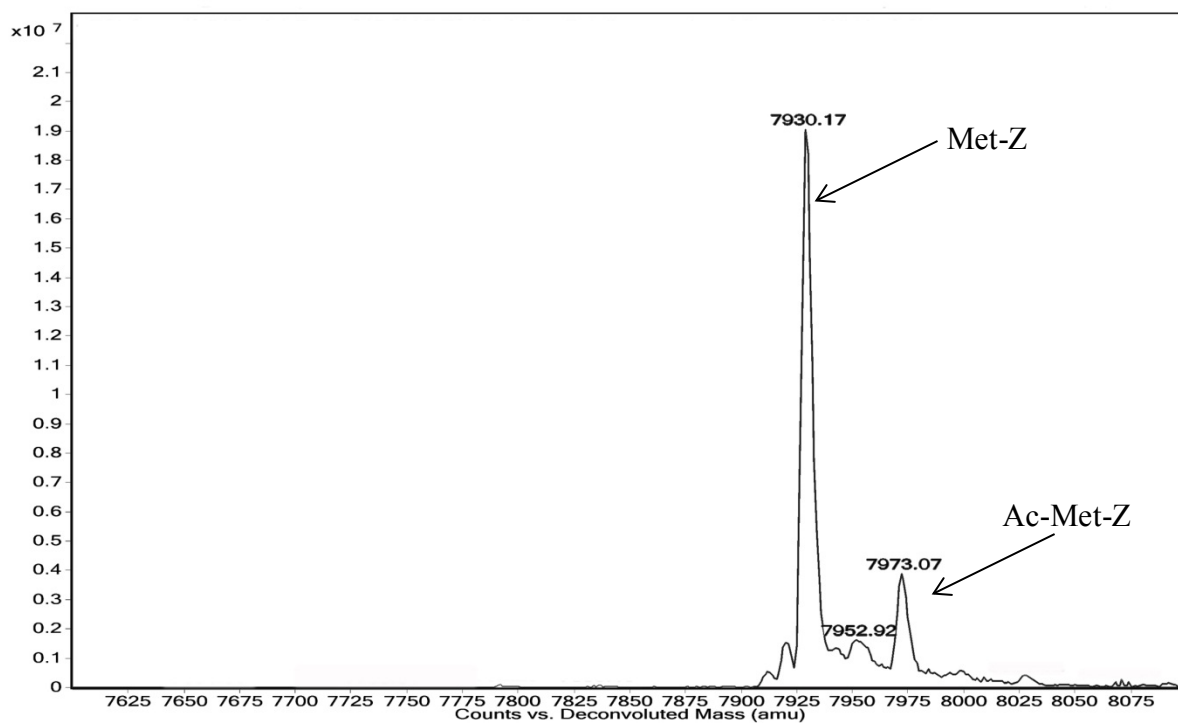


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: m) glutamine and n) glutamic acid.

o)



p)

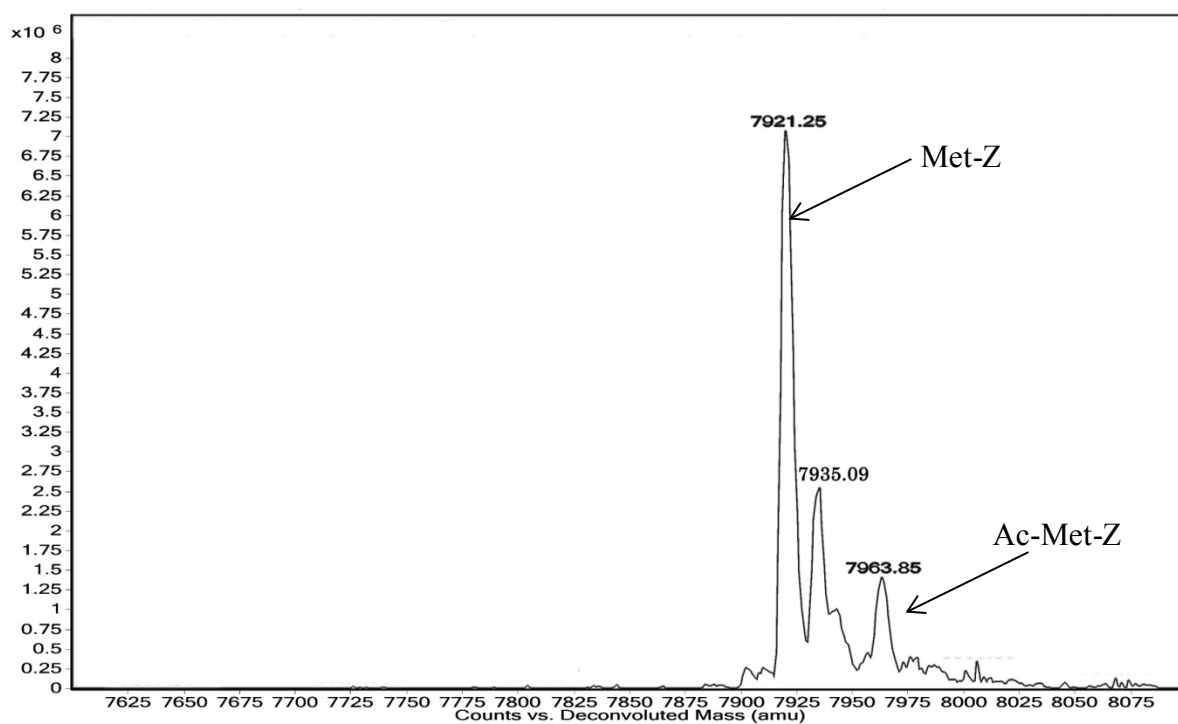
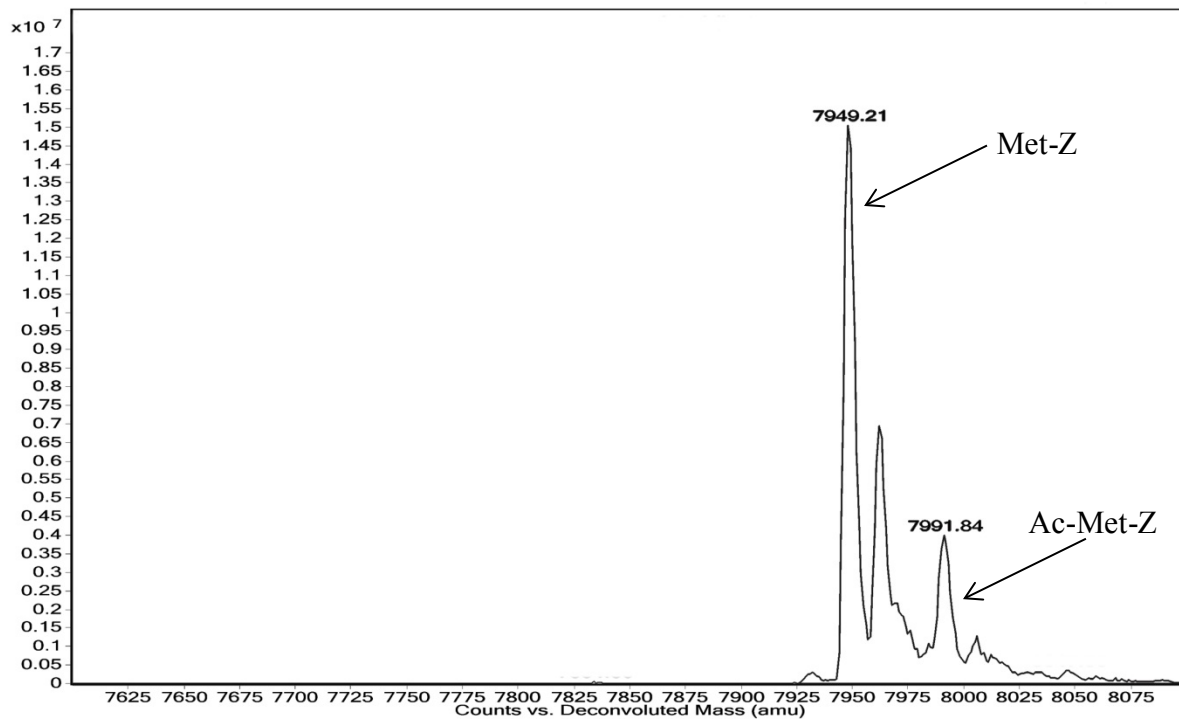


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: o) histidine and p) lysine.

q)



r)

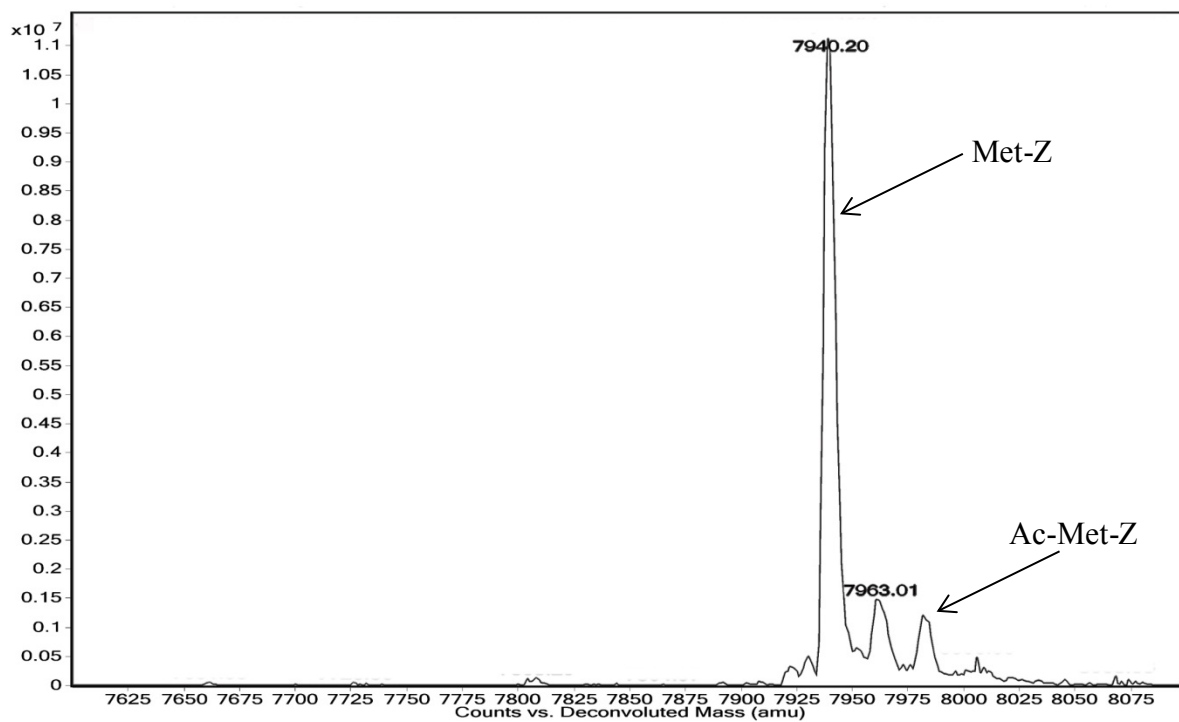
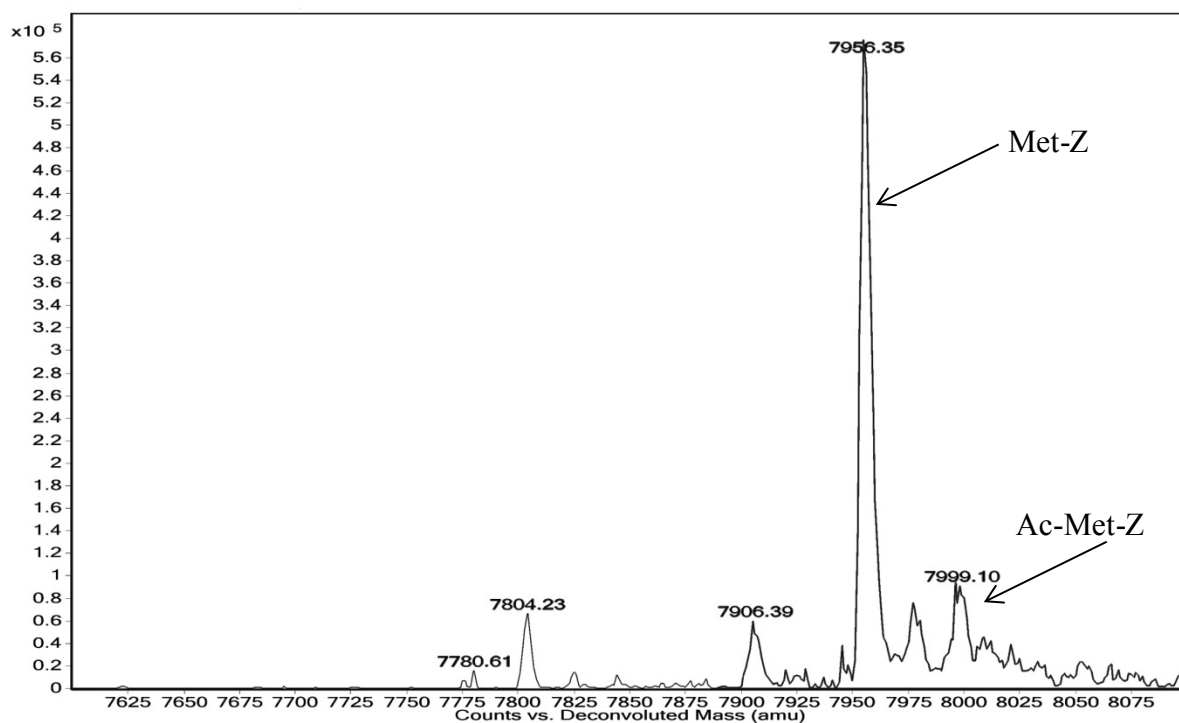


Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: q) arginine and r) phenylalanine.

s)



t)

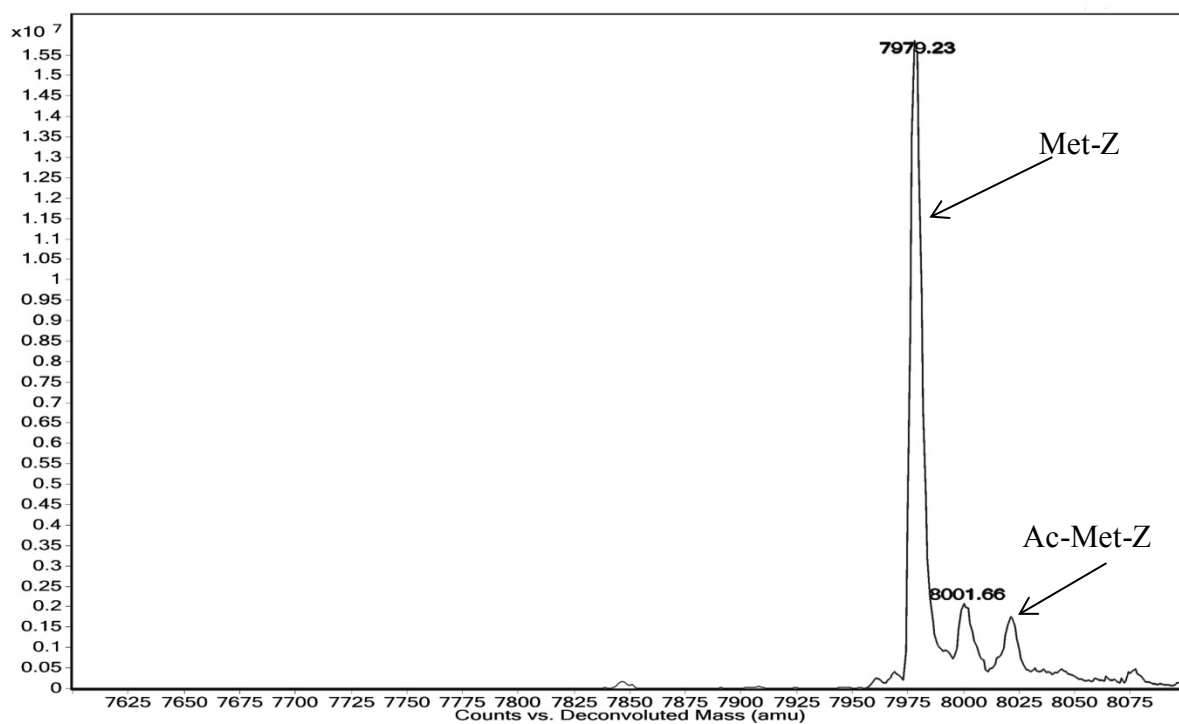


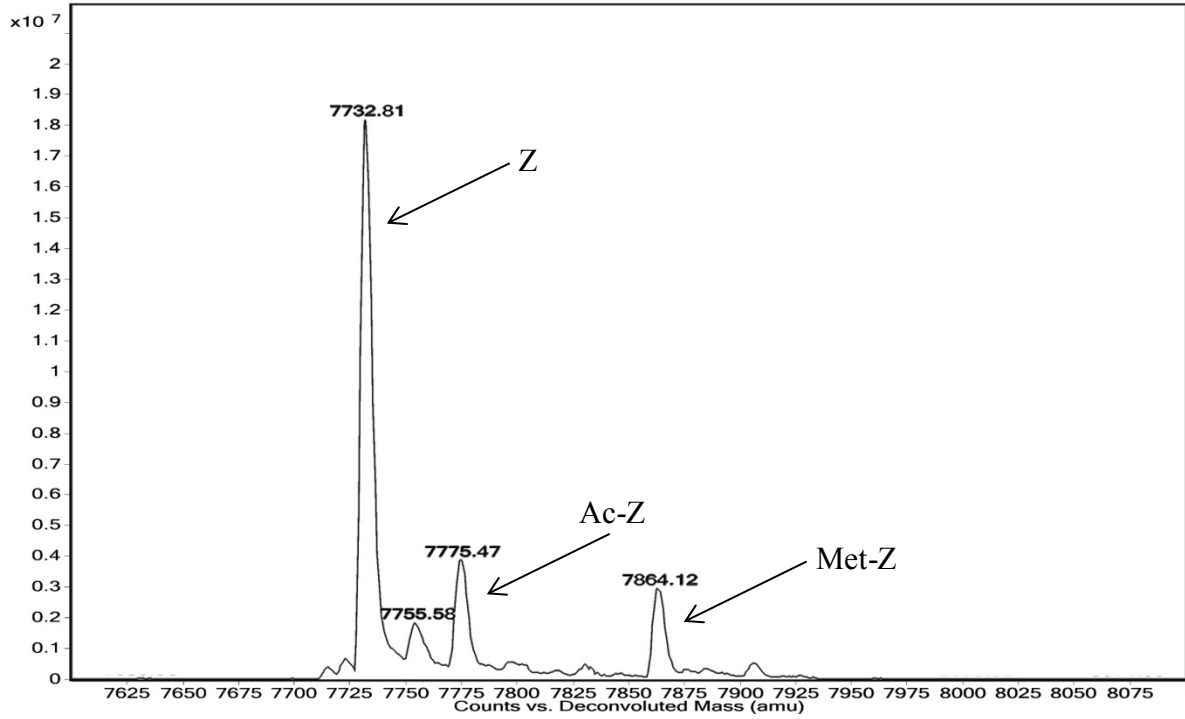
Figure 3.1 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2: s) tyrosine and t) tryptophan.

Table 3.1. Observed masses and their relative intensities of the different N-terminal end forms for each Z-domain variant differing by the second amino acid residue.

Amino acid in position 2	Observed masses, Da (Calculated average mass Da, relative intensity)			
	The Z form ^c	The Ac-Z form ^d	The Met-Z form ^e	The Ac-Met-Z form ^f
Gly	7718.92 (7718.39, 100%)	7761.48 (7760.42, 22%)		
Ala	7732.90 (7732.41, 100%)	7775.29 (7774.45, 26%)		
Pro	7758.93 (7758.45, 100%)	7801.92 (7800.49, 16%)		
Ser	7748.95 (7748.41, 100%)	7791.09 (7790.45, 79%)		
Thr	7762.98 (7762.44, 100%)	7805.25 (7804.48, 45%)	7894.07 (7893.64, 6%)	
Cys	7791.03 (7790.52, 100%) ^g 7835.35 (7834.53, 83%) ^h			
Val	7760.98 (7760.47, 35%)	7803.05 (7802.50, 100%)	7892.15 (7891.66, 61%)	7934.76 (7933.70, 8%)
Leu ^a			7890.27 (7889.65, 100%)	7932.47 (7931.68, 13%)
Ile			7906.10 (7905.69, 100%)	7948.73 (7947.73, 9%)
Met ^b			7912.15 (7911.67, 100%)	7954.65 (7953.71, 16%)
Asn			7906.19 (7906.63, 100%)	7948.39 (7748.67, 10%)
Asp			7908.17 (7907.62, 100%)	7950.23 (7949.66, 34%)
Gln			7921.29 (7920.66, 100%)	7962.91 (7962.70, 16%)
Glu			7922.16 (7921.65, 100%)	7964.36 (7963.68, 16%)
His			7930.17 (7929.67, 100%)	7973.07 (7971.71, 21%)
Lys			7921.25 (7920.71, 100%),	7963.85 (7962.74, 19%)
Arg			7949.21 (7948.72, 100%),	7991.84 (7990.76, 27%)
Phe			7940.20 (7939.71, 100%),	7982.83 (7981.74, 11%)
Tyr			7956.36 (7955.71, 100%),	7999.10 (7997.74, 18%)
Trp			7979.23 (7978.74, 100%),	8021.86 (8020.78, 11%)

Note: ^aDNA sequencing of the expression plasmid revealed a single nucleotide substitution which resulted in an additional amino acid substitution (Leu54Pro); ^bDNA sequencing of the expression plasmid revealed a single nucleotide substitution which resulted in an additional amino acid substitution (Ile19Thr); ^cZ-domain without the Met₁ residue; ^dN^α-Acetylated Z-domain without the Met₁ residue; ^eZdomain with the Met₁ residue; ^fN^α-Acetylated Z-domain with the Met₁ residue; ^g2-Methylthiazolidine derivative of the Z-domain; ^h2-Methyl-2-thiazolidinecarboxylate derivative of the Z-domain.

a)



b)

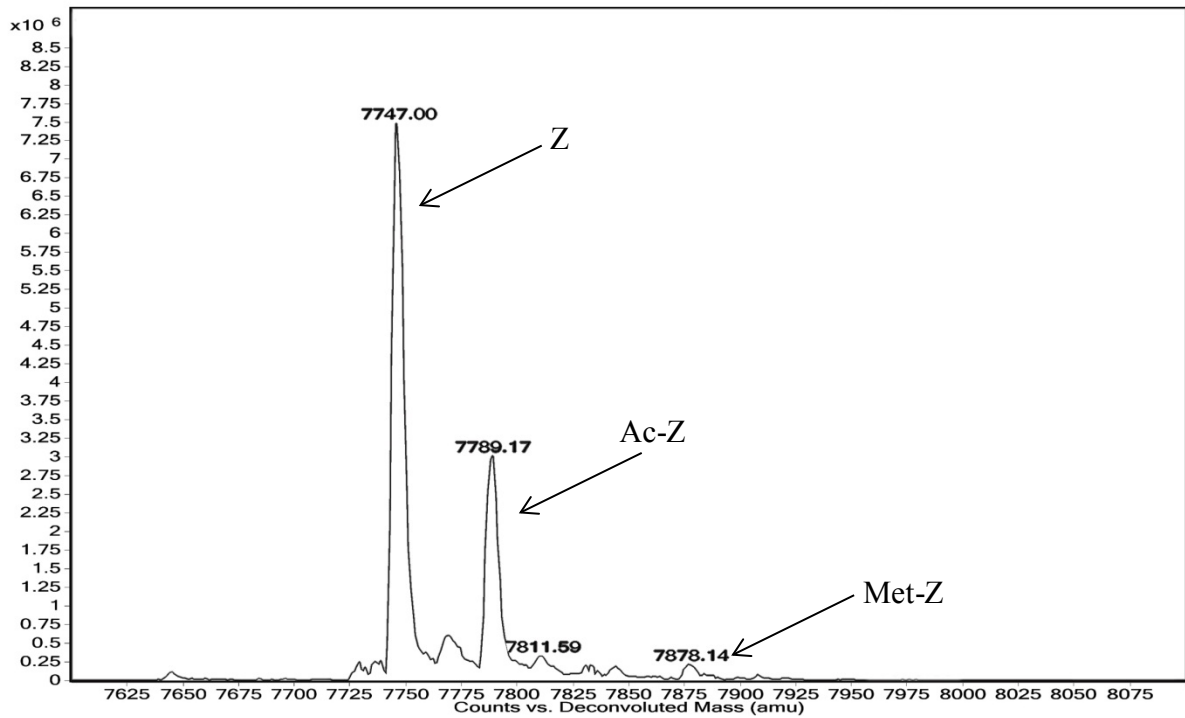
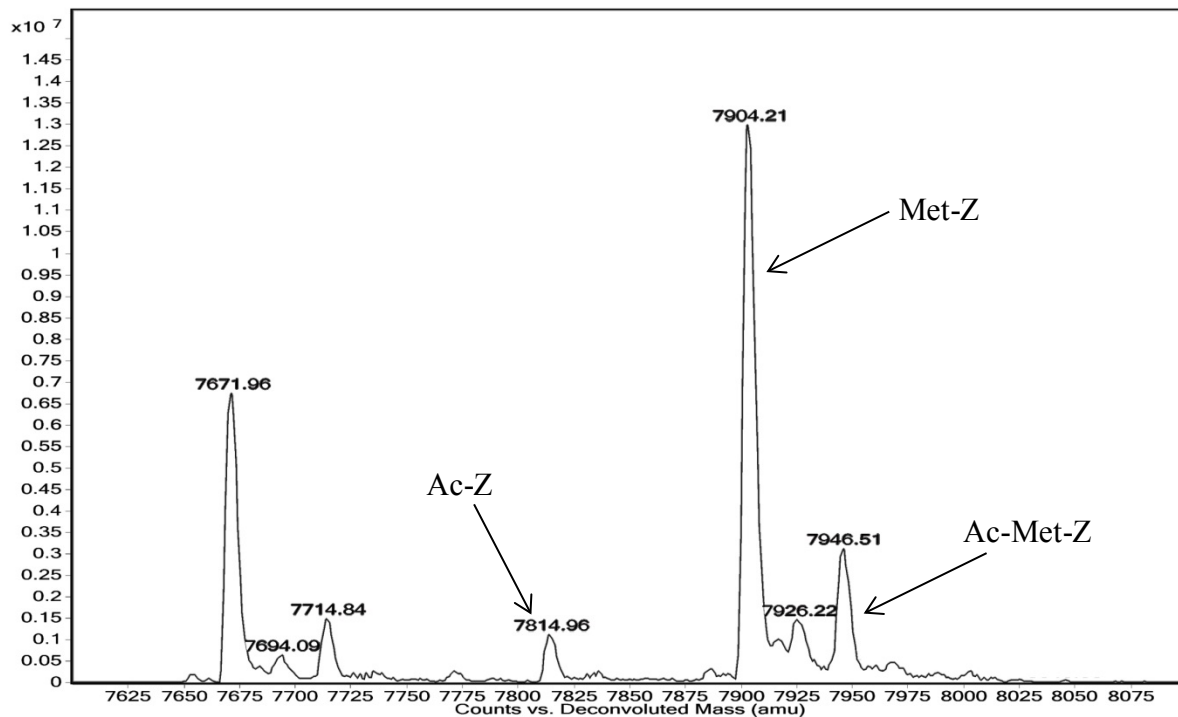


Figure 3.2 Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: a) glycine and b) alanine.

c)



d)

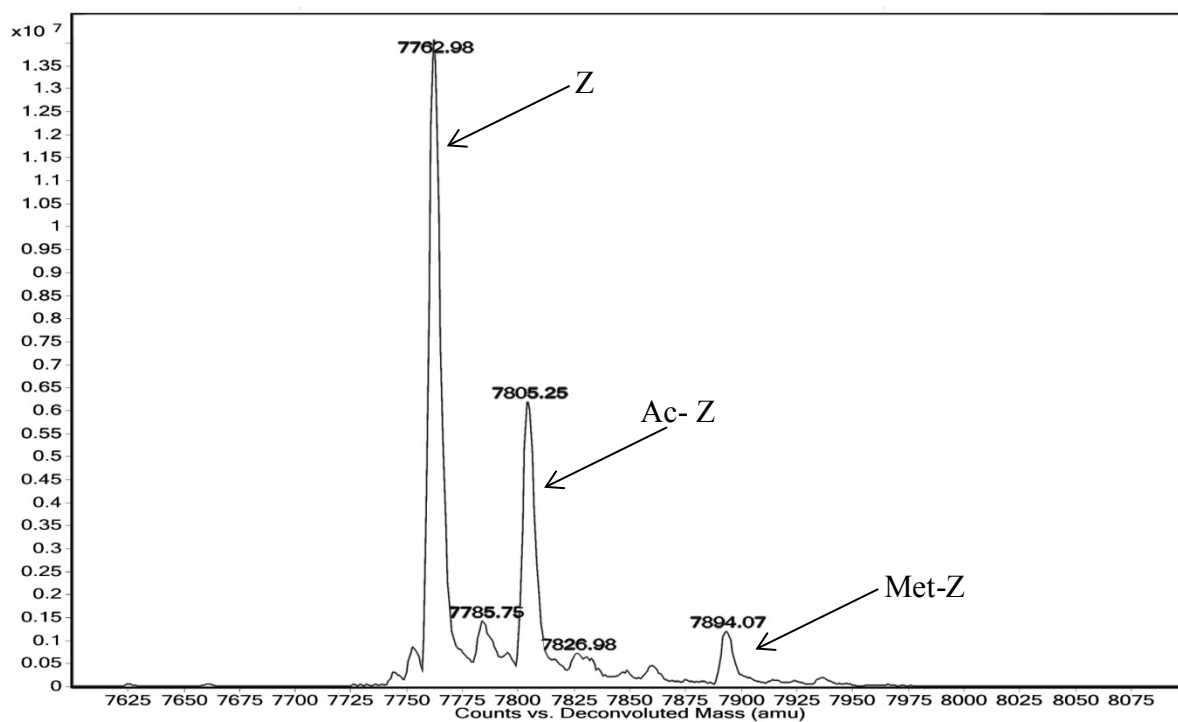
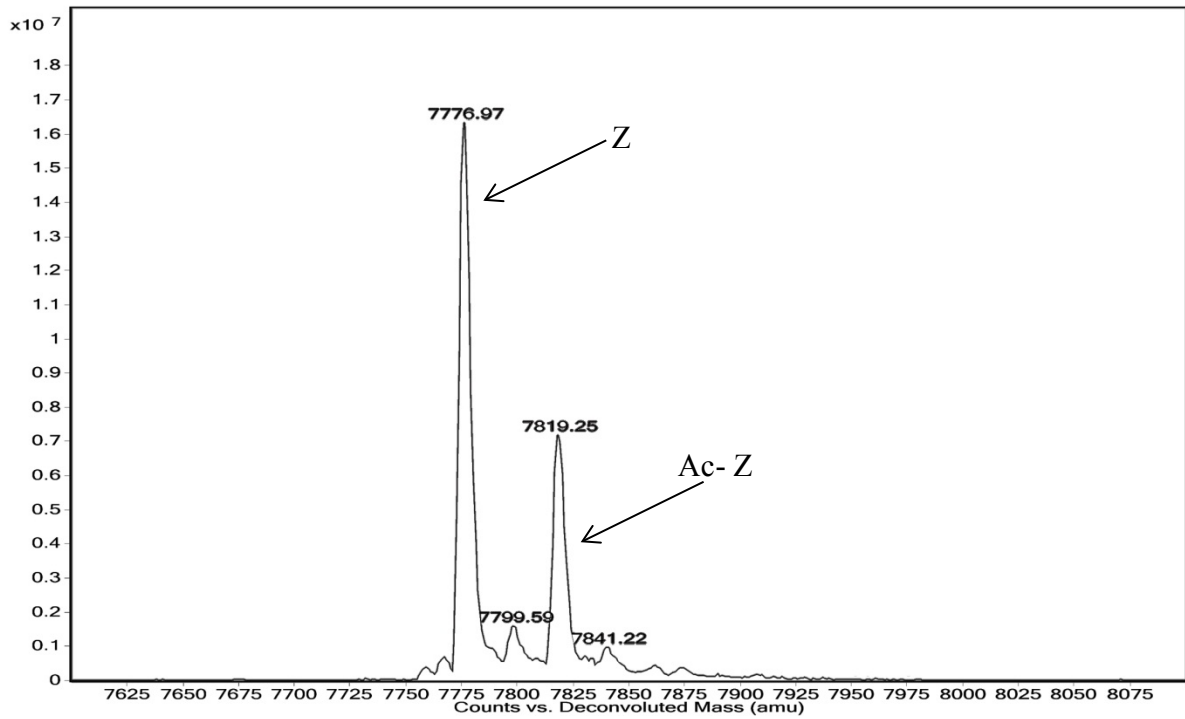


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: c) proline and d) serine.

e)



f)

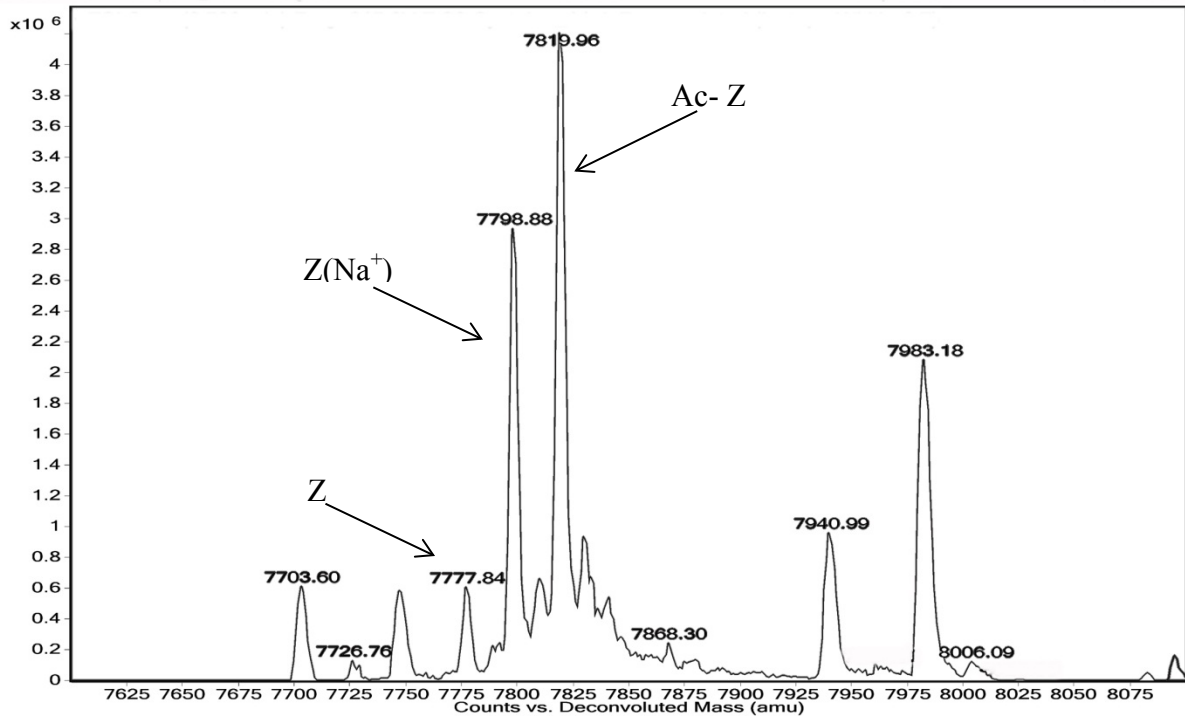
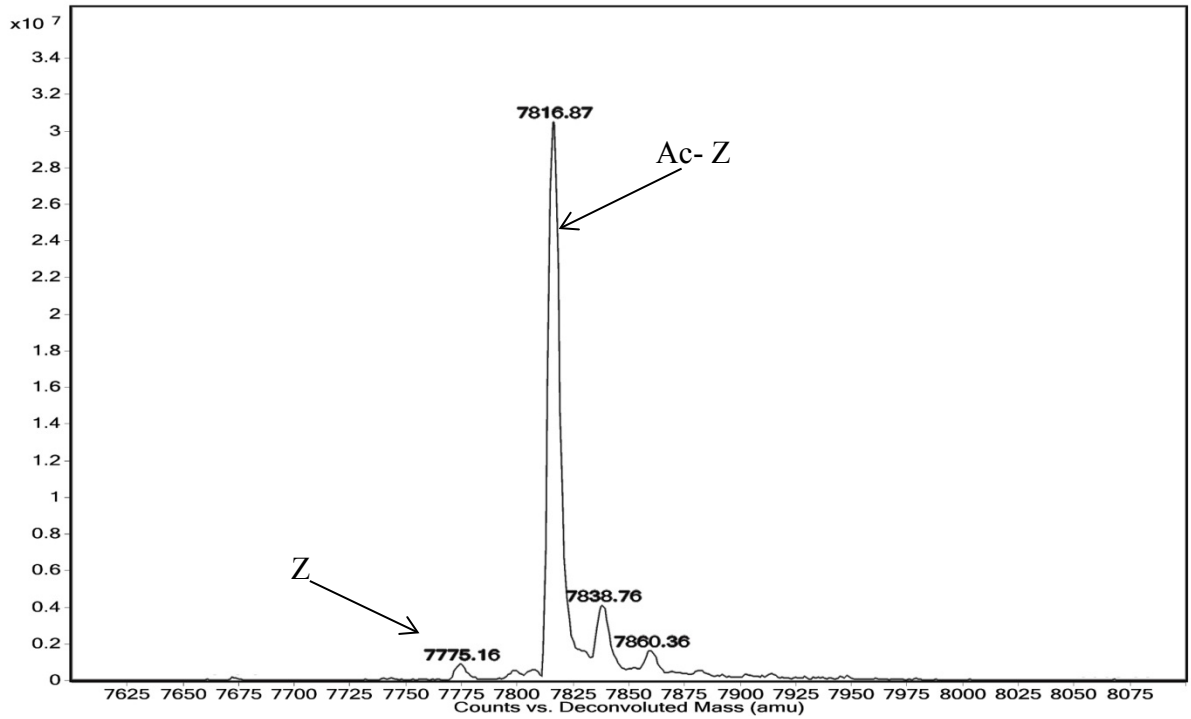


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: e) threonine and f) cysteine.

g)



h)

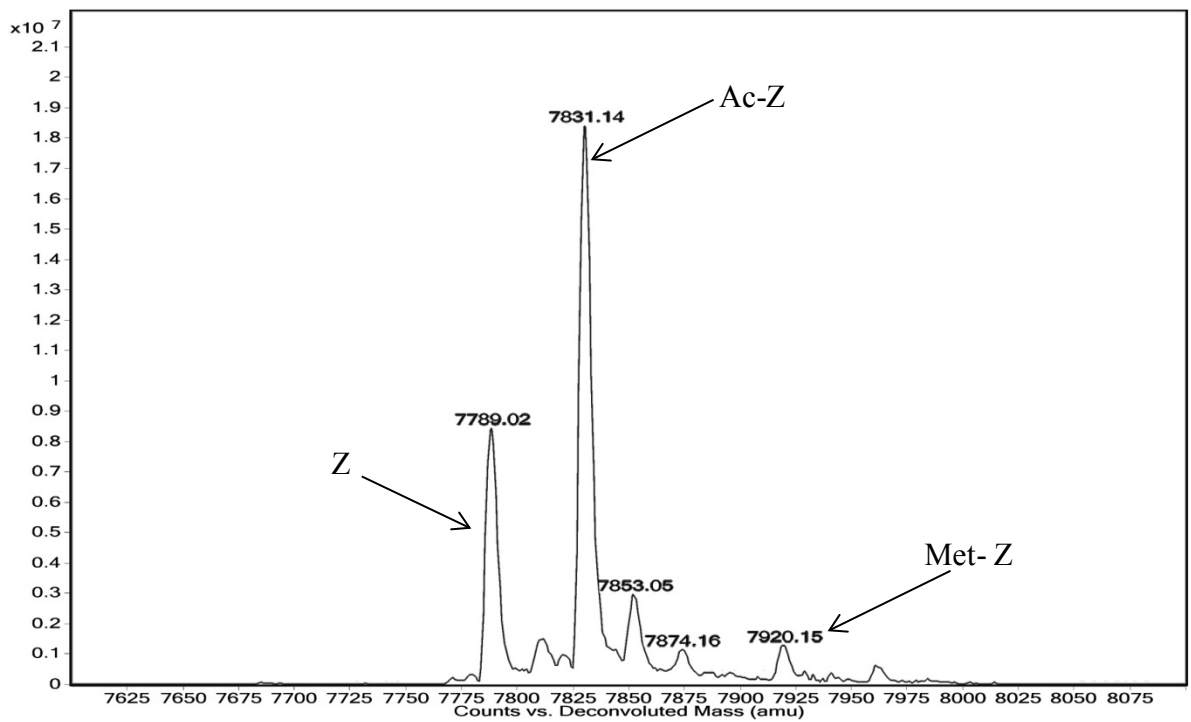
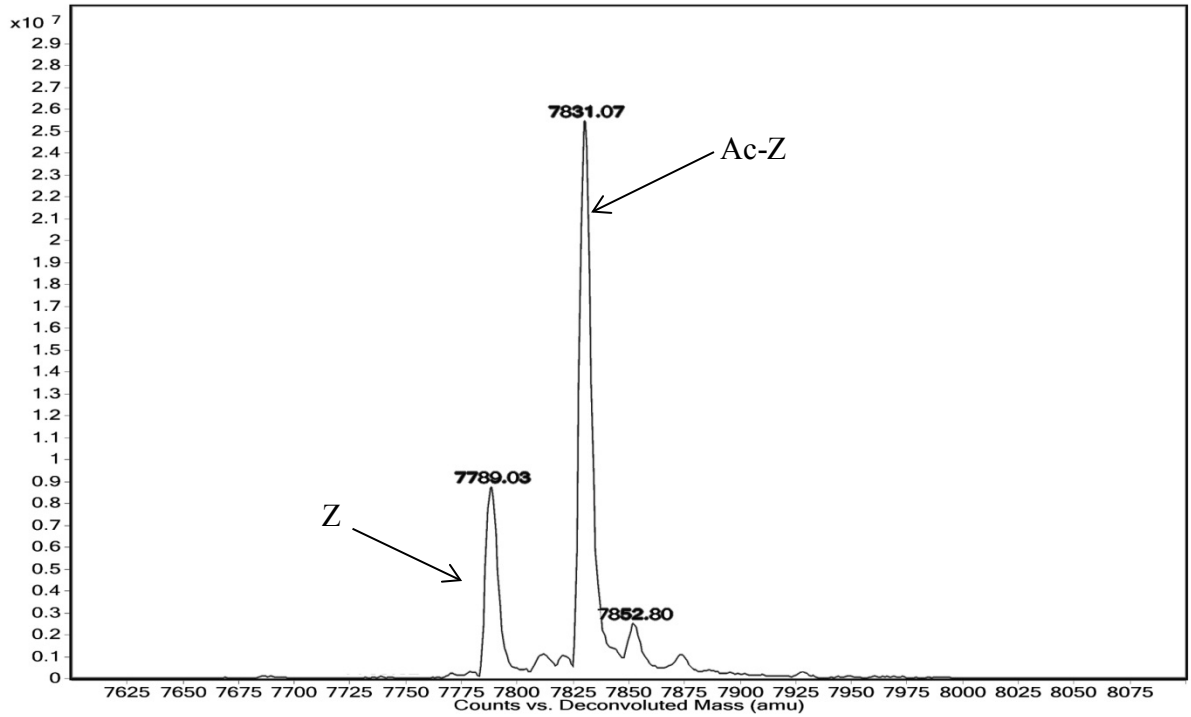


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: g) valine and h) leucine.

i)



j)

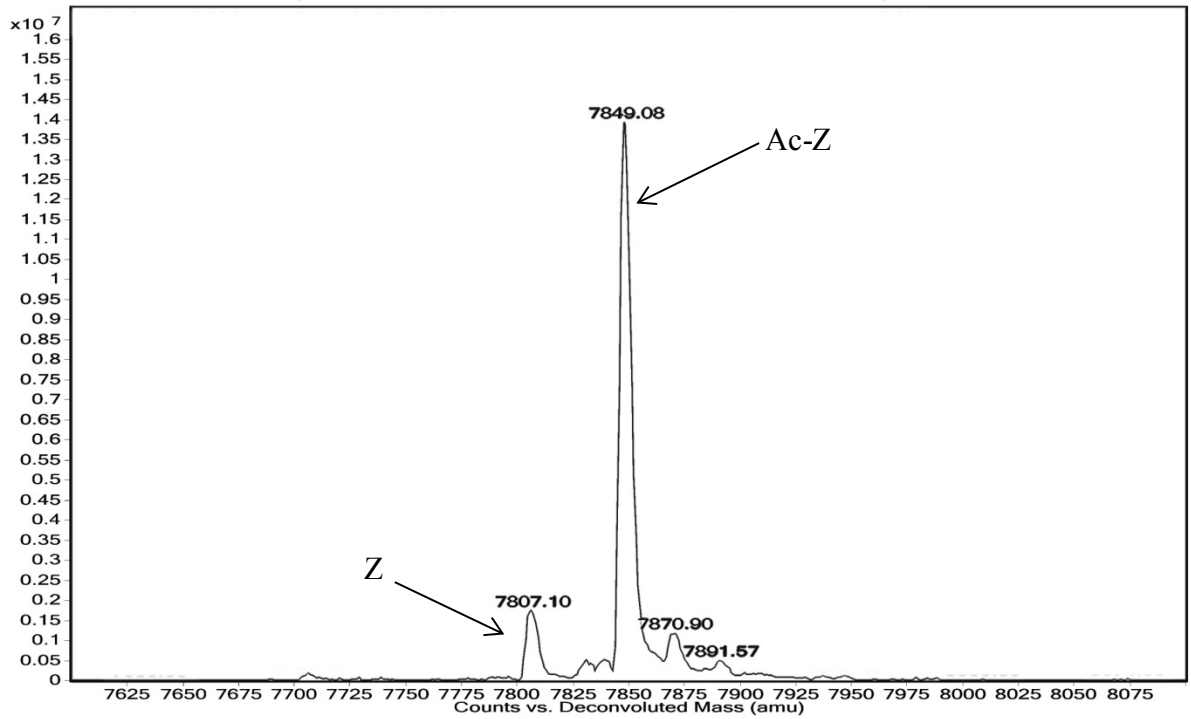
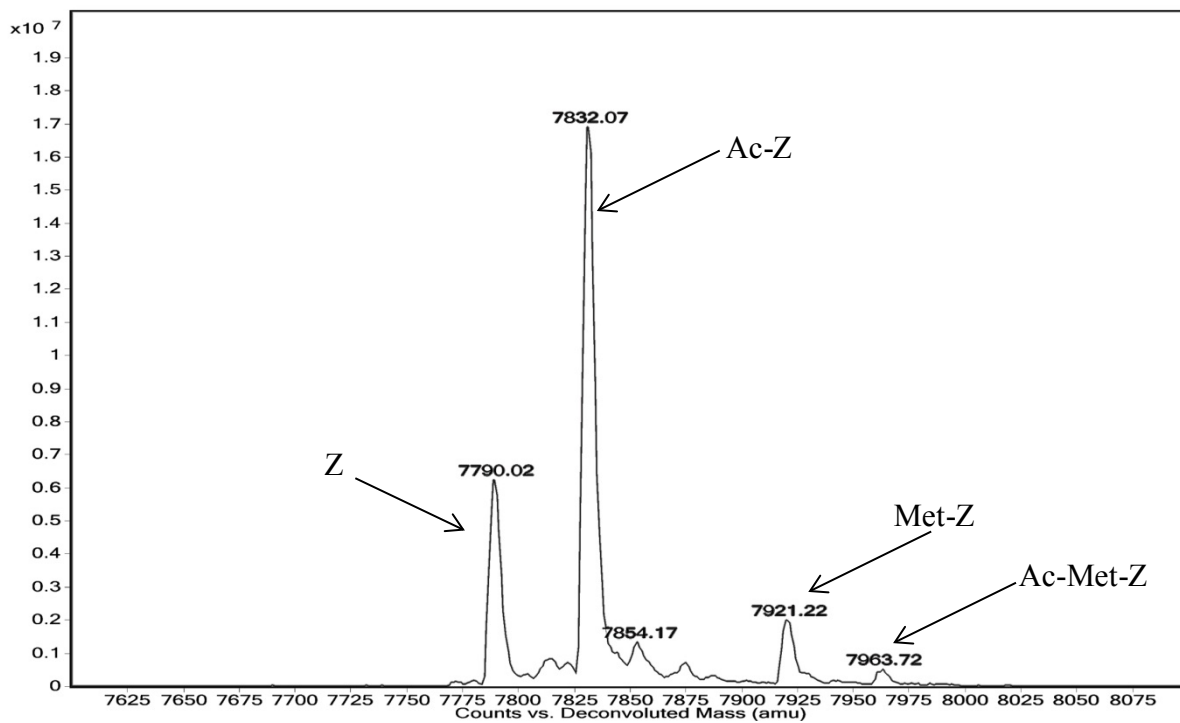


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: i) isoleucine and j) methionine.

k)



l)

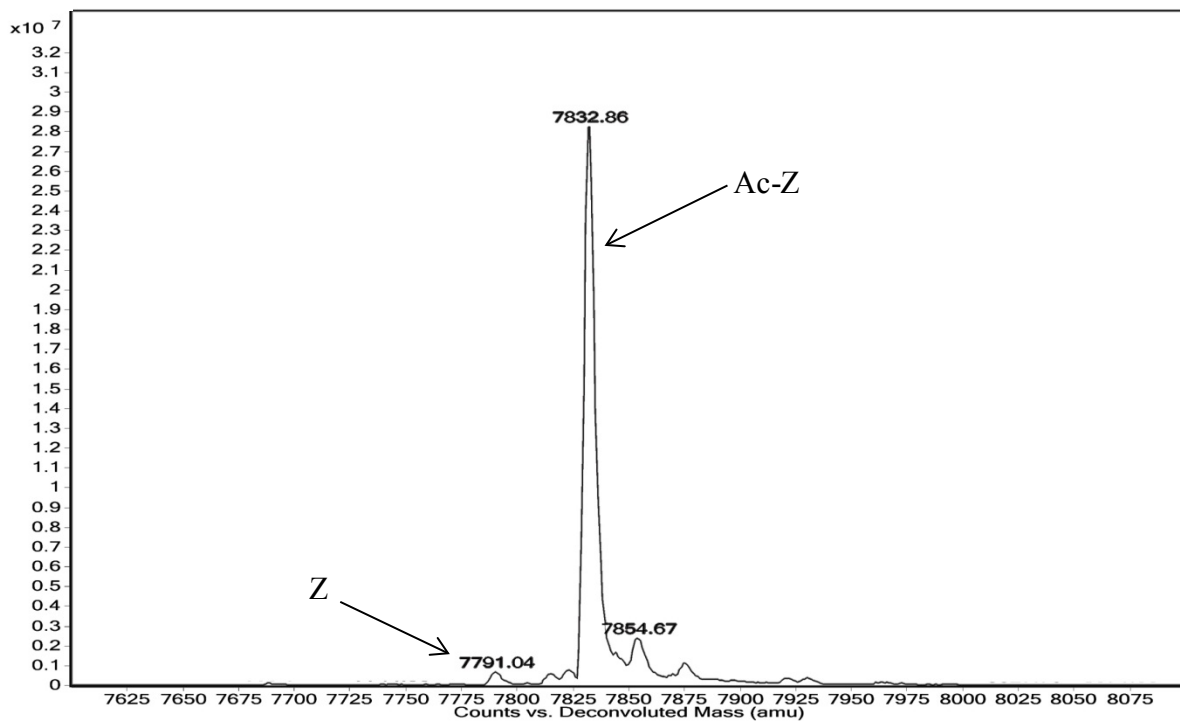
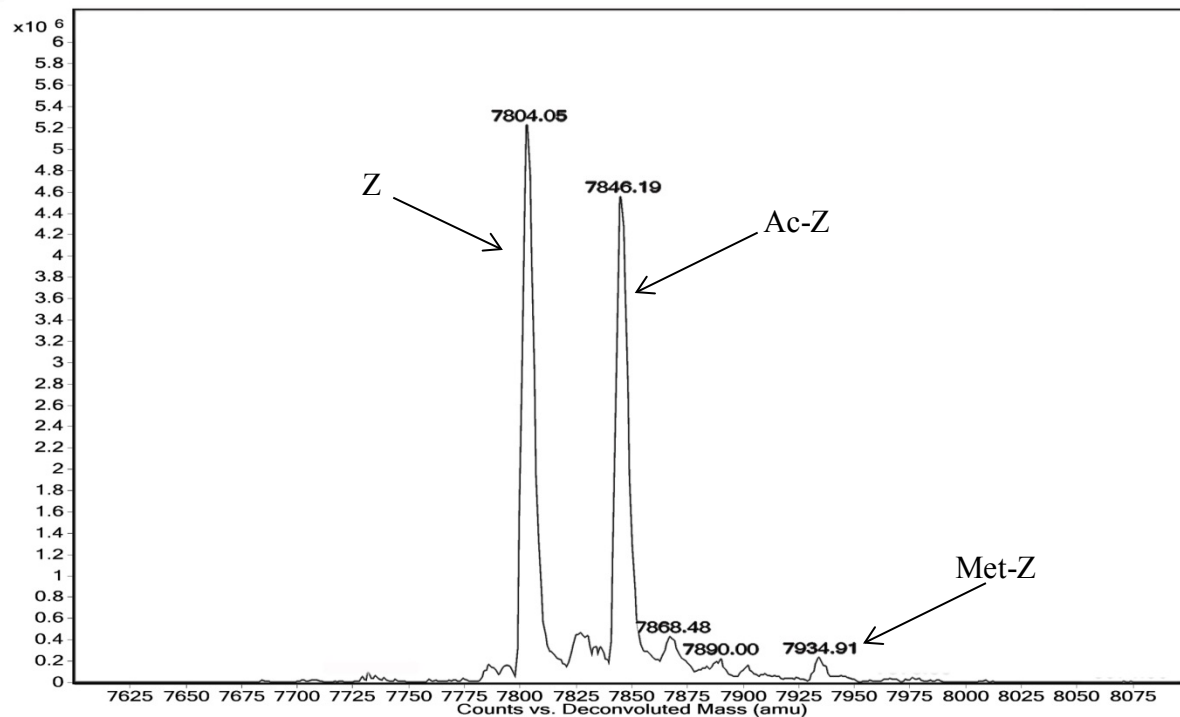


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: k) asparagine and l) aspartic acid.

m)



n)

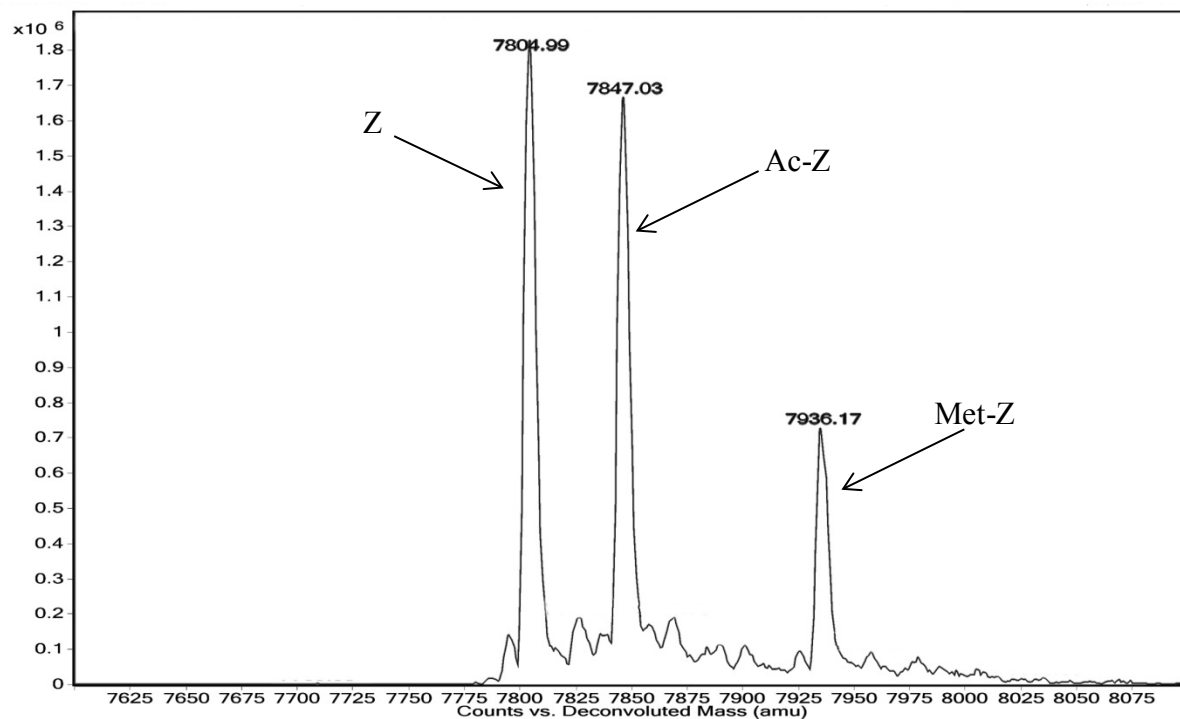
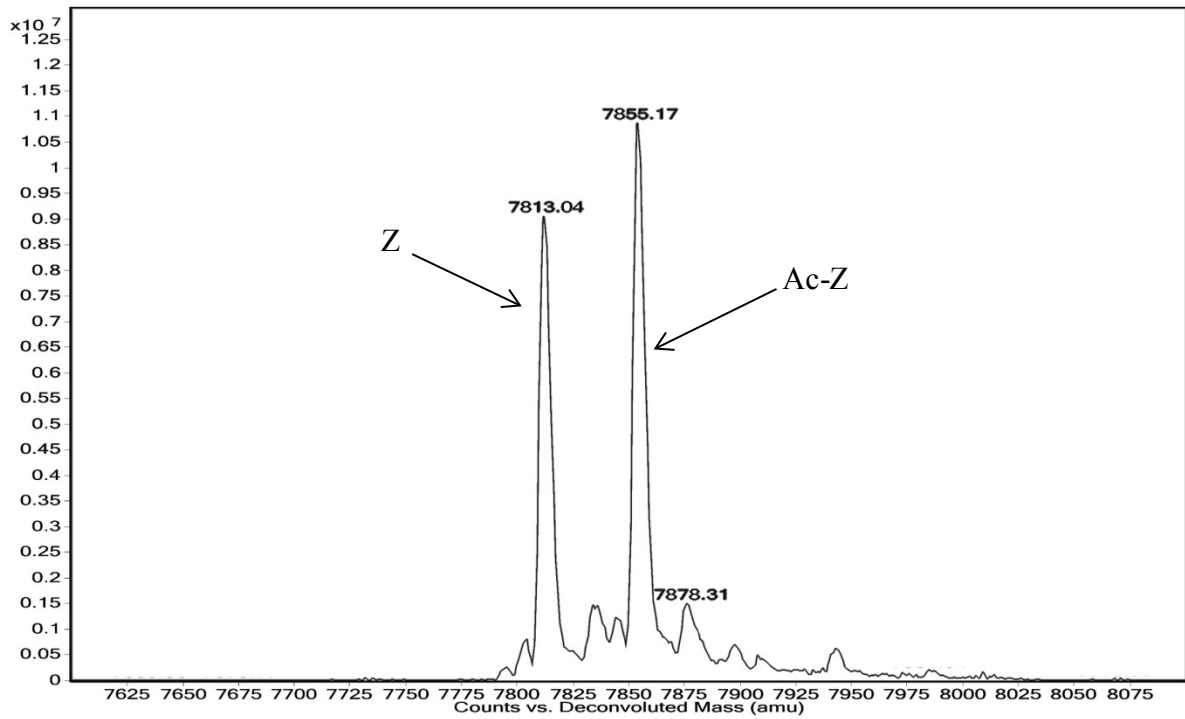


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: m) glutamine and n) glutamic acid.

o)



p)

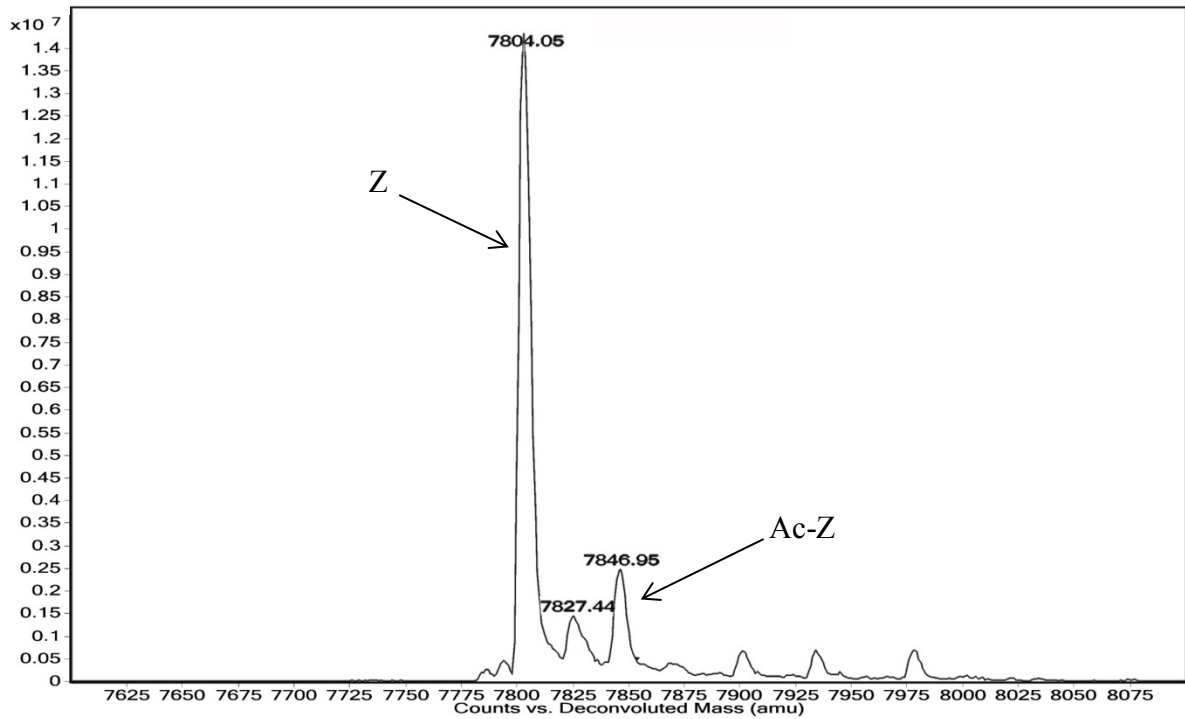


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: o) histidine and p) lysine.

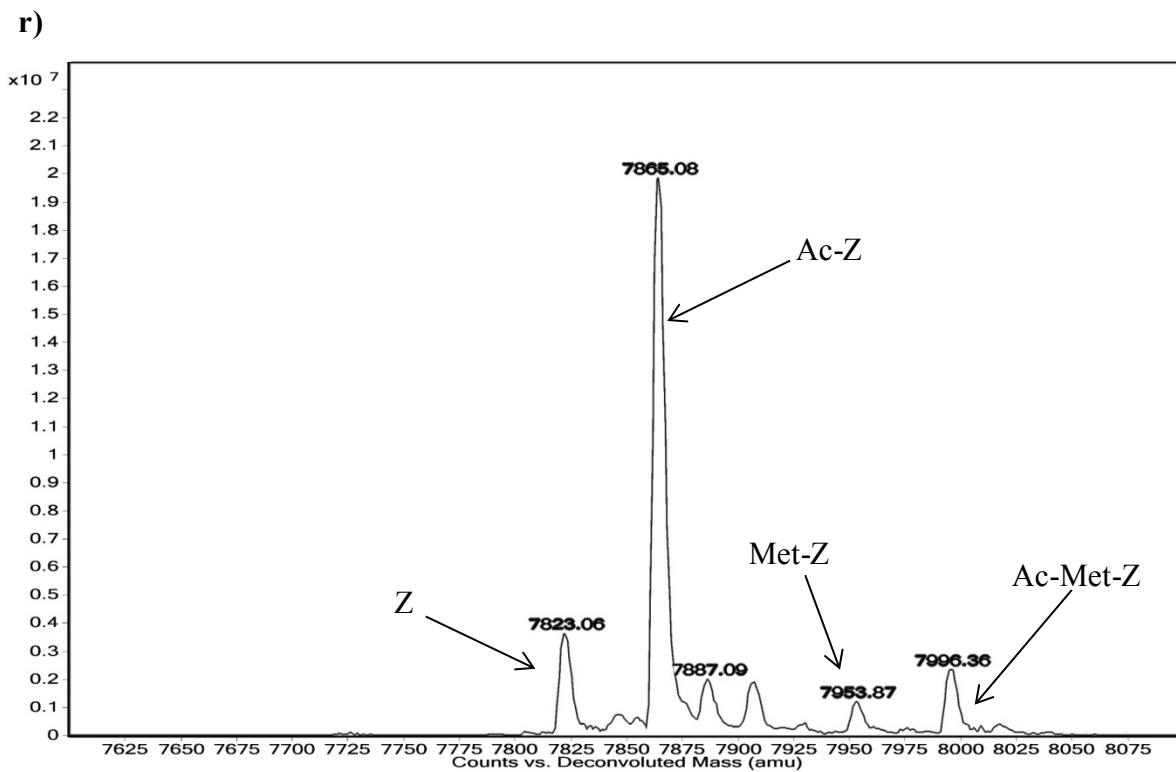
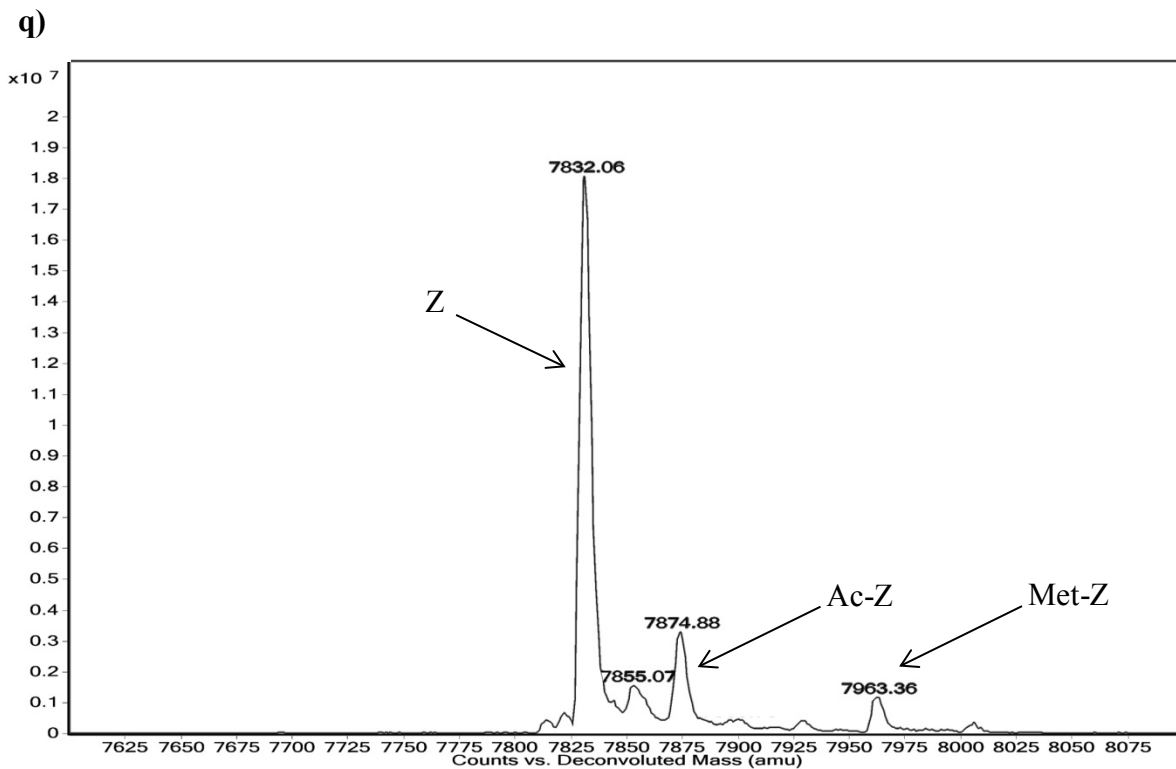
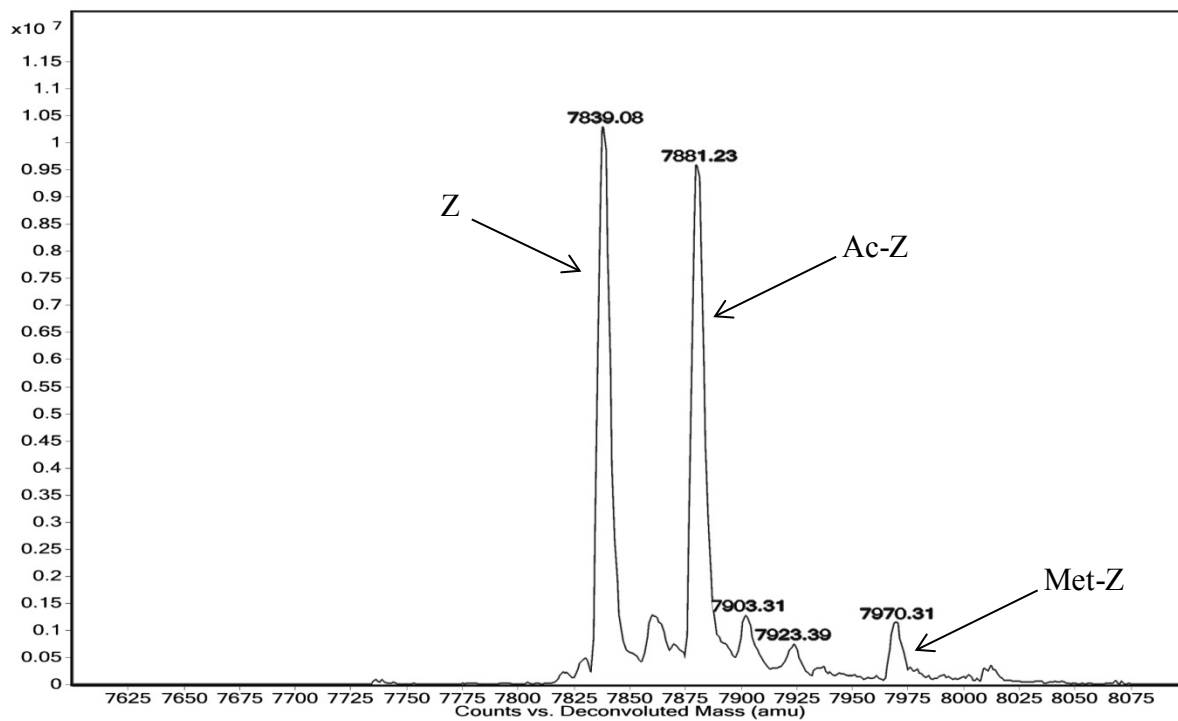


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: q) arginine and r) phenylalanine.

s)



t)

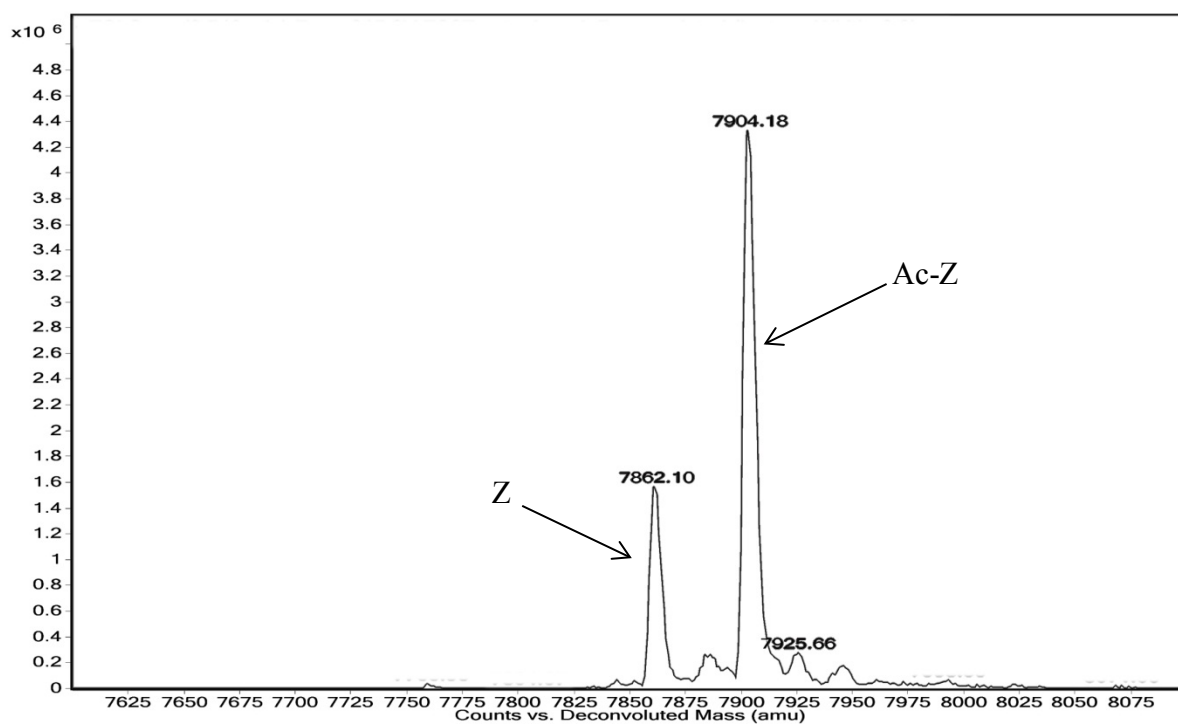


Figure 3.2 Cont. Deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 3: s) tyrosine and t) tryptophan.

Table 3.2. Observed masses and their relative intensities of the different N-terminal end forms for each Z-domain variant differing by the third amino acid residue.

Amino acid in position 3	Observed masses, Da (Calculated average mass Da, relative intensity)			
	The Z form ^a	The Ac-Z form ^b	The Met-Z form ^c	The Ac-Met-Z form ^d
Gly	7732.81 (7732.41, 100%)	7775.47 (7774.45, 20%)	7864.12 (7763.61, 15%)	
Ala	7747.00 (7746.44, 100%)	7789.17 (7788.48, 40%)	7878.14 (7877.64, 3%)	
Pro		7814.96 (7814.52, 7%)	7904.21 (7903.67, 100%)	7946.51 (7945.71, 21%)
Ser	7762.98 (7762.44, 100%)	7805.25 (7804.48, 40%)	7894.07 (7893.64, 6%)	
Thr	7776.97 (7776.47, 100%),	7819.25 (7818.50, 44%)		
Cys	7798.88 (7800.50, 71%) ^e	7819.96 (7820.54, 100%)		
Val	7775.16 (7774.49, 3%)	7816.87 (7816.53, 100%)		
Leu	7789.02 (7788.52, 39%)	7831.14 (7830.56, 100%)	7920.15 (7919.72, 5%)	
Ile	7789.03 (7788.52, 31%)	7831.07 (7830.56, 100%)		
Met	7807.10 (7806.56, 10%)	7849.08 (7848.60, 100%)		
Asn	7790.02 (7789.47, 33%)	7832.07 (7831.50, 100%)	7921.22 (7920.66, 11%)	7963.72 (7962.70, 3%)
Asp	7791.04 (3%)	7832.86 (7832.49, 100%)		
Gln	7804.05 (7803.49, 100%)	7846.19 (7845.53, 85%)	7934.91 (7934.69, 4%)	
Glu	7804.99 (7804.48, 100%)	7847.03 (7846.51, 94%)	7936.17 (7935.67, 39%)	
His	7813.04 (7812.50, 82%)	7855.17 (7854.54, 100%)		
Lys	7804.05 (7804.54, 100%)	7846.95 (7845.57, 16%)		
Arg	7832.06 (7831.55, 100%)	7874.88 (7873.59, 15%)	7963.36 (7962.75, 5%)	
Phe	7823.06 (7822.54, 16%)	7865.08 (7864.58, 100%)	7953.87 (7953.73, 5%)	7996.36 (7995.77, 9%)
Tyr	7839.08 (7838.54, 100%)	7881.23 (7780.57, 90%)	7970.31 (7969.73, 10%)	
Trp	7862.10 (7861.57, 35%)	7904.18 (7903.61, 100%)		

Note: ^aZ-domain without the Met₁ residue; ^bN^α-Acetylated Z-domain without the Met₁ residue; ^cZ-domain with the Met₁ residue; ^dN^α-Acetylated Z-domain with the Met₁ residue; ^eThe Na⁺ adduct of the Z-domain without the Met₁ residue.

3.2.1 Second amino acid substitutions

The nature of the second amino acid had a significant impact on the removal of the Met₁ residue and the N^α-acetylation. The Met₁ residue of the Z-domain was cleaved only when the second amino acid was glycine, alanine, proline, serine, threonine, cysteine or valine (Figure 3.3a). Of these six variants, all but the Val-2 variants showed complete cleavage of the Met₁ residue. For all the other amino acid variants, no cleavage of the Met₁ residue was observed as the Z-domain with the Met₁ residue (the Met-Z form) was mainly detected. These results clearly confirmed the sequence specificity of the *E. coli* MAP, which was previously determined by the N-terminal sequencing of the expressed methionyl-tRNA synthetase mutants differing only by the second amino acid residue and by the computer analysis of the known *E. coli* protein N-terminal sequences.⁸⁹

A substantial amount of the N^α-acetylated Z-domain without Met₁ (the Ac-Z-form) was detected along with the unacetylated Z-domain without Met₁ (the Z-form) when the second amino acid was serine, threonine or valine, each of which is one of the amino acids required for the Met₁ removal in *E. coli*. When the second amino acid was glycine, alanine or proline, only minor N^α-acetylation was observed as the Z-form was mainly detected along with a trace amount of the Ac-Z form despite complete Met₁ removal. This suggests that the RimJ-catalyzed N^α-acetylation occurs only subsequent to the Met₁ cleavage. Notably, only minor acetylation was detected for the Ala-2 variant although the ribosomal protein S5, the natural substrate of RimJ is acetylated at its N-terminal alanine residue. This is probably due to the further influence of the amino acid residues next to the N-terminal group.

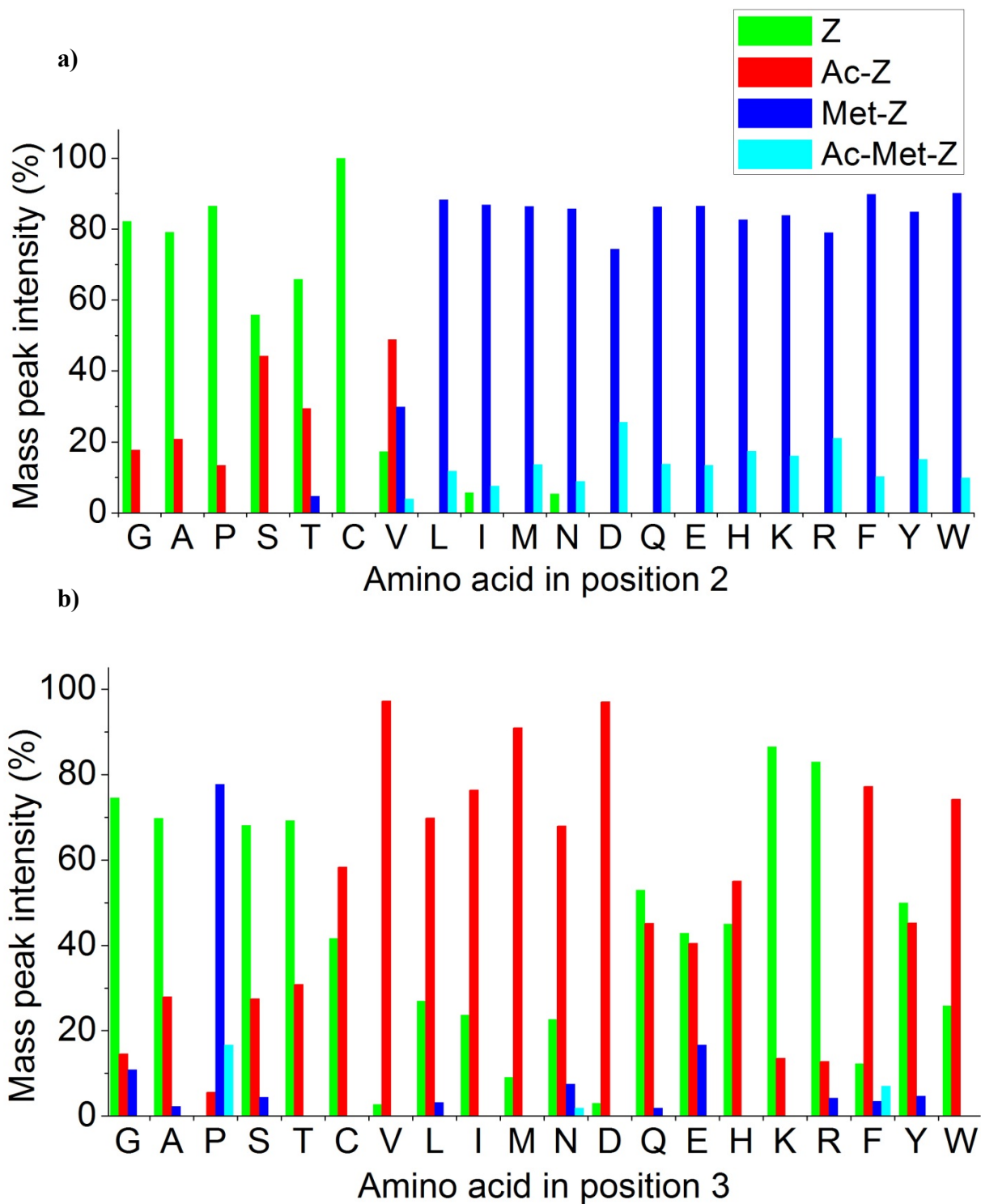


Figure 3.3 Distribution of different N-terminal end forms for the Z-domain variants containing 20 different amino acids: a) in the second position and b) in the third position.

For the Val-2 variant in which the Met₁ residue was incompletely removed, no acetylation was observed when Met₁ remained but N^α-acetylation was detected when Met₁ was removed. When the second amino acid was cysteine, Met₁ was completely removed, but the resulting N-terminal cysteine residue was not acetylated. Instead, the Z-domain forms with +72 and +28 mass shifts were detected, indicating the presence of thiazolidine adducts probably formed during protein expression within the cells by a non-enzymatic reaction of the N-terminal cysteine residue with pyruvate and subsequent decarboxylation, respectively.⁹⁰ It seems unlikely that the lack of N^α-acetylation in the Cys-2 mutants is due to this chemical modification since no increase in N^α-acetylation was observed even when the protein was expressed under conditions that minimized pyruvate formation.⁹⁰ However, it should be noted that N^α-acetylation of the cysteine residue was previously observed for the recombinant human TIMPs and interferons in *E. coli*.^{64,71} Only trace to minor amounts of N^α-acetylation were observed for all other variants with the intact Met₁ residue.

These results clearly suggest that RimJ specifically acetylates the N-terminal serine, threonine or valine residues, subsequent to the Met₁ cleavage by the *E. coli* MAP. In *E. coli*, N^α-acetylation of the serine residue was observed for a few endogenous proteins including the ribosomal protein L12, EF-Tu, and chaperon SecB as well as recombinant human T α fusion proteins.^{45,48,50,72} The ribosomal proteins S18 and S5, the natural substrates of RimI and RimJ, respectively, are acetylated at the N-terminal alanine residue.⁴⁷ The recombinant eglin c was acetylated at the N-terminal threonine residue.⁷³ The preference of small N-terminal amino acid residues by the Rim proteins for N^α-acetylation in *E. coli* is definitely similar to that by NatA, one of three NATs in eukaryotes.⁵ Similar N-terminal sequences specificity for N^α-acetylation was also observed in archaea.³⁵

3.2.2 Third amino acid substitutions

Met₁ removal was little affected by the third amino acid residue in the Z-domain. As shown in Figure 3.3b, Met₁ was completely removed for all but the Pro-3 variants. When the third amino acid was proline, the Met-Z form was mainly detected. This was probably because the structural distortion caused by the proline residue in the third position hindered the *E. coli* MAP from recognizing the Thr-2 residue for the Met₁ cleavage. This result shows to be consistent with previous results on the specificity of *E. coli* MAP in which proline in the second position also caused the Met₁ to be retained.¹¹

The extent of the RimJ-catalyzed N^α-acetylation significantly varied with the nature of the amino acid residue in position 3 (Figure 3.3b). For the Gly-3 variant, the Z-form was mainly detected along with a trace amount of the Ac-Z form. When the third amino acid was proline, a trace amount of the Ac-Z form was also detected along with the major Met-Z form. When the third amino acid was alanine, serine or threonine, a substantial amount of the Ac-Z form was observed although the Z-form was a major product. Almost equal amounts of the Z and Ac-Z forms were detected for the Z-domain variants containing cysteine, glutamine, histidine or tyrosine in position 3. When the third residue was leucine, isoleucine, asparagine, phenylalanine or tryptophan, the Ac-Z form was a major product along with a minor Z-form. For the Met-3 and Val-3 variants, the Ac-Z form was almost exclusively detected. These results suggest that an uncharged, preferably hydrophobic, residue including bulky aromatic amino acids in the penultimate position stimulated the RimJ-mediated N^α-acetylation. This pattern is consistent with our previous observation that Z-domain variant containing Bpa at position 3 showed complete N^α-acetylation. An aromatic amino acid was found next to the N^α-acetylated residue for a few

endogenous proteins in eukaryotes as well as for the recombinant human interferon γ expressed in *E. coli*.^{5,71}

A dramatic difference in N ^{α} -acetylation was observed when a charged residue was in the third position. Whereas very minor N ^{α} -acetylation was evident for the Lys-3 or Arg-3 variants, substantial or almost complete N ^{α} -acetylation was observed for the Glu-3 or Asp-3 variants, respectively. These results clearly suggest that RimJ disfavors a positively charged residue but favors a negatively charged one in the penultimate position. Indeed, many recombinant proteins that were N ^{α} -acetylated in *E. coli* have an acidic residue in position-3: aspartate in interferon A, SLDs, and T α , and glutamate in eglin c.^{70,72-74} This specific preference of a negatively charged penultimate residue on N ^{α} -acetylation was also observed for a number of eukaryotic proteins including the intact Met₁ residue.

3.3 Conclusions

The Met₁ residue of the Z-domain was removed only when the second amino acid was glycine, alanine, proline, serine, threonine, cysteine or valine, consistent with the reported sequence specificity of the *E. coli* MAP. Only subsequent to the Met₁ cleavage, the RimJ-catalyzed N^α-acetylation mainly occurred at the N-terminal serine, threonine, or valine residues. The N^α-acetylation of the Z-domain was significantly decreased by glycine, proline, arginine, or lysine in the penultimate position, but was enhanced by hydrophobic or negatively charged residues in the same position. Practically, this study offers a basis to predict or control Met₁ cleavage and N^α-acetylation of recombinant proteins in *E. coli*, especially when these N-terminal end modifications significantly affect protein stability or activity.

3.4 Materials and Methods

3.4.1 General.

XL1-Blue *E. coli* cells were used for cloning and maintaining plasmids. BL21(DE3) *E. coli* cells were used for protein expression. pET-Z encodes the Z-domain with a C-terminal hexa-histidine tag under control of the T7 promoter. pACYCDuet-RimJ encodes RimJ under control of the T7 promoter. *Taq* DNA polymerase (NEB) was used for polymerase chain reaction (PCR).

3.4.2 Site-directed saturation mutagenesis of the Z-domain gene.

The genes for the Z-domain variants containing all possible 20 amino acids in position 2 or 3 were generated by PCR of pET-Z with the primer ZR (5'-GCAGCCGGATCTCAGTGGTGG-3') and one of the oligonucleotide primers shown in the Table 3.3.

Table 3.3 The Z-domain site directed-mutagenesis oligonucleotide primers.

Primer	Nucleotide sequence
Z2NNK	5'-ATATACATATG NNK AGTGTAGACAACAAAATCAACAAAGAAC-3'
Z2NAK	5'-TATACATATG NAK AGTGTAGACAACAAAATCAACAAAG-3'
Z2HTT	5'-TATACATATG HTT AGTGTAGACAACAAAATCAACAAAG-3'
Z2TAT	5'-GATATACATATG TAT AGTGTAGACAACAAAATCAACAAAG-3'
Z2TGG	5'-TATACATATG TGG AGTGTAGACAACAAAATCAACAAAG-3'
Z3NNK	5'-ATATACATATGACT NNK GTAGACAACAAAATCAACAAAGAACAAC-3'
Z3VAA	5'-TATACATATGACT VAA GTAGACAACAAAATCAACAAAGAAC-3'
Z3TGG	5'-TATACATATGACT TGG GTAGACAACAAAATCAACAAAGAAC-3'
Z3ATG	5'-TATACATATGACT ATG GTAGACAACAAAATCAACAAAGAAC-3'
Z3AAA	5'-GATATACATATGACT AAA GTAGACAACAAAATCAACAAAG-3'

Each of these primers contains either a degenerate or specific codon sequence (in boldface) at the designated position as well as an NdeI restriction site (underlined). The degenerate codon sequences include NNK, NAK, HTT and VAA where N is all four nucleotides; K is G or T; H is

A, C or T; V is A, C or G. Each PCR product was inserted between the NdeI and XhoI sites of pET-Z, and transformed into XL1-Blue *E. coli* cells. Plasmids were isolated from single colonies and sequenced.

3.4.3 Expression and purification of the Z domain variants.

BL21(DE3) *E. coli* cells, co-transformed with each pET-Z variant and pACYCDuet-RimJ were grown overnight at 37°C in the Instant TB autoinduction medium (Novagen) containing 100 µg/mL carbenicillin and 50 µg/mL chloramphenicol. Cells were harvested by centrifugation and lysed by sonication in lysis buffer (50 mM sodium phosphate buffer with 300 mM NaCl, pH 8.0). The Z-domain in the cell lysate was purified with Ni-NTA metal affinity resin (Novagen) under native conditions. Each purified protein was concentrated by ultrafiltration YM-3 Microcon centrifugal filter device (Millipore).

3.4.4 Mass spectrometry.

The Z-domain variants were analyzed by Agilent 6224 Accurate-Mass TOF LC/MS system equipped with an electrospray ionization (ESI) source. Multiple-charged protein ions were deconvoluted by using the MassHunter Workstation software. The deconvoluted mass spectra of the Z-domain variants differing by the amino acid residue in position 2 or 3 are shown in Figures 3.1 and 3.2.

3.4.5 Theoretical mass calculations.

The theoretical mass of each Z-domain variant containing one of 20 amino acids in position 2 or 3 was calculated based on its amino acid sequence:

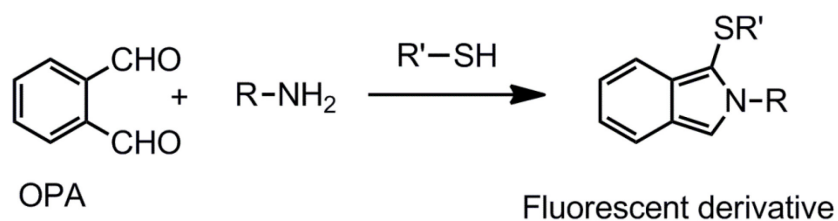
MJOVDNKINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAKKLNDAQA
PKGSHHHHHH where J is one of 20 amino acids and O is S; or J is T and O is one of 20 amino
acids, respectively. The calculated average masses for the different N-terminal end forms of each
Z-domain variant are listed in Tables 3.1 and 3.2.

CHAPTER 4. FLUORESCENCE LABELING OF PROTEINS TO DISTINGUISH N^α-ACETYLATION

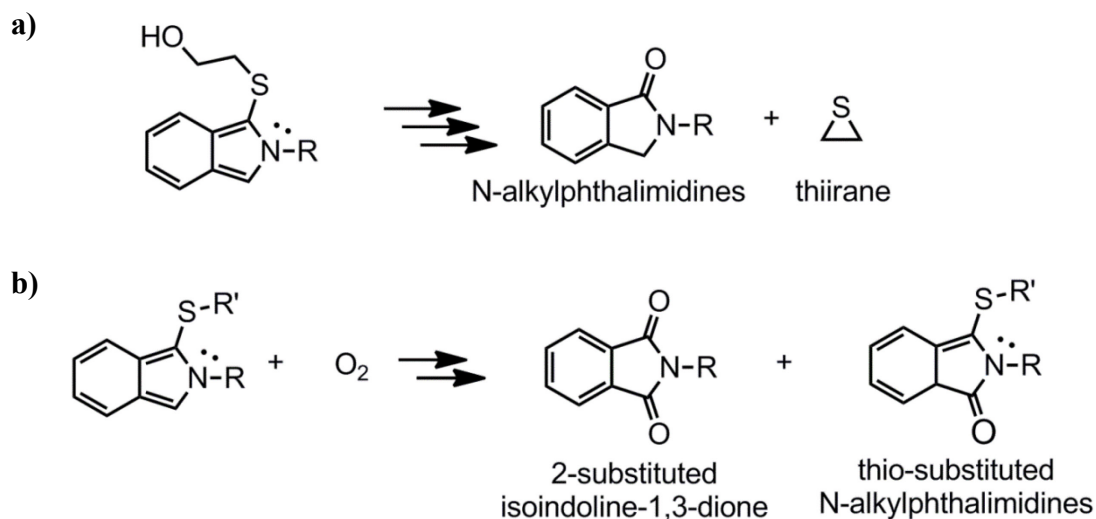
4.1 Introduction

Protein acetylation is typically analyzed by mass spectrometry, by which a 42 Da mass increment can be accurately determined.^{50,64,70,72,74,75,91} Although ideal for the detection of protein acetylation, mass spectrometry of protein samples often requires careful sample preparation and expensive instrumentation. In order to develop a simple and sensitive fluorescence-based method to distinguish the N-terminal acetylation status of proteins, we investigated many fluorogenic reagents including O-phthalaldehyde (OPA) and 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl), which are known to react with primary amino groups.

OPA is a well-known fluorogenic reagent that reacts with primary amines in the presence of a thiol, usually 2-mercaptoethanol (2-ME) to form a 1-alkylthio-2-alkyl substituted isoindole derivative ($\lambda_{\text{ex}}=340$ nm, $\lambda_{\text{em}}=455$ nm) (Scheme 4.1).⁹² The stability and the fluorescence quantum yield of the OPA derivatives depend on several factors including the nature of the amine or thiol in the reaction, and high concentration of OPA.⁹³ The OPA degradation were characterized by Simons and Johnsons in 1978 as N-alkylphthalimidimes and a 3-membered sulfide (Scheme 4.2a).⁹⁴ Autoxidation products were also identified as the major degradation products (Scheme 4.2b).⁹⁵

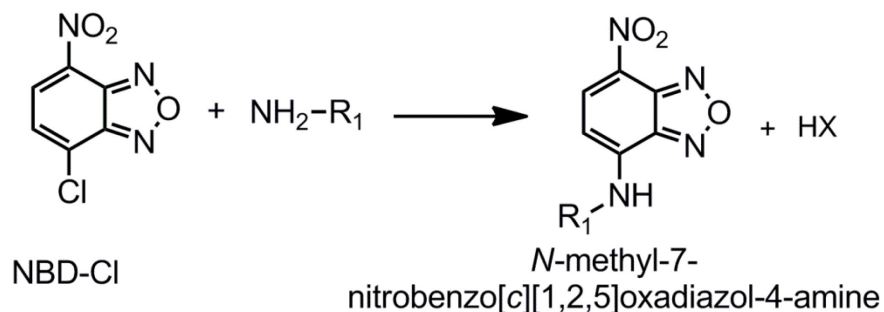


Scheme 4.1 OPA and SH selectively react with primary amines to yield fluorescent adducts.



Scheme 4.2 Isoindole degradation a) obtained from 2-mercaptoethanol and b) autoxidation.

NBD-Cl reacts with aliphatic thiols and amines under alkaline conditions to form highly stable fluorescent derivatives ($\lambda_{ex} = 337 \text{ nm}$, $\lambda_{em} = 515 \text{ nm}$), as shown in Scheme 4.3.⁹⁶



Scheme 4.3 NBD-Cl reacts with primary amines to yield fluorescent adducts.

4.2 Results and discussion

4.2.1 OPA reactions

As we described in Chapter 2, we prepared both the N^α-acetylated and unacetylated forms of the Z-domain, whose acetylation status was confirmed by ESI-MS. In order to distinguish the Z and Ac-Z forms of the Z-domain by fluorescent labeling, we initially tried OPA. When reacted with OPA in the presence of 2-ME in a borate buffer at pH 8, both the Z and Ac-Z forms produced highly fluorescent products within minutes with a maximum emission wavelength at 435 nm (Figure 4.1). However, little difference in the fluorescent intensity was observed for the two forms of the Z-domain. It should be noted that the Z-domain has six internal lysine residues. The solution NMR structure of the Z-domain indicated that all six lysine residues are on the protein surface.⁹⁷ Therefore, the primary amines of these lysine side chains can also react with OPA. In a citrate buffer at neutral pH in which the lysine ε-amino group must be less reactive than the N-terminal amino group since the pKa of the former is much higher than that of the latter, little difference in the fluorescence signals between the Z and Ac-Z forms was observed (Figure 4.2). It was clearly noticeable that the OPA derivatives were very unstable as the fluorescence signal became weaker within 30 minutes.

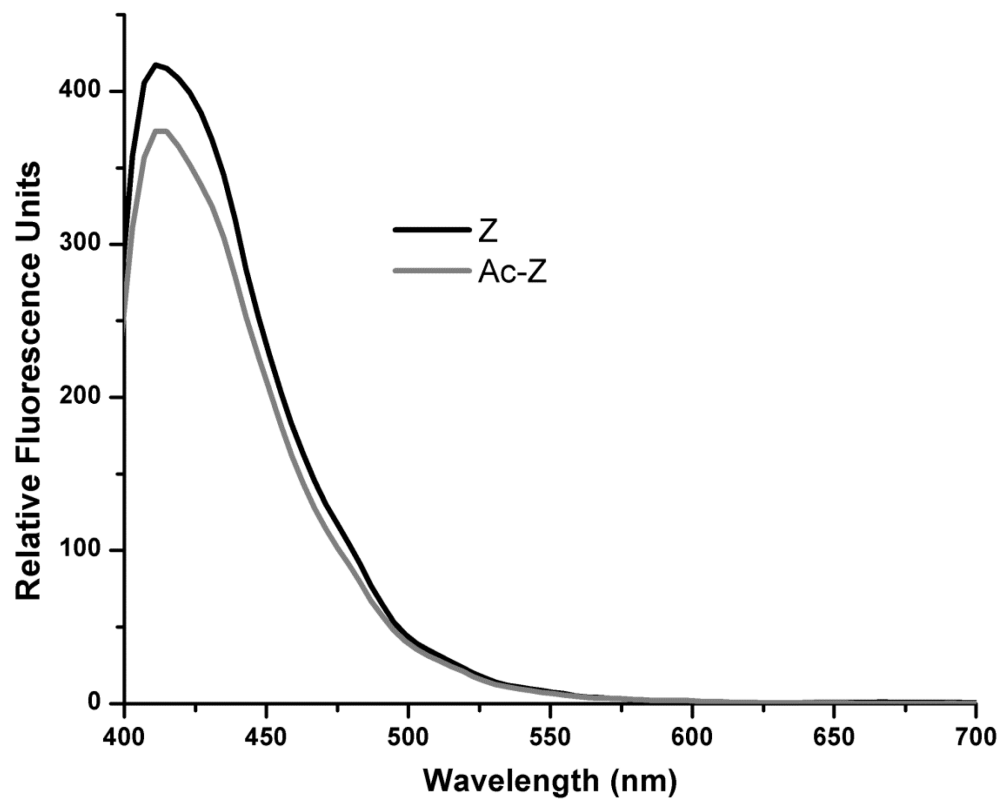


Figure 4.1 Fluorescence spectra for the OPA conjugates of the Ac-Z form and the Z form. Each protein (1.3 μM) was incubated with 50 mM OPA in 100 mM borate buffer, pH 8.0 including 50 mM 2-ME for 5 minutes; $\lambda_{\text{ex}} = 340$ nm.

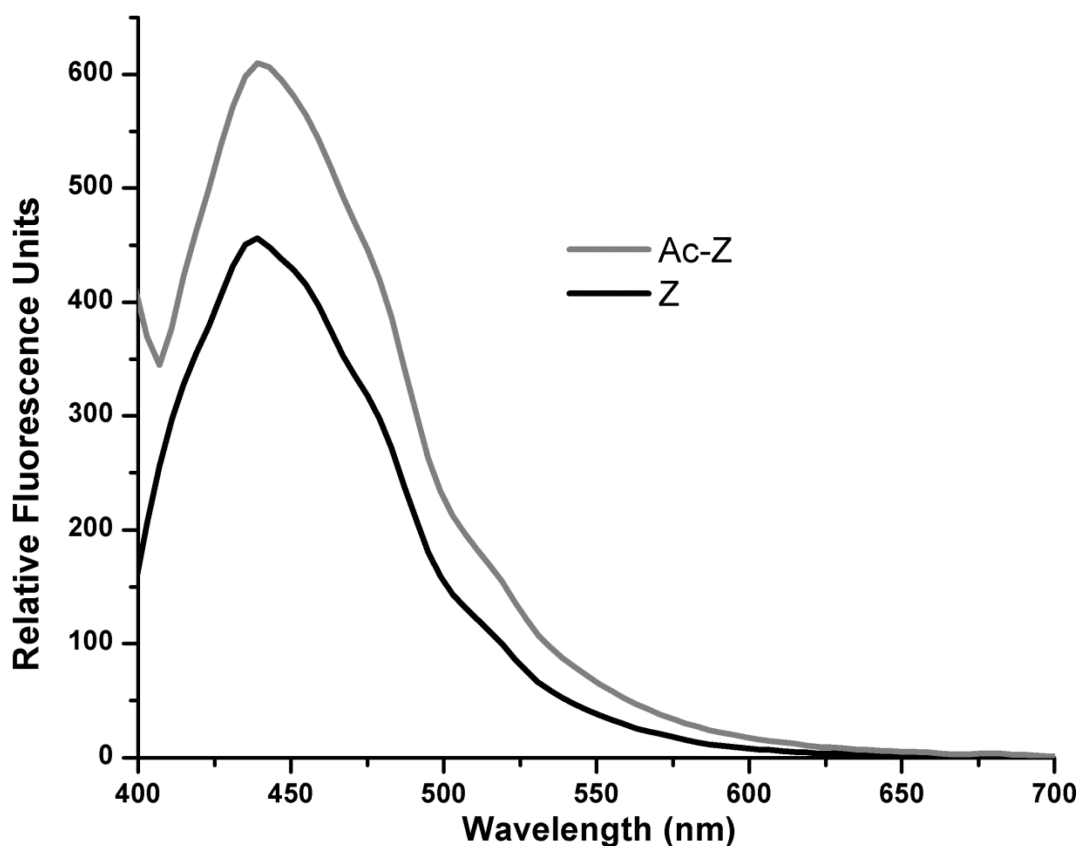
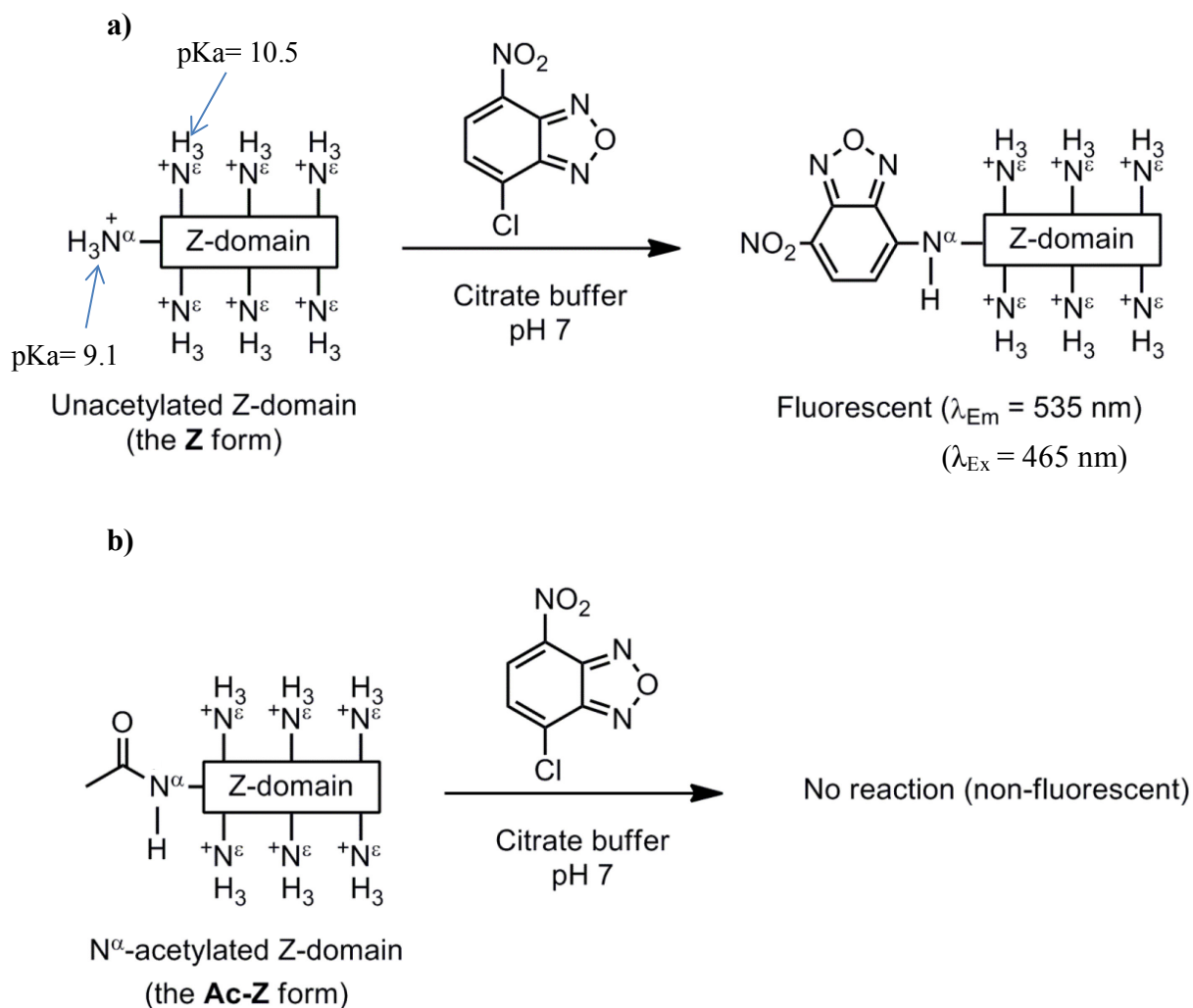


Figure 4.2 Fluorescence spectra for the OPA conjugates of the Ac-Z form and the Z form. Each protein (1.3 μM) was incubated with 50 mM OPA in 50 mM sodium citrate buffer, pH 7.0 including 50 mM 2-ME for 5 minutes; $\lambda_{\text{ex}} = 340$ nm.

4.2.2 NBD-Cl reactions

We next tried NBD-Cl (Scheme 4.4). A stock solution of 50 mM NBD-Cl was prepared in acetonitrile. When 0.5 mM NBD-Cl reacted with 6 μM of the purified Z-domain in 50 mM citrate buffer including 1 mM EDTA, pH 7.0 at room temperature, no immediate reaction was observed. Within 1 hour, however, the Z-form became fluorescent with a maximum fluorescent emission wavelength at 535 nm. In contrast, the Ac-Z form was essentially non-fluorescent. In 7 hours, the fluorescent intensity of the Z form is 28 fold higher than that of the Ac-Z form (Figure

4.3). These results clearly suggest that the free N-terminal amino group in the unacetylated protein selectively reacts with NBD-Cl, whereas the side chain amino groups of the lysine residues are unreactive under these conditions (Scheme 4.4a).



Scheme 4.4 The reaction of NBD-Cl in citrate buffer, pH 7: the unacetylated Z-domain (The Z form), b) N^α -acetylated Z-domain (The Ac-Z form).

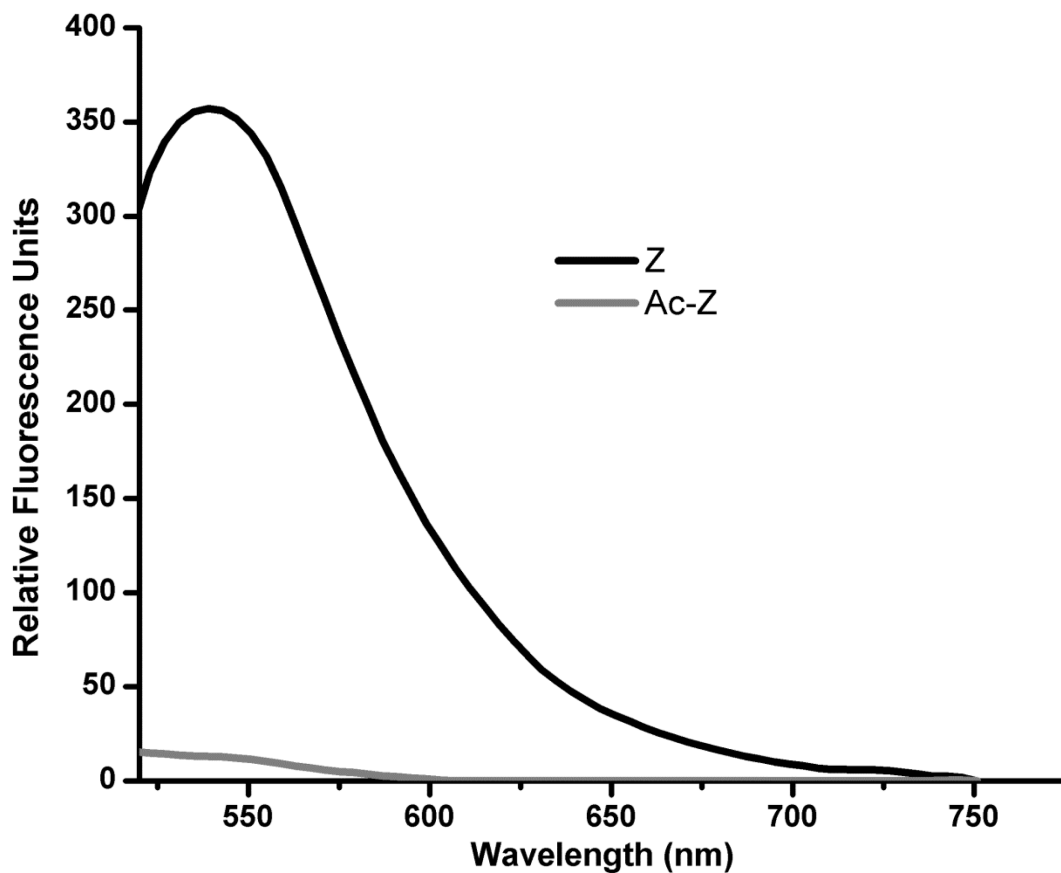


Figure 4.3 Fluorescence spectra for the NBD-Cl conjugates of the Ac-Z form and the Z form. Each protein (6 μM) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA, pH 7.0 for 7 hours.

4.2.2.1 Studies on time-dependent changes in fluorescence of the NBD-Cl reactions

We next evaluated time-dependent changes in fluorescent of the NBD reactions. As illustrated in Figure 4.4, the NBD fluorescence intensities gradually increased over the next 24 hours for the Z-form. Interestingly, no saturation of the fluorescence signal was observed within this timeframe probably due to a slow reaction resulting from mild reactivity of NBD-Cl. In contrast, the fluorescence levels of the Ac-Z form slightly increased over time but reached a plateau within 14 hours. We speculated that this background fluorescence from the Z-form was

due to a slight contamination of the Z-form in the Ac-Z form, possibly resulting from incomplete N^α-acetylation during the Ac-Z form expression. This result clearly suggest that the N^α-acetylated and unacetylated forms can be distinguished by reaction with NBD-Cl over a broad range of time. Also evident was no decrease of fluorescence intensity over a substantial period of time, which indicated a photostability of the probe.

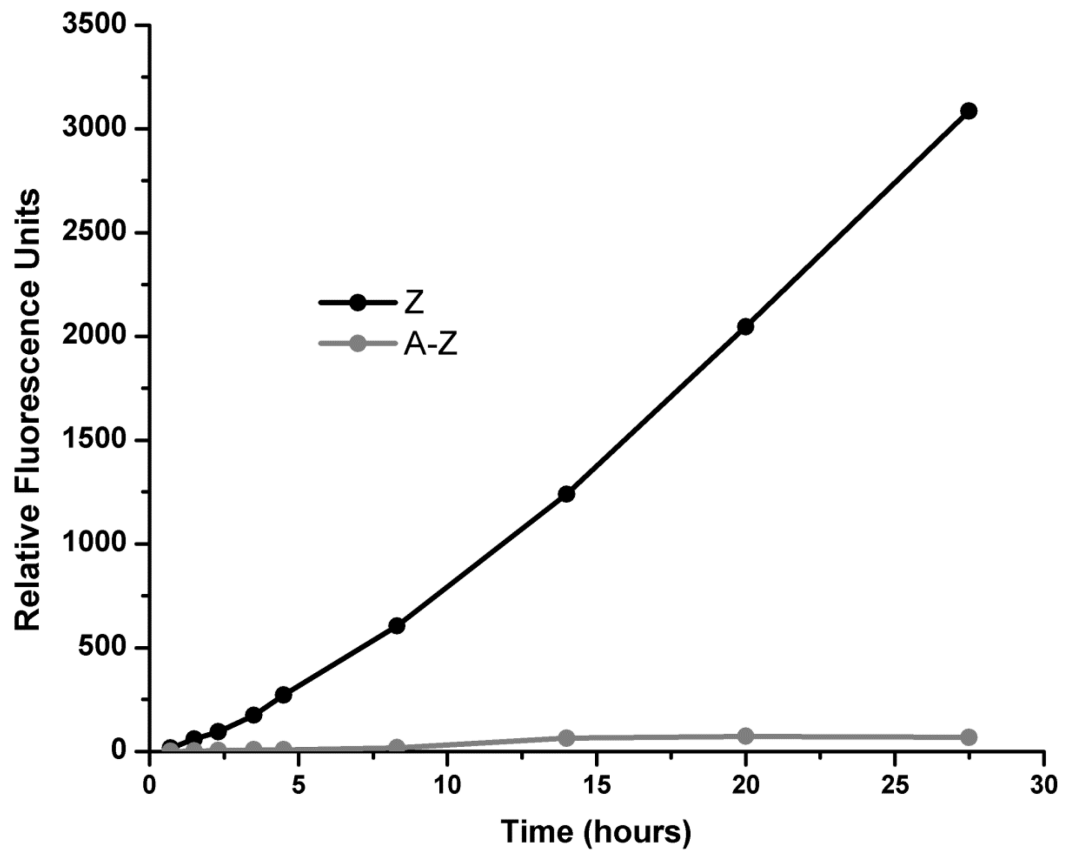


Figure 4.4 Time-dependent NBD-Cl fluorescence intensity changes of the Ac-Z form and the Z form. Each protein (6 μ M) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA at pH 7.0 the indicated times. At each time, an aliquot was removed from the reaction mixture and its fluorescence intensity was measured at 535nm. Results are the average of three independent trials per point.

4.2.2.2 NBD-Cl Fluorescence as a function of different Z-domain protein concentrations

We also examined changes in NBD fluorescence as a function of protein concentrations. As clearly demonstrated in Figure 4.5, fluorescence intensity was linearly proportional to the amount of the Z-domain present in the reactions. Reliable fluorescence emissions were detected using at least 2 μM of the unacetylated form of the Z-domain although a longer reaction time (about 5 hours) was required. Therefore, this method should also be useful to quantitate proteins when their free N-terminal group is available.

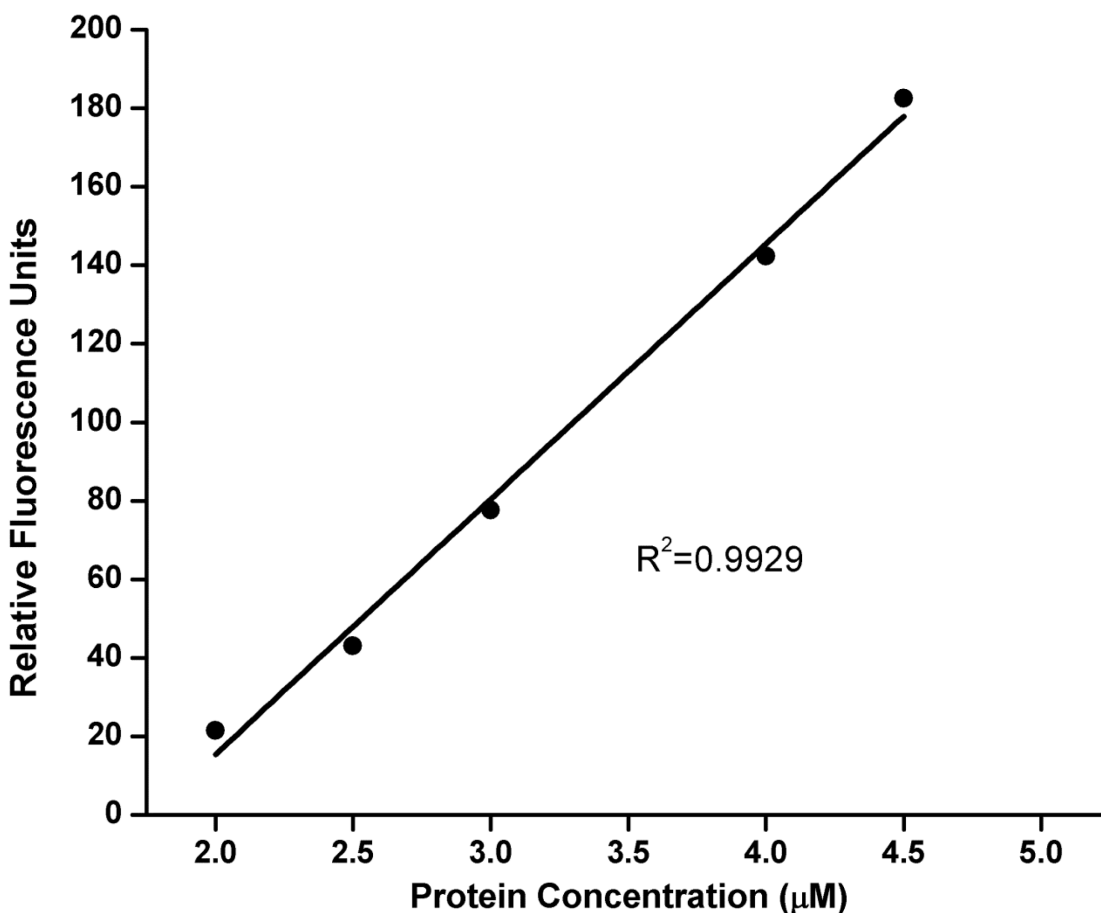


Figure 4.5 Effects of the unacetylated Z-domain concentrations on the NBD-Cl fluorescence emission intensity. The indicated concentrations of protein were incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1 mM EDTA at pH 7.0 for 5.5 h. The fluorescence emission intensity was recorded at 535 nm ($\lambda_{\text{ex}} = 465$ nm). Results are the average of three independent trials per data point.

4.2.2.3 Generality of the NBD-Cl method

To test the generality of this new method, we prepared the N^α-acetylated and unacetylated forms of the thymosin α1-L12 fusion (Tα1-L12) protein according to the procedure reported by Fang *et. al.*⁷⁵ When the Tα1-L12 was expressed from the pET26-Tα1-L12 in JM109(DE3) *E. coli* cells, the ESI-MS of the purified protein gave a major peak at 13200.98 Da and a minor peak at 13243.20 Da, which correspond to the unacetylated form without the initiating fMet₁ residue (the T form, calculated average mass 13200.7 Da) and its N^α-acetylated form (the Ac-T form, calculated average mass 13242.8), respectively (Figure 4.6a). When expressed in BL21(DE3) cells co-transformed with pET26-Tα1-L12 and pACYDuet-RimJ, the Ac-T form was mainly obtained as its ESI-MS showed a single peak at 13242.87 Da (Figure 4.6b).

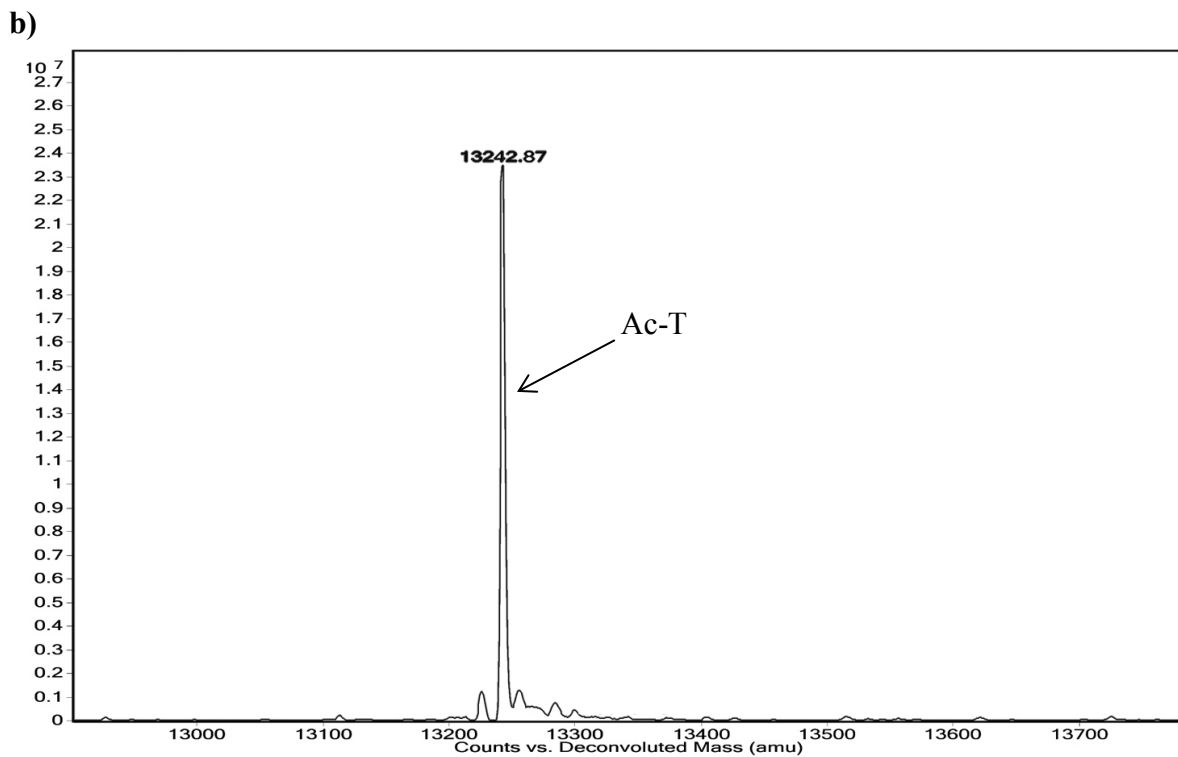
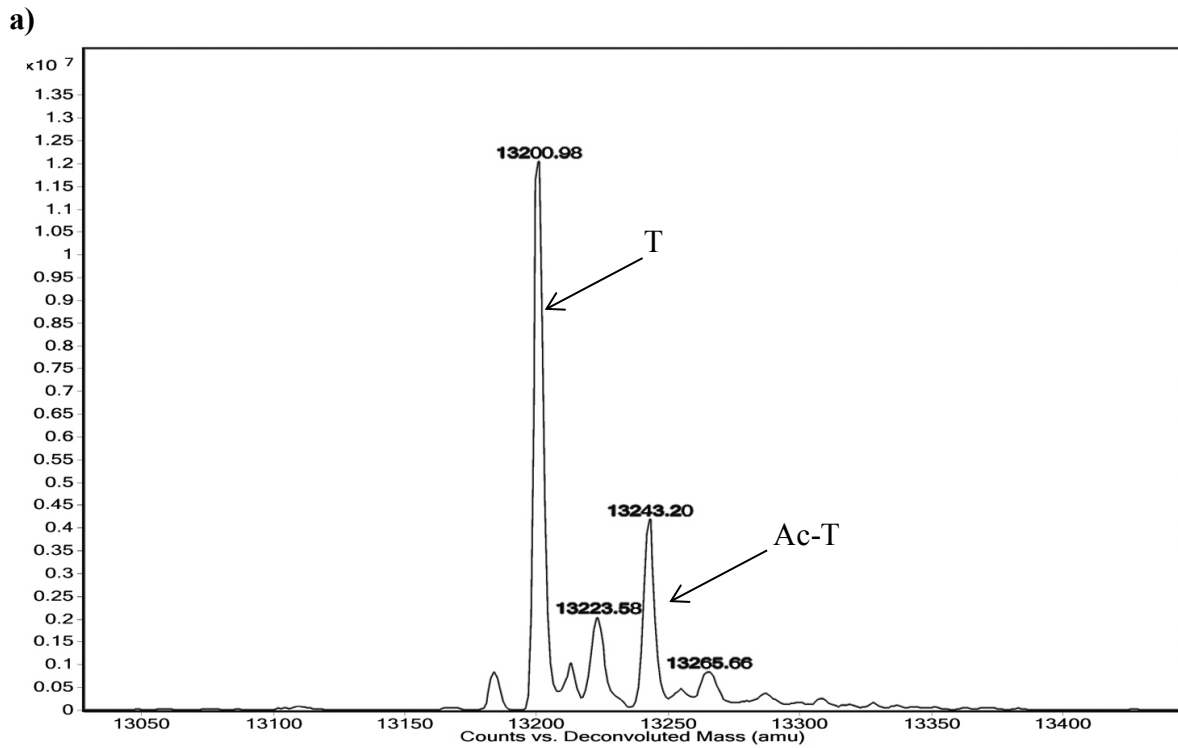


Figure 4.6 Mass-transformed ESI-MS spectra of the thymosin proteins that were expressed in: a) JM109(DE3) cells transformed with pET26-T α 1-L12. b) BL21(DE3) cells co-transformed with pET26-T α 1-L12 and pACYCDuet-RimJ.

As was the case with the Z-domain, the unacetylated form of the T α 1-L12 was highly fluorescent when reacted with NBD-Cl at neutral pH whereas the N ^{α} -acetylated form was much less fluorescent under the same conditions despite the presence of fifteen internal lysine residues (Figure 4.7). The time-dependent fluorescent intensity changes were also consistent with the Z-domain results. As shown in Figure 4.8, the NBD reaction of the T form became highly fluorescent within 1 hour and the fluorescence intensities gradually increased over the next few hours. Within the same timeframe, the fluorescence levels of the Ac-T form reached a saturated background level which probably resulted from a small amount of the T form present in the acetylated protein.

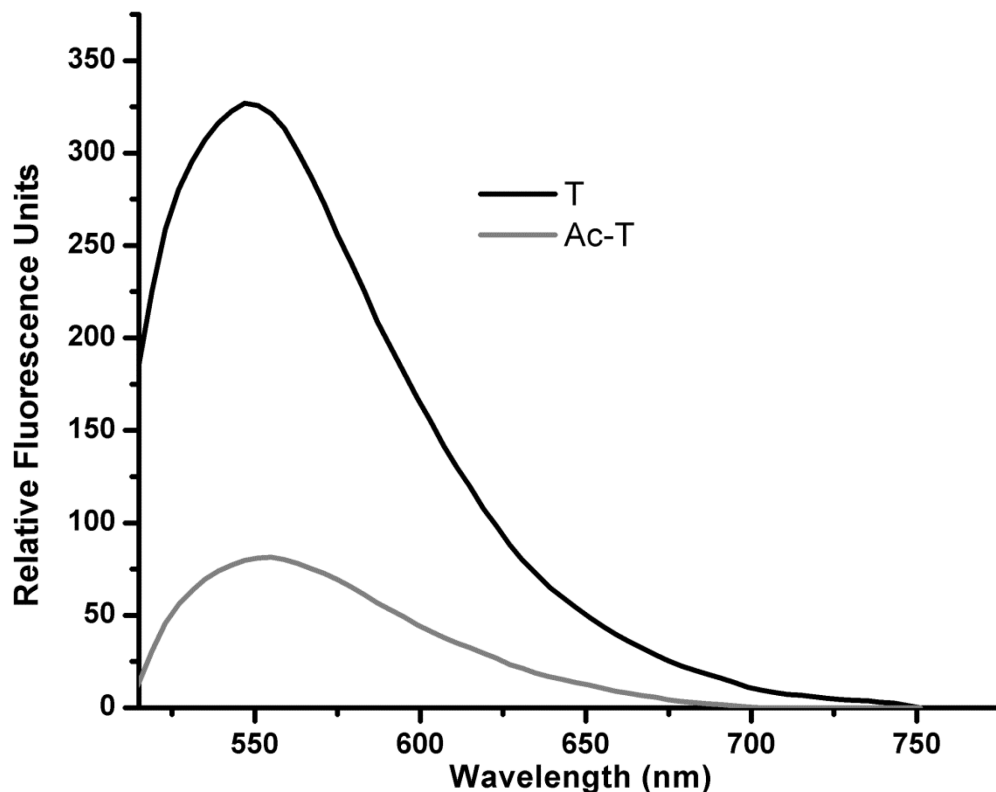


Figure 4.7 Fluorescence spectra for the NBD-Cl conjugates of the N ^{α} -acetylated thymosin α 1-L12 fusion protein (The Ac-T form) and its unacetylated form (the T form). Each protein (6 μ M) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA, pH 7.0 for 5 hours.

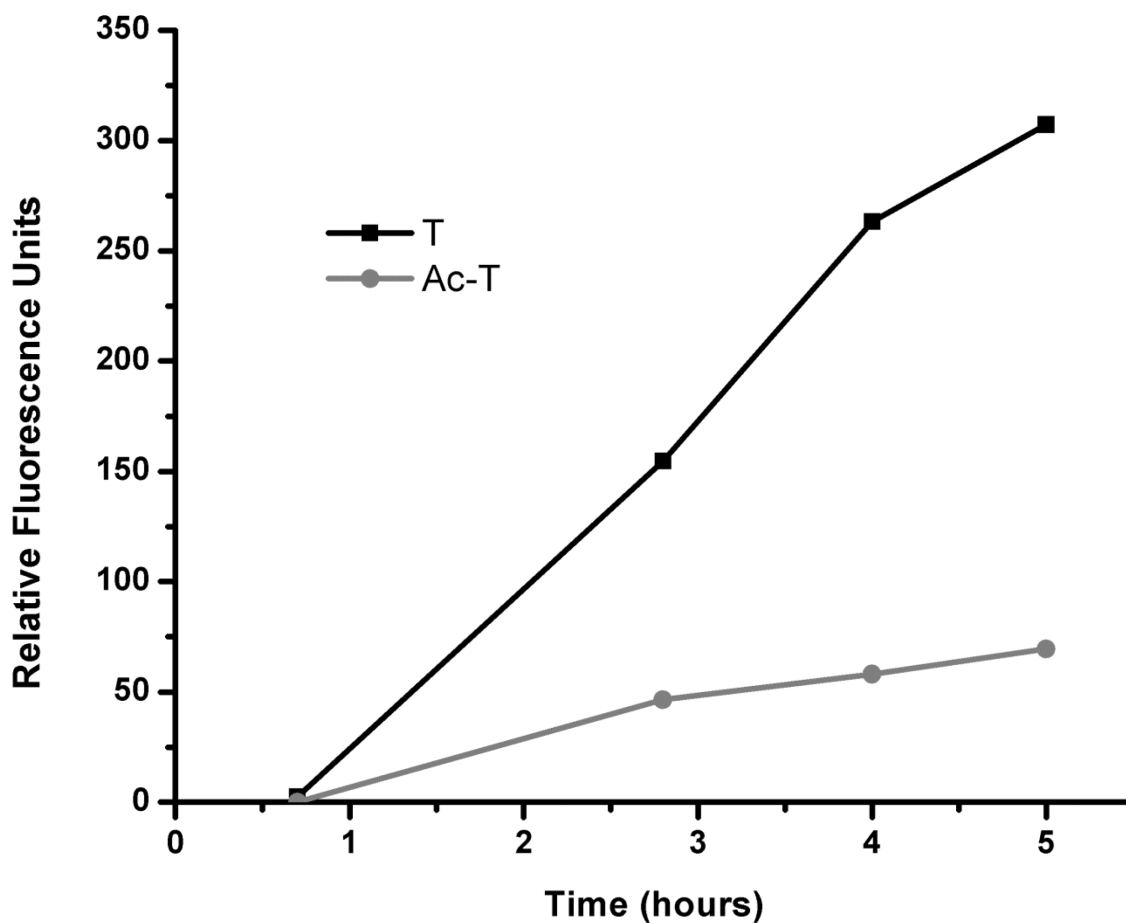


Figure 4.8 Time-dependent NBD-Cl fluorescence intensity changes of the Ac-T form and the T form. Each protein (6 μ M) was incubated with 0.5 mM NBD-Cl in 50 mM sodium citrate buffer including 1mM EDTA at pH 7.0 the indicated times. At each time, an aliquot was removed from the reaction mixture and its fluorescence intensity was measured at 535nm. Results are the average of three independent trials per point.

4.3 Conclusions

Our study provides a simple and sensitive method to distinguish N^α-acetylated protein from the unacetylated form. Despite the presence of many internal lysine residues, only micromolar concentrations of proteins are required for the reliable detection and fluorescence of the NBD labeled proteins which is stable for days. This method provides a general and practical way to quantify proteins when their N-terminal amino group is available. This method is also complementary to the mass spectrometry-based detection of protein N^α-acetylation and should be particularly useful for large scale high-throughput proteomic analysis of protein N^α-acetylation.

4.4 Experimental section

4.4.1 General.

XL1-Blue *E. coli* cells were used for cloning and maintaining plasmids. BL21(DE3) and JM109(DE3) *E. coli* cells were used for protein expression. pET- T α 1-L12 encodes the T α 1-L12 with a C-terminal hexa-histidine tag under control of the T7 promoter. pACYCDuet-RimJ encodes RimJ under control of the T7 promoter. *Taq* DNA polymerase (NEB) was used for polymerase chain reaction (PCR).

4.4.2 Reagents and chemicals

All reagents and solvents were of analytical reagent grade. OPA, NBD-Cl, and 2-ME were purchased from Sigma-Aldrich. A 0.5 M stock solution of OPA was prepared in methanol, and a 0.05 M of NBD-Cl stock solution was prepared in acetonitrile. Solutions were stored at 4° C in the dark.

4.4.3 T α 1-L12 protein expression and purification

Expression of unacetylated pET-T α 1-L12 was performed by transforming the plasmid into JM109(DE3) cell. Acetylated protein was prepared first by co-transformation of pET-T α 1-L12 with pACYDuet-RimJ in BL21(DE3) cells. The cells were grown overnight at 37°C in the instant TB autoinduction medium (Novagen) containing 50 μ g/mL kanamycin (and 50 μ g/mL chloramphenicol for the double transformant). Cells were harvested by centrifugation and lysed by sonication in lysis buffer (50 mM sodium phosphate buffer with 500 mM NaCl, pH 7.5). The T α 1-L12 in the cell lysate was purified with Ni-NTA metal affinity resin (Novagen) under native

conditions. Each purified protein was concentrated by ultrafiltration YM-30 Microcon centrifugal filter device (Millipore).

4.4.4 Mass spectrometry characterization of T α 1-L12

The T α 1-L12 proteins were analyzed by Agilent 6224 Accurate-Mass TOF LC/MS system equipped with an ESI source. Multiple-charged protein ions were deconvoluted by using the MassHunter Workstation software. The deconvoluted mass spectra of the T α 1-L12 are shown in Figure. 4.6.

4.4.5 T α 1-L12 Theoretical mass calculations.

The theoretical mass of T α 1-L12 was calculated based on its amino acid sequence:

MSDAAVDTSSEITTKDLKEKKEVVEEAENGFVSAAAAVAVAAGPVEAAEEKTEFDVI
LKAAGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAG
AEVEVKHHHHHH. The calculated average mass of the T α 1-L12 with the cleavage of the N-terminal methionine (the T form) is 13200.7 Da. The calculated average mass of N^α-acetylated T α 1-L12 with the removal of the N-terminal methionine (the Ac-T form) is 13242.8 Da.

4.4.6 OPA-derivatization reactions

50 μ L reactions in amber microtubes were prepared in 100mM borate buffer, pH 8 or 50 mM sodium citrate buffer, pH 7. 1.3 μ M of Z-domain or the indicated concentrations of amino acids were mixed with 2.5 μ L of 0.05 M 2-ME solution followed by addition of 5 μ L of 0.5 M OPA solution. Reactions were vortexed and fluorescence intensity was measured at the indicated times.

4.4.7 NBD-Cl derivatization reaction

Z-domain or T α 1-L12 proteins were normalized to a concentration of 3 mM. 20 μ L reactions in amber microtubes were prepared in 50 mM sodium citrate buffer containing 1 mM EDTA, pH 7.0. Different protein concentrations in the range of 1.3 μ M- 6.2 μ M were added followed by addition of 500 μ M NBD-Cl (1 μ L of a 5 mM solution).

4.4.8 Fluorescent measurements

Fluorescence spectra were recorded with a Thermo Scientific Nanodrop 300 fluorospectrometer. OPA fluorescence was recorded at an excitation and emission wavelengths of 340 nm and 455 nm, respectively. NBD fluorescence was recorded at an excitation and emission wavelengths of 465 nm and 535 nm, respectively.

References

- (1) 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*, 7342-7372.
- (2) Walsh, C. T. In *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. Roberts & Co., Englewood, CO, **2006**.
- (3) Giglione, C.; Boularot, A.; Meinnel, T. Protein N-terminal methionine excision; *Cell. Mol. Life. Sci.* **2004**, *61*, 1455-1474.
- (4) Lowther, W. T.; Matthews, B. W. Structure and function of the methionine aminopeptidases; *Biochim. Biophys. Acta.* **2000**, *1477*, 157-167.
- (5) Plevoda, B.; Sherman, F. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins; *J. Mol. Biol.* **2003**, *325*, 595-622.
- (6) Meinnel, T.; Mechulam, Y.; Blanquet, S. Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli*; *Biochimie* **1993**, *75*, 1061-1075.
- (7) Giglione, C.; Vallon, O.; Meinnel, T. Control of protein life-span by N-terminal methionine excision; *EMBO. J.* **2003**, *22*, 13-23.
- (8) Liu, S.; Widom, J.; Kemp, C. W.; Crews, C. M.; Clardy, J. Structure of human methionine aminopeptidase-2 complexed with fumagillin; *Science* **1998**, *282*, 1324-1327.
- (9) Roderick, S. L.; Matthews, B. W. Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: a new type of proteolytic enzyme; *Biochemistry* **1993**, *32*, 3907-3912.
- (10) Tahirov, T. H.; Oki, H.; Tsukihara, T.; Ogasahara, K.; Yutani, K.; Libeu, C. P.; Izu, Y.; Tsunasawa, S.; Kato, I. High-resolution crystals of methionine aminopeptidase from *Pyrococcus furiosus* obtained by water-mediated transformation; *J. Struct. Biol.* **1998**, *121*, 68-72.
- (11) 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*, 5588-5599.
- (12) 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*, 263-267.
- (13) Addlagatta, A.; Quillin, M. L.; Omotoso, O.; Liu, J. O.; Matthews, B. W. Identification of an SH3-binding motif in a new class of methionine aminopeptidases from *Mycobacterium tuberculosis* suggests a mode of interaction with the ribosome; *Biochemistry* **2005**, *44*, 7166-7174.

- (14) Alvarado, J. J.; Nemkal, A.; Sauder, J. M.; Russell, M.; Akiyoshi, D. E.; Shi, W.; Almo, S. C.; Weiss, L. M. Structure of a microsporidian methionine aminopeptidase type 2 complexed with fumagillin and TNP-470; *Mol. Biochem. Parasitol.* **2009**, *168*, 158-167.
- (15) D'Souza V, M.; Holz, R. C. The methionyl aminopeptidase from *Escherichia coli* can function as an iron(II) enzyme; *Biochemistry* **1999**, *38*, 11079-11085.
- (16) D'Souza V, M.; Swierczek, S. I.; Cosper, N. J.; Meng, L.; Ruebush, S.; Copik, A. J.; Scott, R. A.; Holz, R. C. Kinetic and structural characterization of manganese(II)-loaded methionyl aminopeptidases; *Biochemistry* **2002**, *41*, 13096-13105.
- (17) Walker, K. W.; Bradshaw, R. A. Yeast methionine aminopeptidase I can utilize either Zn²⁺ or Co²⁺ as a cofactor: a case of mistaken identity?; *Protein. Sci.* **1998**, *7*, 2684-2687.
- (18) Li, J. Y.; Chen, L. L.; Cui, Y. M.; Luo, Q. L.; Li, J.; Nan, F. J.; Ye, Q. Z. Specificity for inhibitors of metal-substituted methionine aminopeptidase; *Biochem. Biophys. Res. Commun.* **2003**, *307*, 172-179.
- (19) Cosper, N. J.; D'Souza V, M.; Scott, R. A.; Holz, R. C. Structural evidence that the methionyl aminopeptidase from *Escherichia coli* is a mononuclear metalloprotease; *Biochemistry.* **2001**, *40*, 13302-13309.
- (20) Wang, J.; Sheppard, G. S.; Lou, P.; Kawai, M.; Park, C.; Egan, D. A.; Schneider, A.; Bouska, J.; Lesniewski, R.; Henkin, J. Physiologically relevant metal cofactor for methionine aminopeptidase-2 is manganese; *Biochemistry.* **2003**, *42*, 5035-5042.
- (21) Lowther, W. T.; Matthews, B. W. Metalloaminopeptidases: common functional themes in disparate structural surroundings; *Chem. Rev.* **2002**, *102*, 4581-4608.
- (22) Chang, S. Y.; McGary, E. C.; Chang, S. Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth; *J. Bacteriol.* **1989**, *171*, 4071-4072.
- (23) Miller, C. G.; Kukral, A. M.; Miller, J. L.; Movva, N. R. pepM is an essential gene in *Salmonella typhimurium*; *J. Bacteriol.* **1989**, *171*, 5215-5217.
- (24) Chang, Y. H.; Teichert, U.; Smith, J. A. Molecular cloning, sequencing, deletion, and overexpression of a methionine aminopeptidase gene from *Saccharomyces cerevisiae*; *J. Biol. Chem.* **1992**, *267*, 8007-8011.
- (25) Li, X.; Chang, Y. H. Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases; *Proc. Natl. Acad. Sci. U S A* **1995**, *92*, 12357-12361.
- (26) Solbiati, J.; Chapman-Smith, A.; Miller, J. L.; Miller, C. G.; Cronan, J. E., Jr. Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal; *J. Mol. Biol.* **1999**, *290*, 607-614.
- (27) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. Mapping protein post-translational modifications with mass spectrometry; *Nat. Methods.* **2007**, *4*, 798-806.

- (28) Driessen, H. P.; de Jong, W. W.; Tesser, G. I.; Bloemendal, H. The mechanism of N-terminal acetylation of proteins; *CRC Crit. Rev. Biochem.* **1985**, *18*, 281-325.
- (29) Bonissone, S.; Gupta, N.; Romine, M.; Bradshaw, R. A.; Pevzner, P. A. N-terminal protein processing: A comparative proteogenomic analysis; *Mol. Cell. Proteomics.* **2012**.
- (30) Arnesen, T.; Van Damme, P.; Polevoda, B.; Helsens, K.; Evjenth, R.; Colaert, N.; Varhaug, J. E.; Vandekerckhove, J.; Lillehaug, J. R.; Sherman, F.; Gevaert, K. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans; *Proc. Natl. Acad. Sci. U S A* **2009**, *106*, 8157-8162.
- (31) Van Damme, P.; Hole, K.; Pimenta-Marques, A.; Helsens, K.; Vandekerckhove, J.; Martinho, R. G.; Gevaert, K.; Arnesen, T. NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation; *PLoS Genet.* **2011**, *7*, e1002169.
- (32) Helbig, A. O.; Gauci, S.; Raijmakers, R.; van Breukelen, B.; Slijper, M.; Mohammed, S.; Heck, A. J. Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome; *Mol. Cell. Proteomics.* **2010**, *9*, 928-939.
- (33) Martinez, A.; Traverso, J. A.; Valot, B.; Ferro, M.; Espagne, C.; Ephritikhine, G.; Zivy, M.; Giglione, C.; Meinnel, T. Extent of N-terminal modifications in cytosolic proteins from eukaryotes; *Proteomics.* **2008**, *8*, 2809-2831.
- (34) Bienvenut, W. V.; Sumpton, D.; Martinez, A.; Lilla, S.; Espagne, C.; Meinnel, T.; Giglione, C. Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features; *Mol. Cell. Proteomics.* **2012**, *11*, M111 015131.
- (35) Falb, M.; Aivaliotis, M.; Garcia-Rizo, C.; Bisle, B.; Tebbe, A.; Klein, C.; Konstantinidis, K.; Siedler, F.; Pfeiffer, F.; Oesterhelt, D. Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey; *J. Mol. Biol.* **2006**, *362*, 915-924.
- (36) Starheim, K. K.; Arnesen, T.; Gromyko, D.; Rynningen, A.; Varhaug, J. E.; Lillehaug, J. R. Identification of the human N(alpha)-acetyltransferase complex B (hNatB): a complex important for cell-cycle progression; *Biochem. J.* **2008**, *415*, 325-331.
- (37) Starheim, K. K.; Gromyko, D.; Evjenth, R.; Rynningen, A.; Varhaug, J. E.; Lillehaug, J. R.; Arnesen, T. Knockdown of human N alpha-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization; *Mol. Cell. Biol.* **2009**, *29*, 3569-3581.
- (38) Song, O. K.; Wang, X.; Waterborg, J. H.; Sternglanz, R. An Nalpha-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A; *J. Biol. Chem.* **2003**, *278*, 38109-38112.

- (39) Evjenth, R.; Hole, K.; Karlsen, O. A.; Ziegler, M.; Arnesen, T.; Lillehaug, J. R. Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity; *J. Biol. Chem.* **2009**, *284*, 31122-31129.
- (40) Yu, M.; Ma, M.; Huang, C.; Yang, H.; Lai, J.; Yan, S.; Li, L.; Xiang, M.; Tan, D. Correlation of expression of human arrest-defective-1 (hARD1) protein with breast cancer; *Cancer. Invest.* **2009**, *27*, 978-983.
- (41) Rope, A. F.; Wang, K.; Evjenth, R.; Xing, J.; Johnston, J. J.; Swensen, J. J.; Johnson, W. E.; Moore, B.; Huff, C. D.; Bird, L. M.; Carey, J. C.; Opitz, J. M.; Stevens, C. A.; Jiang, T.; Schank, C.; Fain, H. D.; Robison, R.; Dalley, B.; Chin, S.; South, S. T.; Pysher, T. J.; Jorde, L. B.; Hakonarson, H.; Lillehaug, J. R.; Biesecker, L. G.; Yandell, M.; Arnesen, T.; Lyon, G. J. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency; *Am. J. Hum. Genet.* **2011**, *89*, 28-43.
- (42) Mullen, J. R.; Kayne, P. S.; Moerschell, R. P.; Tsunasawa, S.; Gribskov, M.; Colavito-Shepanski, M.; Grunstein, M.; Sherman, F.; Sternglanz, R. Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast; *EMBO J.* **1989**, *8*, 2067-2075.
- (43) Hollebeke, J.; Van Damme, P.; Gevaert, K. N-terminal acetylation and other functions of Nalpha-acetyltransferases; *Biol. Chem.* **2012**, *393*, 291-298.
- (44) Starheim, K. K.; Gevaert, K.; Arnesen, T. Protein N-terminal acetyltransferases: when the start matters; *Trends. Biochem. Sci.* **2012**, *37*, 152-161.
- (45) Arai, K.; Clark, B. F.; Duffy, L.; Jones, M. D.; Kaziro, Y.; Laursen, R. A.; L'Italien, J.; Miller, D. L.; Nagarkatti, S.; Nakamura, S.; Nielsen, K. M.; Petersen, T. E.; Takahashi, K.; Wade, M. Primary structure of elongation factor Tu from Escherichia coli; *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1326-1330.
- (46) Dekker, C.; de Kruijff, B.; de Korte-Kool, G.; Kroon, J.; Gros, P. Crystals of acetylated SecB diffract to 2.3-A resolution; *J. Struct. Biol.* **1999**, *128*, 237-242.
- (47) Yoshikawa, A.; Isono, S.; Sheback, A.; Isono, K. Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12; *Mol. Gen. Genet.* **1987**, *209*, 481-488.
- (48) Tanaka, S.; Matsushita, Y.; Yoshikawa, A.; Isono, K. Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12; *Mol. Gen. Genet.* **1989**, *217*, 289-293.
- (49) Jones, M. D.; Petersen, T. E.; Nielsen, K. M.; Magnusson, S.; Sottrup-Jensen, L.; Gausing, K.; Clark, B. F. The complete amino-acid sequence of elongation factor Tu from Escherichia coli; *Eur. J. Biochem.* **1980**, *108*, 507-526.
- (50) Smith, V. F.; Schwartz, B. L.; Randall, L. L.; Smith, R. D. Electrospray mass spectrometric investigation of the chaperone SecB; *Protein. Sci.* **1996**, *5*, 488-494.

- (51) Vetting, M. W.; Bareich, D. C.; Yu, M.; Blanchard, J. S. Crystal structure of RimI from *Salmonella typhimurium* LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18; *Protein. Sci.* **2008**, *17*, 1781-1790.
- (52) Vetting, M. W.; LP, S. d. C.; Yu, M.; Hegde, S. S.; Magnet, S.; Roderick, S. L.; Blanchard, J. S. Structure and functions of the GNAT superfamily of acetyltransferases; *Arch. Biochem. Biophys.* **2005**, *433*, 212-226.
- (53) Hwang, C. S.; Shemorry, A.; Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals; *Science* **2010**, *327*, 973-977.
- (54) Hershko, A.; Heller, H.; Eytan, E.; Kaklij, G.; Rose, I. A. Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown; *Proc. Natl. Acad. Sci. U S A* **1984**, *81*, 7021-7025.
- (55) Bloom, J.; Amador, V.; Bartolini, F.; DeMartino, G.; Pagano, M. Proteasome-mediated degradation of p21 via N-terminal ubiquitinylation; *Cell* **2003**, *115*, 71-82.
- (56) Ben-Saadon, R.; Fajerman, I.; Ziv, T.; Hellman, U.; Schwartz, A. L.; Ciechanover, A. The tumor suppressor protein p16(INK4a) and the human papillomavirus oncoprotein-58 E7 are naturally occurring lysine-less proteins that are degraded by the ubiquitin system. Direct evidence for ubiquitination at the N-terminal residue; *J. Biol. Chem.* **2004**, *279*, 41414-41421.
- (57) Ciechanover, A.; Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in; *Trends. Cell. Biol.* **2004**, *14*, 103-106.
- (58) Coulton, A. T.; East, D. A.; Galinska-Rakoczy, A.; Lehman, W.; Mulvihill, D. P. The recruitment of acetylated and unacetylated tropomyosin to distinct actin polymers permits the discrete regulation of specific myosins in fission yeast; *J. Cell. Sci.* **2010**, *123*, 3235-3243.
- (59) Guo, L.; Munzberg, H.; Stuart, R. C.; Nilni, E. A.; Bjorbaek, C. N-acetylation of hypothalamic alpha-melanocyte-stimulating hormone and regulation by leptin; *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 11797-11802.
- (60) Mima, J.; Kondo, T.; Hayashi, R. N-terminal acetyl group is essential for the inhibitory function of carboxypeptidase Y inhibitor (I(C)); *FEBS Lett.* **2002**, *532*, 207-210.
- (61) Scott, D. C.; Monda, J. K.; Bennett, E. J.; Harper, J. W.; Schulman, B. A. N-Terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex; *Science* **2011**, *334*, 674-678.
- (62) Manning, L. R.; Manning, J. M. The acetylation state of human fetal hemoglobin modulates the strength of its subunit interactions: long-range effects and implications for histone interactions in the nucleosome; *Biochemistry* **2001**, *40*, 1635-1639.
- (63) Ashiuchi, M.; Yagami, T.; Willey, R. J.; Padovan, J. C.; Chait, B. T.; Popowicz, A.; Manning, L. R.; Manning, J. M. N-terminal acetylation and protonation of individual hemoglobin subunits: position-dependent effects on tetramer strength and cooperativity; *Protein. Sci.* **2005**, *14*, 1458-1471.

- (64) Van Doren, S. R.; Wei, S.; Gao, G.; DaGue, B. B.; Palmier, M. O.; Bahudhanapati, H.; Brew, K. Inactivation of N-TIMP-1 by N-Terminal Acetylation When Expressed in Bacteria; *Biopolymers* **2008**, *89*, 960-968.
- (65) Ogawa, H.; Gomi, T.; Takata, Y.; Date, T.; Fujioka, M. Recombinant expression of rat glycine N-methyltransferase and evidence for contribution of N-terminal acetylation to cooperative binding of S-adenosylmethionine; *Biochem. J.* **1997**, *327 (Pt 2)*, 407-412.
- (66) Jackson, C. L. N-terminal acetylation targets GTPases to membranes; *Nat. Cell. Biol.* **2004**, *6*, 379-380.
- (67) Behnia, R.; Panic, B.; Whyte, J. R.; Munro, S. Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p; *Nat. Cell. Biol.* **2004**, *6*, 405-413.
- (68) Setty, S. R.; Strohlic, T. I.; Tong, A. H.; Boone, C.; Burd, C. G. Golgi targeting of ARF-like GTPase Arl3p requires its N-terminal acetylation and the integral membrane protein Sys1p; *Nat. Cell. Biol.* **2004**, *6*, 414-419.
- (69) Forte, G. M.; Pool, M. R.; Stirling, C. J. N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum; *PLoS Biol.* **2011**, *9*, e1001073.
- (70) Charbaut, E.; Redeker, V.; Rossier, J.; Sobel, A. N-terminal acetylation of ectopic recombinant proteins in Escherichia coli; *FEBS Lett.* **2002**, *529*, 341-345.
- (71) Honda, S.; Asano, T.; Kajio, T.; Nishimura, O. Escherichia coli-derived human interferon-gamma with Cys-Tyr-Cys at the N-terminus is partially N-terminal acetylated; *Arch. Biochem. Biophys.* **1989**, *269*, 612-622.
- (72) Wu, J.; Chang, S.; Gong, X.; Liu, D.; Ma, Q. Identification of N-terminal acetylation of recombinant human prothymosin alpha in Escherichia coli; *Biochim. Biophys. Acta.* **2006**, *1760*, 1241-1247.
- (73) Grütter, M. G.; Märki, W.; Walliser, H.-P. Crystals of the Complex between Recombinant N-Acetylglycyl-L-histidyl-L-leucyl-L-phenylalanine and Subtilisin; *J. Biol. Chem.* **1985**, *260*, 11436-11437.
- (74) Takao, T.; Kobayashi, M.; Nishimura, O.; Shimo, Y. Chemical Characterization of Recombinant Human Leukocyte Interferon A Using Fast Atom Bombardment Mass Spectrometry; *J. Biol. Chem.* **1987**, *262*, 3541-3547.
- (75) Fang, H.; Zhang, X.; Shen, L.; Si, X.; Ren, Y.; Dai, H.; Li, S.; Zhou, C.; Chen, H. RimJ is responsible for N-terminal acetylation of thymosin α 1 in Escherichia coli; *Appl. Microbiol. Biotechnol.* **2009**, *84*, 99-104.
- (76) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications; *Nat. Biotechnol.* **2003**, *21*, 255-261.
- (77) Parker, C. E.; Mocanu, V.; Mocanu, M.; Dicheva, N.; Warren, M. R. In *Mass Spectrometry for Post-Translational Modifications*; CRC Press: Boca Raton, **2010**.

- (78) Seo, J.; Lee, K.-J. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches; *J. Biochem. Mol. Biol.* **2004**, *37*, 35-44.
- (79) Reid, G. E.; McLuckey, S. A. 'Top down' protein characterization via tandem mass spectrometry; *J. Mass. Spectrom.* **2002**, *37*, 663-675.
- (80) Lin, H.; Cornish, V. W. Screening and selection methods for large-scale analysis of protein function; *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 4402-4425.
- (81) Nord, K.; Gunneriusson, E.; Ringdahl, J.; Stahl, S.; Uhlen, M.; Nygren, P. A. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain; *Nat. Biotechnol.* **1997**, *15*, 772-777.
- (82) Tashiro, M.; Tejero, R.; Zimmerman, D. E.; Celda, B.; Nilsson, B.; Montelione, G. T. High-resolution solution NMR structure of the Z domain of staphylococcal protein A; *J. Mol. Biol.* **1997**, *272*, 573-590.
- (83) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli *Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 11020-11024.
- (84) Barak, R.; Yan, J.; Shainskaya, A.; Eisenbach, M. The chemotaxis response regulator CheY can catalyze its own acetylation; *J. Mol. Biol.* **2006**, *359*, 251-265.
- (85) Waegeman, H.; Soetaert, W. Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering; *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 1891-1910.
- (86) Ryu, Y.; Schultz, P. G. Efficient incorporation of unnatural amino acids into proteins in Escherichia coli; *Nat. Methods.* **2006**, *3*, 263-265.
- (87) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products; *Proc. Natl. Acad. Sci. U S A* **2000**, *97*, 6640-6645.
- (88) 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*, 1128-1130.
- (89) Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid; *Proc. Natl. Acad. Sci. U S A* **1989**, *86*, 8247-8251.
- (90) 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*, 658-663.
- (91) Bariola, P. A.; Russell, B. A.; Monahan, S. J.; Stroop, S. D. Identification and quantification of N alpha-acetylated Y. pestis fusion protein F1-V expressed in Escherichia coli using LCMS E; *J. Biotechnol.* **2007**, *130*, 11-23.

- (92) Roth, M. Fluorescence reaction for amino acids; *Anal. Chem.* **1971**, *43*, 880-882.
- (93) Garcia Alvarez-Coque, M. C.; Medina Hernandez, M. J.; Villanueva Camanas, R. M.; Mongay Fernandez, C. Formation and instability of o-phthalaldehyde derivatives of amino acids; *Anal. Biochem.* **1989**, *178*, 1-7.
- (94) S. Stoney Simons Jr.; Johnson, D. F. Reaction of o-phthalaldehyde and thiols with primary amines: formation of 1-alkyl(and aryl)thio-2-alkylisoindoles; *J. Org. Chem.* **1978**, *43*, 2886-2891.
- (95) Stobaugh, J. F.; Repta, A. J.; Sternson, L. A.; Garren, K. W. Factors affecting the stability of fluorescent isoindoles derived from reaction of o-phthalaldehyde and hydroxyalkylthiols with primary amines; *Anal. Biochem.* **1983**, *135*, 495-504.
- (96) Bernal-Perez, L. F.; Prokai, L.; Ryu, Y. Selective N-terminal fluorescent labeling of proteins using 4-chloro-7-nitrobenzofurazan: A method to distinguish protein N-terminal acetylation; *Anal. Biochem.* **2012**, *428*, 13-15.
- (97) Zheng, D.; Aramini, J. M.; Montelione, G. T. Validation of helical tilt angles in the solution NMR structure of the Z domain of Staphylococcal protein A by combined analysis of residual dipolar coupling and NOE data; *Protein. Sci.* **2004**, *13*, 549-554.

VITA

LINA F. BERNAL-PEREZ

EDUCATION

Ph.D. Chemistry, Texas Christian University, Fort Worth, TX, 2012
Research advisor: Dr. Youngha Ryu

B.S. Biochemistry, Texas Wesleyan University, Fort Worth, TX, 2005
Advisor: Dr. Ricardo Rodriguez

RESEARCH EXPERIENCE

Associate Scientist, Alcon Laboratories, Inc., Fort Worth, TX,
March 2007- January 2008

Quality Assurance Chemist, Alcon Laboratories, Inc., Fort Worth, TX,
October 2006- March 2007.

Associate Scientist, Alcon Laboratories, Inc., Fort Worth, TX,
August 2006- September 2006

Chemistry Intern, Alcon Laboratories, Inc., Fort Worth, TX,
October 2004- December 2005

AWARDS

1. Air Force Summer Faculty Fellowship Program (SFFP) Accompanying Graduate Student Award, 2011 & 2012.
2. 1st Place Graduate Research Poster in Chemistry, The Michael and Sally McCracken SRS, 2011.
3. Dean's Graduate Student Teaching Award in Chemistry, TCU, 2010.
4. *Cum laude*, Texas Wesleyan University, 2005.
5. American Chemical Society Award, 2005.
6. Academic Dean's List, Texas Wesleyan University, 2004 and 2005.
7. National Dean's List, Texas Wesleyan University, 2002-2003, 2003-2004, and 2004-2005.

PUBLICATIONS

1. Budhathoki, P.; Bernal-Perez, L.F.; Annunziata, O.; Ryu, Y. "Rationally-designed fluorescent lysine riboswitch probes." *Org. Biomol. Chem.* **2012**, *10*, 7872-7874.
2. Bernal-Perez, L.F.; Prokai, L.; Ryu, Y. "Selective N-terminal fluorescent labeling of proteins: a method to distinguish protein N-terminal acetylation." *Anal. Biochem.* **2012**, *428*, 13-15.
3. 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*, 1128-1130.

4. Scoper, S. V.; Kabat, A. G.; Owen, G. R.; Stroman, D. W.; Kabra, B. P.; Faulkner, R.; Kulshreshtha, A. K.; Rusk, C.; Bell, B.; Jamison, T.; Bernal-Perez, L. F.; Brooks, A. C.; Nguyen, V. A. "Ocular distribution, bactericidal activity and settling characteristics of TobraDex® ST ophthalmic suspension compared with TobraDex® ophthalmic suspension." *Adv. Ther.* **2008**, *25*, 77-88.

PATENTS

1. Owen, G. R.; Brooks, A. C.; Bernal-Perez, L. F.; Stroman, D.W.; Dajcs, J.J. "Fluoroquinolone derivatives for ophthalmic applications." WO/2009/103053.

CONFERENCE PRESENTATIONS

1. Lee, Trang; Bernal-Perez, L.F.; Ryu, Y.; "Amino acid sequence requirements for N-terminal acetylation of the Z-domain protein in *E.coli*." The Michael and Sally McCracken Student Research Symposium (SRS); Fort Worth, Texas; April, 2012.
2. Budhathoki, P.S.; Bernal-Perez, L.F.; Ryu, Y.; "Structure-based rational design and synthesis of fluorescent lysine riboswitch probes." The Michael and Sally McCracken Student Research Symposium (SRS); Fort Worth, Texas; April, 2011.
3. Bernal-Perez, L.F.; Prokai, L.; Ryu, Y.; "Fluorescence-based detection of N-terminal protein acetylation." The Michael and Sally McCracken SRS; Fort Worth, Texas; April, 2011.
4. Bernal-Perez, L.F.; Prokai, L.; Ryu, Y.; "Fluorescence-based detection of N-terminal protein acetylation." 44th Annual Meeting in Miniature; Stephenville, Texas; April, 2011.
5. Ryu, Y.; Bernal-Perez, L.F.; Sahyouni, F.; Prokai, L.; "Context-dependent N-terminal acetylation of recombinant proteins in *Escherichia coli*." The 24th Annual Symposium of The Protein Society; San Diego, California; August, 2010.
6. Ryu, Y.; Bernal-Perez, L.F.; Sahyouni, F.; Prokai, L.; "Context-dependent N-terminal acetylation of recombinant proteins in *Escherichia coli*." 43th Annual Meeting in Miniature; Dallas, Texas; April, 2010.
7. Bernal-Perez, L.F.; Morales-Alfaro, A.; Ryu, Y.; "Photochemical and genetic approaches to identify recombinant protein N-acetyltransferases in *E. coli*." SRS; Fort Worth, Texas; April, 2010.
8. Budhathoki, P.S.; Bernal-Perez, L.F.; Ryu, Y.; "*In vitro* selection of peptides that recognize riboswitch conformational changes from combinatorial phage display libraries." SRS; Fort Worth, Texas; April, 2009.
9. Owen G.R.; Brooks A.C.; Bernal-Perez L.F.; Campbell-Furtick M.B. "The ocular penetration of antibiotics using a rabbit model that emulates human topical dosing." EHRlich II Conference; Nurnberg, Germany; October, 2008.
10. Brubaker, M.J.; Bernal-Perez, L.F.; Brooks, A.C.; Owen, G.R. "The correlation of an innovative *in vivo* dosing method with clinical trial results of the ocular penetration of dexamethasone using a new TobraDex formulation." Association for Research in Vision and Ophthalmology (ARVO); May, 2008.
11. Tran, D.; Bernal-Perez, L.F., Ryu, Y. "Directed evolution of leucyl-tRNA synthetase for incorporation of unnatural amino acids into proteins." SRS; Fort Worth, Texas; April, 2008.
12. Owen, G. R.; Brooks, A. C.; Bernal-Perez, L. F.; "The ocular distribution and kinetics of moxifloxacin following prophylactic dosing regimens and an intracameral injection in rabbits." The Ocular Microbiology and Immunology Group (OMIG) 41st Annual Meeting; New Orleans, Louisiana; November, 2007.

ABSTRACT

STUDIES ON PROTEIN N-TERMINAL ACETYLATION IN BACTERIA

By

Lina Fernanda Bernal-Perez, Ph.D., 2012

Department of Chemistry

Texas Christian University

Dissertation Advisor: Dr. Youngha Ryu

N-terminal (N^{α}) protein acetylation, one of the most common post-translational modifications in eukaryotes, plays a pivotal role in the stability, activity and targeting of certain proteins (Chapter 1). This protein modification is significantly less frequent in prokaryotes. In *Escherichia coli*, the only N^{α} -acetyltransferases identified are RimI, RimJ, and RimL, which acetylate the ribosomal proteins S18, S5 and L7/L12, respectively. Although most eukaryotic proteins are not acetylated when ectopically expressed in *E. coli*, partial or complete N^{α} -acetylation has been reported for several recombinant proteins. Just recently, it was demonstrated that N^{α} -acetylation of the thymosin $\alpha 1$ fusion proteins is catalyzed by RimJ. For most other proteins, however, the underlying mechanism of N^{α} -acetylation remains unknown.

We recently observed that the Z-domain protein, a small three-helix bundle protein derived from the *Staphylococcal* protein A, is N^{α} -acetylated only under certain conditions. We decided to use the Z-domain as a model protein to study the N^{α} -acetylation in *E. coli*. We revealed that the N^{α} -acetylation of the Z-domain depends on the *E. coli* strains, expression vectors and amino acid residues near the N-terminus, and is enhanced by high cellular levels of RimJ (Chapter 2). In order to systematically study the sequence dependence of the N-terminal methionine cleavage

and RimJ-mediated N^α-acetylation *in E. coli*, the Z-domain variants differing by the second or third amino acid residue were expressed and analyzed by mass spectrometry (Chapter 3). The initiating methionine residue of the Z-domain was removed only when a small and uncharged amino acid residue was in the second position. Only subsequent to the cleavage of the initiating methionine residue, the RimJ-catalyzed N-terminal acetylation mainly occurred at the N-terminal serine and threonine residues and was significantly enhanced by a hydrophobic or negatively charged residue in the penultimate position.

Although primarily used for analysis of N-terminal acetylation, mass spectrometry often requires careful sample preparation and expensive instrumentation. Therefore, in order to find a simple and sensitive method to analyze the acetylation status of proteins, we developed a fluorogenic derivatization method using 4-chloro-7-nitrobenzofurazan (NBD-Cl) (Chapter 4). The unacetylated protein selectively reacted with NBD-Cl at neutral pH to provide high fluorescence. In contrast, the N^α-acetylated protein was essentially non-fluorescent under the same conditions despite the presence of many internal lysine residues. This method should be particularly useful for a large scale high-throughput proteomic analysis of protein N^α-acetylation.