

# New Methods for Synthesis and Modification of Peptides and Proteins

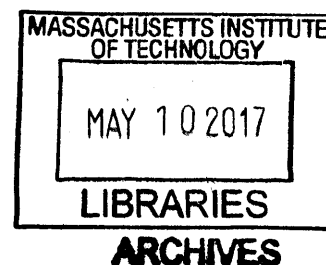
by  
Alexander Alexandrovich Vinogradov

Specialist in Chemistry  
Lomonosov Moscow State University, Russia, 2012

Submitted to the Department of Chemistry  
in Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology  
February 2017



© 2017 Massachusetts Institute of Technology  
All rights reserved

**Signature redacted**

Signature of Author.....

Department of Chemistry  
January, 17, 2017

**Signature redacted**

Certified by.. ..

Bradley L. Pentelute  
Pfizer-Laubach Career Development Professor of Chemistry  
Thesis Supervisor

**Signature redacted**

Accepted by.....

Robert W. Field  
Haslam and Dewey Professor of Chemistry  
Chairman, Departmental Committee for Graduate Students



This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Signature redacted

---

JoAnne Stubbe  
Novartis Professor of Chemistry  
Thesis Committee Chair

Signature redacted

---

Bradley L. Pentelute  
Pfizer-Laubach Career Development Professor of Chemistry  
Thesis Supervisor

Signature redacted

---

Elizabeth M. Nolan  
Associate Professor



# New Methods for Synthesis and Modification of Peptides and Proteins

by

Alexander Vinogradov

Submitted to the Department of Chemistry  
on January 30<sup>th</sup>, 2017 in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

## Abstract

Chemical modification of peptides and proteins is an enabling suite of tools for tailoring the properties of these biomolecules to specific applications. A number of bio-conjugation reactions allows fine-tuning of the biological activity, proteolytic stability, and immunogenicity of peptides and proteins, as well as equipping them with completely novel functions such as cell penetration, fluorescence, unique chemical reactivity, and much more.

Described herein are a number of new methods for the synthesis of modified peptides and proteins, and an approach to the discovery of such methodologies. Applications of fast-flow solid phase peptide synthesis — a technique recently developed to accelerate and improve peptide synthesis— towards the synthesis of difficult sequences and the refinement of associated protocols is described. The utility of the system is demonstrated via rapid total synthesis of barnase, a model 110-residue RNase, in the L- and D-forms. Systematic characterization of the biochemical properties of the synthesized proteins revealed that barnase is able to hydrolyze substrates of various chiralities, and that D-barnase is fully proteolytically stable.

Separately, a method for the preparation and utilization of unprotected peptide isocyanates in water was developed. It was shown that easily accessible C-terminal peptide isocyanates can be conjugated to a number of strong nucleophiles in the presence of unprotected amino acid side chains for peptides and proteins of various structures. Two-component macrocyclization of peptide isocyanates with bifunctional linkers was developed as an extension of the described chemistry. The resulting cyclic peptides were shown to be more proteolytically stable and more bioactive than their linear analogs.

In pursuit of generalizing the C-terminal protein modification chemistry to fully proteogenic peptides and proteins, a number of library screening approaches was developed. Liquid chromatography coupled to tandem mass spectrometry was employed to screen and reliably decode synthetic peptide libraries in a high-throughput manner. These protocols were used to discover proteogenic sequence tags reactive towards substituted hydrazine derivatives in a transpeptidation reaction. The discovered C-terminal tripeptide tag His-Gly-Cys underwent transpeptidation with a number of structurally different nucleophiles in various sequence contexts.

Thesis Supervisor: Bradley L. Pentelute  
Pfizer-Laubach Career Development Professor of Chemistry



## Acknowledgements

Graduate school has been an eye-opening experience. My time here transformed the way I perceive the world, think about problems in science or outside of it, and even interact with people. Of course, these changes did not happen overnight, nor did they occur spontaneously: a lot of people influenced and contributed to my professional and personal development.

First and foremost, I thank Prof. Bradley Pentelute, my thesis advisor, for the careful guidance he provided all this time. Walking the fine line, Brad gave me an incredible amount of intellectual freedom to explore ideas I was passionate about, while still providing necessary guidance and support to ensure the success of my undertakings. Brad's enthusiasm is contagious: every time I talked to him I felt reinvigorated and filled with determination. And that was a big deal.

I would like to extend my gratitude to my thesis committee chair, Prof. JoAnne Stubbe, Prof. Elizabeth Nolan, and Prof. Matthew Shoulders for helpful suggestions and insightful discussions throughout my time here. As far as I can tell, most of these recommendations greatly improved the overall quality of the work, and enabled a number of neat experiments. Additionally, I would like to thank Liz for providing continuous access to some of the essential lab equipment.

I spent countless hours talking about science, arguing and sharing ideas with my peers: Dr. Dmitry Vinichenko, Mark Simon, Dr. Zak Gates, and Chi Zhang. These are some of the brightest people I know of: they taught me a tremendous amount, helped through the low points of graduate school, and brightened up the overall experience. I also enjoyed working with and learning from other Pentelute lab members: Amy, Alex M., Alex L., Anthony Q., Ethan, Justin, Jingjing, Surin, Zi-Ning and everybody else. The Pentelute lab has been an exceptionally nurturing environment; hopefully it stays the same.

I am grateful to my parents, who have been very understanding of my passions. They supported my fairly spontaneous moving to another country, and encouraged me ever since.

Finally, if, by any chance, any of you reads it, I thank everyone from /djt/. Inspired by your approach and mutual encouragement, I persevered towards what seemed like an unattainable goal. This sense of direction definitely helped me survive graduate school. The goal is still distant as ever, but I am not giving up.





# Table of Contents

Abstract.....	5
Acknowledgements.....	7
Table of Contents.....	9
List of Figures.....	13
List of Tables.....	17
Preface.....	19
References.....	24
<b>Chapter 1. Rapid Total Synthesis of RNase <i>B. a.</i> ....</b>	<b>26</b>
1.1. Introduction.....	27
1.2. Results.....	29
1.2.1. Initial Peptide Synthesis.....	29
1.2.2. Protein Synthesis via Native Chemical Ligation.....	32
1.3. Discussion.....	35
1.3.1. Deletion Products.....	37
1.3.2. Aspartimide Formation.....	38
1.3.3. Miscellaneous Problems.....	40
1.3.4. Oxidation of Peptide Hydrazides and Native Chemical Ligation.....	40
1.3.5 Conclusion.....	42
1.4. Experimental.....	43
1.4.1. General.....	43
1.4.2. Peptide Synthesis Protocols.....	43
1.4.3. Functionalization of Resins.....	44
1.4.4. Analytical Methods.....	46
1.4.5. Synthesis of and Purification of Barnase Fragments.....	48
1.4.6. Native Chemical Ligation of Barnase Peptides.....	66
1.5. Acknowledgements.....	78
1.6. References.....	78
<b>Chapter 2. Total Synthesis and Biochemical Characterization of Mirror Image Barnase.....</b>	<b>81</b>
2.1. Introduction.....	82
2.2. Results and Discussion.....	82
2.2.1. Protein Synthesis.....	82
2.2.2. Enzymatic Characterization of D-Barnase.....	84
2.2.3. Proteolytic Stability of D-Barnase.....	89
2.3.4. Conclusions.....	93
2.3. Experimental.....	93
2.3.1. General.....	93

2.3.2. Methods.....	94
2.3.3. Protein Expression and Purification.....	95
2.3.4. Native RNA Digestion by D-Barnase .....	97
2.3.5. Proteolysis Experiments.....	97
2.4. Acknowledgements.....	102
2.5. References.....	102
<b>Chapter 3. C-terminal Modification of Fully Unprotected Peptide Hydrazides via <i>in situ</i> Generation of Isocyanates .....</b>	<b>105</b>
3.1. Introduction.....	106
3.2. Results and Discussion .....	106
3.2.1. Investigation of Conjugation Conditions .....	106
3.2.2. Investigation of Conjugation Scope .....	108
3.3. Experimental.....	118
3.3.1. General.....	118
3.3.2. Peptide Synthesis .....	118
3.3.3. HPLC-MS Analysis .....	120
3.3.4. Discovery of Isocyanate Generation .....	120
3.3.5. Optimization of Conjugation Conditions.....	127
3.3.6. Side Reactions Associated with Some C-terminal Amino Acids .....	130
3.3.7. Sortase A Mediated Ligation of Peptide Hydrazide to LF <sub>N</sub> -DTA .....	133
3.3.8. Peptide Synthesis: Analytical Data .....	135
3.3.9. Nucleophile Scope of Conjugation: Analytical Data .....	138
3.5. Acknowledgements.....	142
3.6. References.....	142
<b>Chapter 4. Macrocyclization of Unprotected Peptide Isocyanates.....</b>	<b>145</b>
4.1. Introduction.....	146
4.2. Results and Discussion .....	148
4.2.1. Establishment of the Macrocyclization Conditions .....	148
4.2.2. Properties of Macrocyclic Peptides.....	154
4.3. Experimental .....	158
4.3.1. General.....	158
4.3.2. Fmoc Solid Phase Peptide Synthesis .....	158
4.3.3. Peptide Macrocyclization.....	159
4.3.4. Protein Expression .....	160
4.3.5. Analytical Techniques.....	162
4.3.6. Peptide Synthesis: Analytical Data .....	164
4.3.7. Peptide Macrocyclization: Analytical Data .....	165
4.3.8. Bio-Layer Interferometry: Analytical Data.....	169
4.3.9. NMR Spectra .....	169

4.4. Acknowledgements.....	177
4.5. References.....	177
<b>Chapter 5. Strategies for High-Throughput Sequencing of Synthetic Peptide Libraries .....</b>	<b>179</b>
5.1. Introduction.....	180
5.2. Results and Discussion .....	184
5.2.1. Library Design .....	184
5.2.2. Decoding of a Model Library.....	187
5.2.3. Critical Evaluation of the Alternating Monomer Set Design .....	193
5.2.4. Analysis of Method Throughput .....	195
5.2.5. Application to Libraries Containing Non-proteogenic Amino Acid Residues .....	197
5.2.6. Conclusion .....	199
5.3. Experimental .....	200
5.3.1. General.....	200
5.3.2. Peptide and Library Synthesis.....	200
5.3.3. Library Cleavage and Nano-LC/MSMS Analysis .....	201
5.3.4. Amino Acid Frequency Distribution $\chi^2$ -Test for the Sample of 660 Library 1 Beads.....	203
5.3.5. Analysis of a Library 1 Sample Containing 306 Beads .....	204
5.3.6. Matching Fragmentation Spectra from Library and Authentic Peptides .....	205
5.4. Acknowledgements.....	213
5.5. References.....	213
<b>Chapter 6. Discovery of Reactive Sequence Tags for C-Terminal Peptide and Protein Conjugation via Reactivity Screening of Intact Peptide Libraries .....</b>	<b>215</b>
6.1. Introduction.....	216
6.2. Results and Discussion .....	219
6.2.1. Establishing Conditions for Library Reactivity Screening .....	219
6.2.2. Library 1 Screen.....	224
6.2.3. Library 2 Screen.....	228
6.2.4. Investigation of the Reaction Scope.....	231
6.2.5. Conclusion .....	235
6.3. Experimental .....	237
6.3.1. General.....	237
6.3.2. Library and Peptide Synthesis.....	237
6.3.3. Preparation of LF <sub>N</sub> -HKTHGC-COOH.....	238
6.3.4. Library Analysis.....	239
6.3.5. HPLC-MS Analysis .....	240
6.3.6. Conjugation to Different Nucleophiles .....	241
6.3.7. Peptide Synthesis: Analytical Data.....	242
6.3.8. Library 1 PFPh Dataset.....	251
6.3.9. Library 1 Mesna Dataset.....	252

6.3.10. Library 2 PFPh Dataset.....	256
6.4. Acknowledgements.....	261
6.5. References.....	261

## List of Figures

Figure 1.1. Synthetic scheme of barnase.....	28
Figure 1.2. HPLC-MS data (total ion current vs. time) of crude barnase fragments.....	30
Figure 1.3. HPLC-MS data (total ion current vs time) of purified barnase ligation products.....	33
Figure 1.4. Characterization of the synthesized protein.....	36
Figure 1.5. Base-promoted formation of aspartimide and its hydrolysis during Fmoc SPPS.....	39
Figure 1.6. Transformations of C-terminal peptide hydrazides during oxidation with sodium nitrite.	41
Figure 1.7. The kinetic data for the RNA hydrolysis assay.....	47
Figure 1.8. LC-MS total ion current vs. time for the synthesis of B[1+2] using a standard 3-minute cycle (HBTU, 50% (v/v) piperidine).....	49
Figure 1.9. LC-MS total ion current vs. time for the synthesis of B[1+2] using a standard 3-minute cycle (HATU, 50% (v/v) piperidine).....	49
Figure 1.10. LC-MS total ion current vs. time for the synthesis of B[1+2] using 10 equivalents of amino acids (HATU, 50% (v/v) piperidine).....	51
Figure 1.11. LC-MS total ion current vs. time for the synthesis of B[1+2] using a standard 3-minute cycle (HATU, 50% (v/v) piperidine) and less DIEA.....	51
Figure 1.12. LC-MS total ion current vs. time for the synthesis of B[1+2] on PEG resin using a standard 3-minute cycle (HATU, 50% (v/v) piperidine).....	52
Figure 1.13. LC-MS total ion current vs. time for the first twenty residues of B[1+2].....	52
Figure 1.14. LC-MS total ion current vs. time for the synthesis of B[1+2] (HATU, 50% (v/v) piperidine) using exhaustive deprotection on the 11 <sup>th</sup> -13 <sup>th</sup> residues coupled.....	53
Figure 1.15. LC-MS total ion current vs. time of purified G <sub>5</sub> -B[1].....	53
Figure 1.16. LC-MS total ion current vs. time the synthesis G <sub>3</sub> -B[1] on a standard 3-minute cycle (HATU, 20% (v/v) piperidine).....	55
Figure 1.17. LC-MS total ion current vs. time of purified G <sub>3</sub> -B[1].....	55
Figure 1.18. LC-MS total ion current vs. time for the original synthesis of B[2] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine).....	56
Figure 1.19. LC-MS total ion current vs. time for the resynthesis of B[2] (HATU, 20% (v/v) piperidine) using extended deprotection and couplings.....	56
Figure 1.20. LC-MS analysis of B[2].....	58
Figure 1.21. LC-MS total ion current vs. time for the synthesis B[3] using a standard 3-minute cycle (HATU, 50% (v/v) piperidine).....	59
Figure 1.22. LC-MS total ion current vs. time for the synthesis B[3] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine).....	59
Figure 1.23. LC-MS total ion current vs. time of purified B[3].....	60
Figure 1.24. LC-MS total ion current vs. time for the synthesis of B[4] on AM resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine with 0.1M HOBt).....	60
Figure 1.25. LC-MS total ion current vs. time for the synthesis of B[4] on PEG resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine with 0.1M HOBt).....	63
Figure 1.26. B[4] after incubation in 50% A / 50% B for 5 hours.....	63

Figure 1.27. LC-MS total ion current vs. time for the synthesis B[4] using PEG resin, a standard 3-minute cycle (HATU, 20% (v/v) piperidine), and DCC/HOBt activation of <sup>90</sup> Arg.....	64
Figure 1.28. LC-MS total ion current vs. time for the synthesis of B[4] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine and slightly modified coupling solutions .....	64
Figure 1.29. LC-MS total ion current vs. time for purified B[4].....	65
Figure 1.30. LC-MS total ion current vs. time for G <sub>5</sub> -B[1] oxidation reaction .....	67
Figure 1.31. LC-MS total ion current vs. time for G <sub>5</sub> -B[1] thioesterification reaction .....	67
Figure 1.32. LC-MS total ion current vs. time for improved G <sub>5</sub> -B[1] MPAA formation reaction.....	68
Figure 1.33. LC-MS total ion current vs. time for the ligation of B[1] to B[2] after 19 hours .....	68
Figure 1.34. LC-MS total ion current vs. time for purified fragment B[5] .....	70
Figure 1.35. LC-MS total ion current vs. time for the ligation of B[3] to B[4] after 10 hours .....	70
Figure 1.36. LC-MS total ion current vs. time of purified (Acm,Cy)-B[6].....	71
Figure 1.37. LC-MS total ion current vs. time for Acm removal of (Acm,Cy)-B[6].....	71
Figure 1.38. LC-MS total ion current vs. time of crude B[6] with residual <i>p</i> -cresol .....	72
Figure 1.39. LC-MS total ion current vs. time of purified B[6] .....	72
Figure 1.40. LC-MS total ion current vs. time for the oxidation and thioesterification of B[5] .....	74
Figure 1.41. LC-MS total ion current vs. time of purified MPAA thioester of B[5].....	74
Figure 1.42. LC-MS total ion current vs. time for the ligation of B[5]-MPAA thioester to B[6] after 36 hours.....	75
Figure 1.43. LC-MS total ion current vs. time of purified B[7] .....	75
Figure 1.44. LC-MS total ion current vs. time for isolated, desulfurized full-length barnase.....	77
Figure 2.1. Protein synthesis .....	83
Figure 2.2. Enzymatic characterization of D-barnase.....	85
Figure 2.3. Digestion of native RNA by L- and D-barnase.....	88
Figure 2.4. Comparison of the proteolytic stability of L- and D-barnase.....	90
Figure 2.5. HPLC-MS characterization of recombinant barnase .....	96
Figure 2.6. HPLC-MS characterization of recombinant barstar.....	96
Figure 2.7. Relative proteolytic stabilities of L- and D-barnase .....	99
Figure 2.8. Relative proteolytic stabilities of L- and D-barnase .....	100
Figure 2.9. Proteolytic stability of D-barnase towards denaturing proteinase K digestion.....	101
Figure 2.10. Proteolytic stability of D-barnase towards extended proteinase K digestion .....	101
Figure 3.1. <i>In situ</i> generation of peptide isocyanates .....	107
Figure 3.2. Influence of the parameters on the outcome of the reaction .....	109
Figure 3.3. HPLC-MS analysis of two diastereomeric conjugation products .....	112
Figure 3.4. Perfluorophenylhydrazine conjugation to longer peptides.....	116
Figure 3.5. C-terminal labeling of 53 kDa protein with perfluorophenylhydrazine.....	117
Figure 3.6. HPLC-MS analysis of the isocyanate formation.....	122
Figure 3.7. HPLC-MS analysis of a crude conjugation reaction with MPAA .....	123
Figure 3.8. Transformations of the C-terminal isocyanate in water .....	123

Figure 3.9. HPLC-MS analysis of a crude conjugation reaction with perfluorophenylhydrazine .....	125
Figure 3.10. MS/MS analysis of the synthesized semicarbazide.....	126
Figure 3.11. Optimization of reaction conditions.....	128
Figure 3.12. The effect of solvent on the outcome of the conjugation .....	129
Figure 3.13. Formation of oxazolidin-2-ones for Ser/Thr-containing peptides.....	131
Figure 3.14. Oxidation of peptides bearing C-terminal Glu or His hydrazides.....	132
Figure 3.15. HPLC-MS of C-terminal hydrazide of LF <sub>N</sub> -DTA.....	134
Figure 4.1. The concept of the study .....	147
Figure 4.2. On resin synthesis of glutamic acid $\gamma$ -hydrazides .....	149
Figure 4.3. Development of isocyanate macrocyclization reaction.....	150
Figure 4.4. Attempts to macrocyclize acyl azides .....	152
Figure 4.5. Substrate and linker scope of the isocyanate macrocyclization .....	153
Figure 4.6. Proteolytic stability of macrocyclic peptides .....	155
Figure 4.7. The effects of the macrocyclization on the biological properties of NYAD-1 and its macrocyclized analogues .....	157
Figure 4.8. HPLC-MS analysis of expressed C-CA.....	161
Figure 4.9. HPLC-MS analysis of macrocyclized peptide 4a .....	165
Figure 4.10. HPLC-MS analysis of macrocyclized peptide 4b .....	165
Figure 4.11. HPLC-MS analysis of macrocyclized peptide 4c .....	166
Figure 4.12. HPLC-MS analysis of macrocyclized peptide 4d .....	166
Figure 4.13. HPLC-MS analysis of macrocyclized peptide 1a .....	167
Figure 4.14. HPLC-MS analysis of macrocyclized peptide 1b .....	167
Figure 4.15. HPLC-MS analysis of macrocyclized peptide 1c .....	168
Figure 4.16. HPLC-MS analysis of macrocyclized peptide 1d .....	168
Figure 5.1. Comparison of various decoding strategies .....	181
Figure 5.2. The overall experimental workflow for the proposed sequencing strategy .....	183
Figure 5.3. Recovery of sequences with incomplete fragmentation ladders using the AMS principle .....	186
Figure 5.4. Analysis of nLC efficiency .....	189
Figure 5.5. Evaluation of the role of AMS in increasing the confidence in sequencing results.....	192
Figure 5.6. Evaluating the reliability of library decoding .....	194
Figure 5.7. Analysis of the method throughput.....	196
Figure 5.8. Decoding of the peptide libraries containing non-canonical amino acids.....	198
Figure 5.9. Evaluation of the role of AMS in increasing the confidence in sequencing results.....	204
Figure 5.10. HPLC-MS analysis of crude H <sub>2</sub> N- GC $\beta$ FADASEFPHG-COOH .....	206
Figure 5.11. HPLC-MS analysis of crude H <sub>2</sub> N- GC $\beta$ FLDEGYGPWG-COOH .....	206
Figure 5.12. HPLC-MS analysis of crude H <sub>2</sub> N- GC $\beta$ FLDEVEFPHG-COOH .....	206
Figure 5.13. HPLC-MS analysis of crude H <sub>2</sub> N- GC $\beta$ WADESAGPWG-COOH .....	207
Figure 5.14. HPLC-MS analysis of crude H <sub>2</sub> N- GC $\beta$ WLDEDTFMGG-COOH .....	207

Figure 5.15. HPLC-MS analysis of crude H <sub>2</sub> N- GCβWLDPSLFPHG-COOH.....	207
Figure 5.16. HPLC-MS analysis of crude H <sub>2</sub> N- GCβWLDTDPFPHG-COOH.....	208
Figure 5.17. HPLC-MS analysis of crude H <sub>2</sub> N- GCβWLDTSEFPHG -COOH .....	208
Figure 5.18. HPLC-MS analysis of crude H <sub>2</sub> N- GCβWLSADEFPHG-COOH .....	208
Figure 5.19. Overlaid raw CID fragmentation spectra for GCβWLSADEFPHG .....	209
Figure 5.20. Overlaid raw CID fragmentation spectra for GCβFLDEGYGPWG .....	209
Figure 5.21. Overlaid raw CID fragmentation spectra for GCβWLDEDTFMGG.....	210
Figure 5.22. Overlaid raw CID fragmentation spectra for GCβWLDTDPFPHG .....	210
Figure 5.23. Overlaid raw CID fragmentation spectra for GCβWLDTSEFPHG.....	211
Figure 5.24. Overlaid raw CID fragmentation spectra for GCβWLDPSLFPHG .....	211
Figure 5.25. Overlaid raw CID fragmentation spectra for GCβWADESAGPWG .....	212
Figure 6.1. The general idea of the study .....	217
Figure 6.2. Site-selective hydrazinolysis of peptides and proteins.....	218
Figure 6.3. The design of library 1 .....	220
Figure 6.4. Analysis of the naïve library 1 dataset .....	222
Figure 6.5. Sequencing quality heat maps for the datasets containing erroneous assignments obtained from a naïve library analysis.....	223
Figure 6.6. Library 1 reaction screening .....	225
Figure 6.7. Analysis of the naïve library 2 dataset .....	229
Figure 6.8. Design and reactivity screening of library 2 .....	230
Figure 6.9. Results of peptide 14 alanine scan experiments.....	232
Figure 6.10. Analysis of the reactivity exhibited by the discovered sequence tag .....	233
Figure 6.11. Synthesis of a C-terminal protein hydrazide using the discovered chemistry .....	234



## List of Tables

Table 1.1. Yields of peptides and overall yield of synthesized barnase .....	34
Table 2.1. Catalytic activities values ( $k_{cat}/K_M$ in $M^{-1}\cdot s^{-1}$ ) of select RNases towards different fluorogenic substrates.....	87
Table 2.2. Remaining RNase catalytic activity of L- and D-barnase after the proteolytic digestion with select proteases.....	92
Table 3.1. Nucleophile scope of the reaction .....	111
Table 3.2. C-terminal amino acid scope of the reaction .....	114
Table 5.1. Monoisotopic masses in Da of all dipeptides encodable with the alternating monomer set library design .....	185
Table 5.2. Data analysis summary for a sample consisting of 660 beads bearing library 1 peptides.	191
Table 5.3. $\chi^2$ -test statistic table for the amino acid distribution in the sample of 660 library 1 beads	203
Table 6.1. Summary of the raw PEAKS output for the sequencing analysis of naïve library 1 .....	222
Table 6.2. Summary of the final (parsed) dataset for the sequencing analysis of naïve library 1 .....	222
Table 6.3. Summary of the FDR experiment .....	223
Table 6.4. Reactivity of authentic hit peptides towards nicotinic hydrazide.....	227
Table 6.5. Summary of the raw PEAKS output for the sequencing analysis of naïve library 2.....	229
Table 6.6. Summary of the final (parsed) dataset for the sequencing analysis of naïve library 2 .....	229



## Preface

“Life is the mode of existence of protein bodies”.<sup>1</sup> Frederick Engels put forth this idea almost 150 years ago, and despite numerous scientific discoveries that completely changed our understanding of Life, his definition still stands firm. Indeed, proteins are implicated in all facets of Life: they build our skin and muscles, transport oxygen throughout the body, metabolize nutrients, regulate gene expression levels, and replicate our cells. As these functions became clear, chemical technology was there to take advantage of these fascinating properties of proteins outside of living systems. These days, we use proteins to make some of the strongest materials on Earth,<sup>2</sup> use them as exquisitely selective catalysts,<sup>3</sup> medications,<sup>4,5</sup> and so much more.

However, the fact that proteins are the basis of living systems also introduces limitations to their utility in science in technology. Specifically, proteins can interact with living systems, which may lead, for instance, to their premature proteolytic degradation or an immune response from the host organism. Synthesis of modified peptides and proteins exists to alleviate these limitations. A myriad of approaches to the creation of “unnatural” biomolecules with improved functions has been developed. D-Peptides,<sup>6</sup> peptoids,<sup>7-9</sup>  $\beta$ -peptides,<sup>10</sup> macrocyclic peptides<sup>11</sup> represent but a fraction of peptidomimetic biopolymers developed to build on the foundation laid down by the traditional protein science. Peptide and protein modification also helps us in empowering these molecules with completely new functions. These functions are numerous; some examples include selectively cytotoxic antibody-drug conjugates,<sup>12</sup> fluorophore-labeled proteins for imaging applications,<sup>13</sup> and cell-penetration achieved via the conjugation of polycationic moieties.<sup>14,15</sup>

Generally speaking, there are two ways to prepare “unnatural” peptides and proteins: either through *de novo* chemical synthesis or via point modification of existing biomolecules obtained, primarily, via biological expression or isolation from host organism.

Of these two ways, chemical synthesis undoubtedly is a more enabling technique: any number of chemical modifications in any position can be installed if a peptide is *de novo* chemically assembled. As powerful as chemical synthesis may be, it is also a lot more challenging — both in terms of required time and the overall synthesis cost — than biological expression methods, at least for the production of reasonably sized proteins. The state of the art protocols for protein synthesis<sup>16</sup> make use of native chemical ligation reactions<sup>17</sup> supplemented by a number of auxiliary reactions, such as desulfurization,<sup>18,19</sup> orthogonal protecting group removal, and thioester forming reactions,<sup>20</sup> to stitch together peptides comprising the sequence of a protein of interest. In turn, solid phase peptide synthesis (SPPS) is a method of choice for the production of peptide fragments. This is the technique that allows almost unlimited incorporation of modified amino acids and other structural elements into peptide structure. SPPS of  $\alpha$ - (L- and D-amino acids,

proteogenic and non-proteogenic) and  $\beta$ -peptides, peptoids, ureas, and many others kinds of structures is now commonplace.

Two general strategies for SPPS, different in their approach to the side chain and  $N^\alpha$ -protecting schemes are known. In Boc strategy,<sup>21</sup> TFA-labile  $N^\alpha$ -Boc-protected amino acids are used, and the side chains protection scheme is HF-labile. In Fmoc strategy,<sup>22</sup> base removable  $N^\alpha$  Fmoc-amino acids bear TFA-labile side chain protecting groups. Although this opinion may be controversial, the Boc strategy is superior in terms of the resulting peptide quality and the speed of synthesis, whereas Fmoc synthesis is cheaper and safer. Recently, in an effort to overcome the challenges associated with Fmoc peptide synthesis (primarily, the speed of synthesis), the Pentelute lab developed an Fmoc-based fast-flow peptide synthesis platform.<sup>23</sup> The platform is flow-based; it relies on a continuous flow of reagents through a resin bed confined inside a reaction vessel. Rapid delivery and exchange of reagents allows the maintenance of high concentration of fresh reagents in the resin bed at all points in time, which was hypothesized to accelerate and improve synthesis outcomes. Additionally, the reactor is kept heated at 60-90 °C at all times further speeding the process, and potentially giving extra benefit by alleviating the effects of on-resin peptide aggregation,<sup>24</sup> which increases the efficiency of the chain elongation process.

In *Chapter 1*, we discuss the refinement of the fast-flow peptide synthesis platform. Working towards total chemical synthesis of *B. a.* RNase,<sup>25</sup> a model bacterial ribonuclease, we worked through a number of challenges traditionally encountered in Fmoc SPPS. Namely, we had to overcome persistent formation of deletion products linked to on-resin peptide aggregation,<sup>26</sup> and base-promoted aspartimide formation for sequences containing Asp-Gly, Asp-Ala, and Asp-Asn dipeptide motifs.<sup>27</sup> We systematically evaluated synthetic parameters and previously published solutions in terms of their applicability to the flow protocols, and, based on the results of these investigations, we were able to efficiently synthesize and purify a number of “difficult” sequences. These new synthetic protocols enabled rapid and efficient convergent synthesis of barnase from four peptide fragments with three native chemical ligation steps.

As an application of the developed techniques, in *Chapter 2* we prepared and characterized a D-enantiomer of barnase. D-Proteins, which are mirror images of their native L-counterparts, are an intriguing class of “unnatural” biomolecules. They possess a number of distinguishing features, as D-proteins were shown to be proteolytically stable and non-immunogenic,<sup>28,29</sup> yet still biologically active<sup>6,30</sup> — although the biological activity is manifested towards an enantiomer target (the so-called principle of reciprocal chiral specificity). However, D-proteins are only accessible via total chemical synthesis. We sought to take advantage of the fast-flow peptide synthesis protocols to deepen our knowledge of mirror image proteins. To this end, we further refined synthetic protocols for D-barnase (specifically, we improved several ligation steps), and used the obtained protein to characterize it enzymatically as well as to study its proteolytic stability. Interestingly, we found that barnase is not fully stereospecific towards its substrates, and therefore

we observed some activity of mirror image barnase towards native RNA. In the proteolysis study, we found that a variety of digestive proteases, including those that are specific towards achiral glycine residues, were unable to digest mirror image barnase, confirming its full stability towards proteolysis.

As mentioned above, point modification of biologically produced peptides and proteins is another way of tailoring the properties of these biomolecules. This field is much too large to cover its current state of the art methodologies here; additionally, a number of excellent reviews exist.<sup>31-33</sup> Briefly, protein modification can be either site-specific (chemo- and regioselective) or not. Considering that all proteins are comprised of just twenty distinct amino acids, and considering the size of an average protein (200 or so amino acid residues), each chemical functionality associated with amino acid side chains are found multiple times on each protein, which makes the task of site-selective protein modification difficult.

Non-specific protein labeling is, of course, easier to achieve, and for many applications (protein PEGylation<sup>34</sup> and antibody drug conjugate<sup>35</sup> synthesis are just two examples) non-specifically labeled protein mixtures do the job. At the same time, there is no denying that site-selective protein modification is desirable in most fields of research and technology, as the resulting constructs are homogeneous in their chemical, physical and biological properties, which in turn simplifies protein purification, characterization and improves the overall reproducibility of the protocols.

How does one achieve site-selective protein modification? A number of approaches exist. Perhaps the most straightforward one is to take advantage of amino acids found on protein surfaces at extremely low frequency, such as cysteine,<sup>36</sup> although this method is not generalizable. Another approach is to introduce amino acids containing chemical moieties of unique reactivity during or after the expression step.<sup>37-40</sup> Examples include expression of proteins containing azides, aldehydes, aminophenols, alkynes, and many other functional groups. These functionalities may later be introduced into bio-orthogonal reactions to prepare site-selectively modified constructs. However, expression of proteins containing non-proteogenic amino acids is not necessarily straightforward, which somewhat limits the use of the approach.

Another strategy relies on making use of unique reactivities of proteogenic amino acids found in local sequence environments. This approach is exemplified by the use of a cysteine residue inside a tetrapeptide tag to achieve site-selective perfluoroarylation, a method recently developed in the Pentelute lab,<sup>41</sup> as well as few other similar techniques.<sup>42,43</sup> If it is possible to discriminate amino acid residues based on their chemical environment, should it not be possible to take advantage of protein termini, which also display unique functionalities present in every peptide and protein (with the few exceptions of unusual cyclic systems)? Indeed, a number of N-terminal protein modification methods are also known. These include native chemical ligation,<sup>17,44</sup> Srt A-mediated ligation,<sup>45</sup> Ser/Thr oxidation to generate aldehydes<sup>46</sup> and many other techniques.<sup>33,47,48</sup> C-terminal modification techniques are less numerous. These methods mostly rely

on the synthesis of thioesters and acyl hydrazides to perform native chemical ligation with an N-terminal cysteine containing ligation partner.

In *Chapter 3*, we discuss a reaction serendipitously discovered during barnase synthesis, which led to the formation of peptide isocyanates from precursor C-terminal hydrazides via a sequence of hydrazide oxidation followed by a Curtius rearrangement. The reaction proceeded in water for fully unprotected peptide isocyanates, which enabled direct conjugation of strong nucleophiles to peptides and proteins. This transformation represents a new method for the C-terminal modification of peptides and proteins, with a number of characteristic features. First, in this reaction, peptides act as electrophiles during the reaction, while the conjugation partners are nucleophilic, which is not very common,<sup>49,50</sup> primarily due to the fact that proteins inherently are nucleophilic in nature. Second, to us, the reaction seemed interesting due the incredibly dynamic nature of the product distribution. In fact, yields and the very nature of the reaction products greatly depend on the pH and the reagent concentrations. The extraordinary reactivity of isocyanates highlighted the unique reactivity of various hydrazine derivatives, which reacted with each other at low concentrations and low pH (3-4), where every other functional group in the substrate proved to be unreactive. Almost every amino acid side chain is capable of reacting with isocyanates, yet under appropriate reaction conditions none react.

In *Chapter 4*, this chemistry was further elaborated and extended to the synthesis of side-chain peptide isocyanates, generated from  $\gamma$ -glutamic acid hydrazides. Peptides bearing two of these functionalities were cyclized with bifunctional dihydrazides of dicarboxylic acids to offer a two component modular approach to the generation of macrocyclic peptides bridged by a rigid hydrophilic linker. As mentioned above, macrocyclization is one of the most popular peptide modification tools,<sup>11,51-55</sup> which restricts the conformational freedom of the biomolecule, and traps it in a particular conformation. When macrocyclized in an  $i,i+4$  fashion, peptides often display a higher degree of  $\alpha$ -helicity, which improves a number of biologically relevant properties of the resulting constructs. Thus, macrocyclized peptides were demonstrated to be more resistant to proteolysis and be more bioactive than their linear counterparts. Our experiments revealed that stapled peptides prepared via the isocyanate/hydrazide cyclization demonstrate similar characteristics.

Nucleophile conjugation to C-terminal isocyanates is naturally limited by the availability of corresponding protein hydrazides.<sup>56</sup> As mentioned above, a number of methods to prepare these exist, but each of them is also associated with practical limitations. Therefore, we sought to expand the reaction to the case of fully proteogenic proteins, in an attempt to increase the utility of uncovered transformations. We hypothesized that combining the sequence tag approach with what we learned from the isocyanate conjugation (namely, the unique reactivity of various hydrazine derivatives) may be productive. We thought that it may possible to discover a site-selective C-terminal sequence tag that would engage in a direct or

indirect reaction with an external nucleophile, leading to a cleavage of peptide backbone with the nucleophile. In other words, what we looked for was a non-enzymatic transpeptidation reaction to achieve C-terminal peptide and protein modification.

In *Chapter 5*, we discuss the development of high-throughput strategies for the decoding of synthetic peptide libraries, an approach that we took in *Chapter 6* to discover the transpeptidation tag. We found that the combination of liquid chromatography and tandem mass spectrometry can be utilized to analyze mixtures containing thousands of peptides at once. After establishing necessary experimental protocols and data analysis routines, we benchmarked and verified the performance of our approach and used it to directly profile the outcomes of the reaction between tri- and tetrapeptide libraries with various hydrazine-derived nucleophiles. This workflow yielded multiple reactive sequences, which were validated in a traditional batch format and studied in greater detail. The results of our analysis point towards His-Gly-Cys as a reactive C-terminal tripeptide, which is able to engage in a reaction with hydrazides in the presence of external thiol in various sequence contexts.

---

As the brief summary provided above illustrates, the work presented in this thesis primarily revolves around the development of new tools for the synthesis of unnatural peptides and proteins. The thesis explores different approaches towards the goal of creating modified biomolecules with designer properties.

The field of protein modification is still in its infancy. It is my personal belief that one day the chemistry to manipulate proteins in all imaginable ways (for instance: ligation, cleavage, insertion, and mutagenesis of protein sequences in ways similar to those of the DNA technology) will exist. It is my hope that this thesis brings these times a tiny bit closer to the reality.

## References

- (1) Engels, F. *Dialectics of Nature*; Progress Publishers, Moscow, 1934.
- (2) Altman, G.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. *Biomaterials* **2003**, *24*, 401–416.
- (3) Lobedanz, S.; Damhus, T.; Borchert, T.; Hansen, T.; Lund, H.; Lai, W.; Lin, M.; Leclerc, M.; Kirk, O. *Enzymes in Industrial Biotechnology*; John Wiley & Sons, Inc., 2016.
- (4) Vellard, M. *Curr. Opin. Biotechnol.* **2003**, *14*, 444–450.
- (5) Craik, D.; Fairlie, D.; Liras, S.; Price, D. *Chem. Biol. Drug Des.* **2013**, *81*, 136–147.
- (6) Milton, R.; Milton, S.; Kent, S. *Science.* **1990**, *256*, 1445–1448.
- (7) Udugamasooriya, D.; Ritchie, C.; Brekken, R.; Kodadek, T. *Bioorganic Med. Chem.* **2009**, *16*, 6338–6343.
- (8) Zuckermann, R.; Kodadek, T. *Curr. Opin. Mol. Ther.* **2009**, *11*, 299–307.
- (9) Fowler, S.; Blackwell, H. *Org. Biomol. Chem.* **2009**, *7*, 1508–1524.
- (10) Cheng, R.; Gellman, S.; DeGrado, W. *Chem. Rev.* **2001**, *101*, 3219–3232.
- (11) Verdine, G.; Hilinski, G. *Stapled Peptides for Intracellular Drug Targets*; 1<sup>st</sup> ed.; Elsevier Inc., **2012**; Vol. 503.
- (12) Sievers, E.; Senter, P. *Annu. Rev. Med.* **2013**, *64*, 15–29.
- (13) Umezawa, K.; Citterio, D.; Suzuki, K. *Anal. Sci.* **2014**, *30*, 327–349.
- (14) McNaughton, B.; Cronican, J.; Thompson, D.; Liu, D. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6111–6116.
- (15) Li, M.; Tao, Y.; Shu, Y.; Larochelle, J.; Steinauer, A.; Thompson, D.; Schepartz, A.; Chen, Z.-Y.; Liu, D. *J. Am. Chem. Soc.* **2015**, *137*, 14084–14093.
- (16) Dawson, P.; Kent, S. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
- (17) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science.* **1994**, *266*, 776–778.
- (18) Wan, Q.; Danishefsky, S. *Angew. Chem. Int.; Ed.* **2007**, *46*, 9248–9252.
- (19) Pentelute, B.; Kent, S. *Org. Lett.* **2007**, *9*, 687–690.
- (20) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7645–7649.
- (21) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. *Int. J. Pept. Protein Res.* **1992**, *40*, 180–193.
- (22) Fields, G.; Noble, R. *Int. J. Pept. Protein Res.* **1990**, *35*, 161–214.
- (23) Simon, M.; Heider, P.; Adamo, A.; Vinogradov, A.; Mong, S.; Li, X.; Berger, T.; Policarpo, R.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A.; Jensen, K.; Pentelute, B. *Chembiochem* **2014**, *15*, 713–720.
- (24) Hyde, C.; Johnson, T.; Owen, D.; Quibell, M.; Sheppard, R. *Int. J. Pept. Protein Res.* **1994**, *43*, 431–440.
- (25) Hartley, R. W. *Trends Biochem. Sci.* **1989**, *14*, 450–454.
- (26) Coin, I.; Beyermann, M.; Bienert, M. *Nat. Protoc.* **2007**, *2*, 3247–3256.
- (27) Mergler, M.; Dick, F.; Sax, B.; Weiler, P.; Vorherr, T. *J. Pept. Sci.* **2003**, *46*, 36–46.
- (28) Dintzis, H.; Symer, D.; Dintzis, R.; Zawadzke, L.; Berg, J. *Proteins Struct. Funct. Genet.* **1993**, *16*, 306–308.
- (29) Zawadzke, L.; Berg, J. *J. Am. Chem. Soc.* **1992**, *114*, 4002–4003.
- (30) Fitzgerald, M.; Chernushevich, I.; Standing, K.; Kent, S.; Whitman, C. *J. Am. Chem. Soc.* **1995**, *117*, 11075–11080.
- (31) Kalia, J.; Raines, R. *Curr. Org. Chem.* **2010**, *14*, 138–147.
- (32) Hermanson, G. *Bioconjugate Techniques*; Academic Press: London, 2013.
- (33) Sletten, E.; Bertozzi, C. *Angew. Chem.; Int. Ed.* **2009**, *48*, 6974–6998.
- (34) Roberts, M.; Bentley, M.; Harris, J. *Adv. Drug Deliv. Rev.* **2012**, *64*, 116–127.
- (35) Alley, S.; Okeley, N.; Senter, P. *Curr. Opin. Chem. Biol.* **2010**, *14*, 529–537.
- (36) Chalker, J.; Bernardes, G.; Lin, Y.; Davis, B. *Chem. - An Asian J.* **2009**, *4*, 630–640.



- (37) Jackson, D.; King, D.; Chmielewski, J.; Singh, S.; Schultz, P. *J. Am. Chem. Soc.* **1991**, *113*, 9391–9392.
- (38) Xie, J.; Schultz, P. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548–554.
- (39) Liu, C.; Schultz, P. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
- (40) Goto, Y.; Kato, T.; Suga, H. *Nat. Protoc.* **2011**, *6*, 779–790.
- (41) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N.; Santos, M.; Van Voorhis, T.; Pentelute, B. *Nat. Chem.* **2015**, *8*, 120–128.
- (42) Ramil, C.; An, P.; Yu, Z.; Lin, Q. *J. Am. Chem. Soc.* **2016**, *138*, 5499–5502.
- (43) Martin, B.; Giepmans, B.; Adams, S.; Tsien, R. *Nat. Biotechnol.* **2005**, *23*, 1308–1314.
- (44) Bang, D.; Pentelute, B.; Kent, S. *Angew. Chem.; Int. Ed.* **2006**, *45*, 3985–3988.
- (45) Navarre, W.; Schneewind, O. *Mol. Microbiol.* **1994**, *14*, 115–121.
- (46) Davies, M. *Biochem. J.* **2016**, *473*, 805–825.
- (47) Obermeyer, A.; Jarman, J.; Netirojjanakul, C.; El Muslemany, K.; Francis, M. *Angew. Chem.; Int. Ed.* **2013**, *53*, 1057–1061.
- (48) Behrens, C.; Hooker, J.; Obermeyer, A.; Romanini, D.; Katz, E.; Francis, M. *J. Am. Chem. Soc.* **2011**, *133*, 16398–16401.
- (49) Besret, S.; Vicogne, J.; Dahmani, F.; Fafeur, V.; Desmet, R.; Drobecq, H.; Romieu, A.; Melnyk, P.; Melnyk, O. *Bioconjug. Chem.* **2014**, *25*, 1000–1010.
- (50) Carrico, I.; Carlson, B.; Bertozzi, C. *Nat. Chem. Biol.* **2007**, *3*, 321–322.
- (51) Leshchiner, E.; Parkhitko, A.; Bird, G.; Luccarelli, J.; Bellairs, J.; Escudero, S.; Opoku-Nsiah, K.; Godes, M.; Perrimon, N.; Walensky, L. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 1761–1766.
- (52) Bird, G.; Bernal, F.; Pitter, K.; Walensky, L. *Methods Enzymol.* **2008**, *446*, 369–386.
- (53) Stewart, M.; Fire, E.; Keating, A.; Walensky, L. *Nat. Chem. Biol.* **2010**, *6*, 595–601.
- (54) Schafmeister, C.; Po, J.; Verdine, G. *J. Am. Chem. Soc.* **2000**, *122*, 5891–5892.
- (55) Spokoiny, A.; Zou, Y.; Ling, J.; Yu, H.; Lin, Y.; Pentelute, B. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.
- (56) Huang, Y.; Fang, G.; Liu, L. *Natl. Sci. Rev.* **2016**, *3*, 107–116.

## Chapter 1. Rapid Total Synthesis of RNase *B. a.*

The work presented in this chapter is a part of the following manuscript and is reproduced here with permission from John Wiley & Sons.

Mong\*, S.; Vinogradov\*, A.; Simon, M.; Pentelute B. Rapid Total Synthesis of DARPin pE59 and RNase *B. a.* *ChemBioChem.* **2014**, 15, 721–733. DOI: 10.1002/cbic.201300797

\*: authors contributed equally

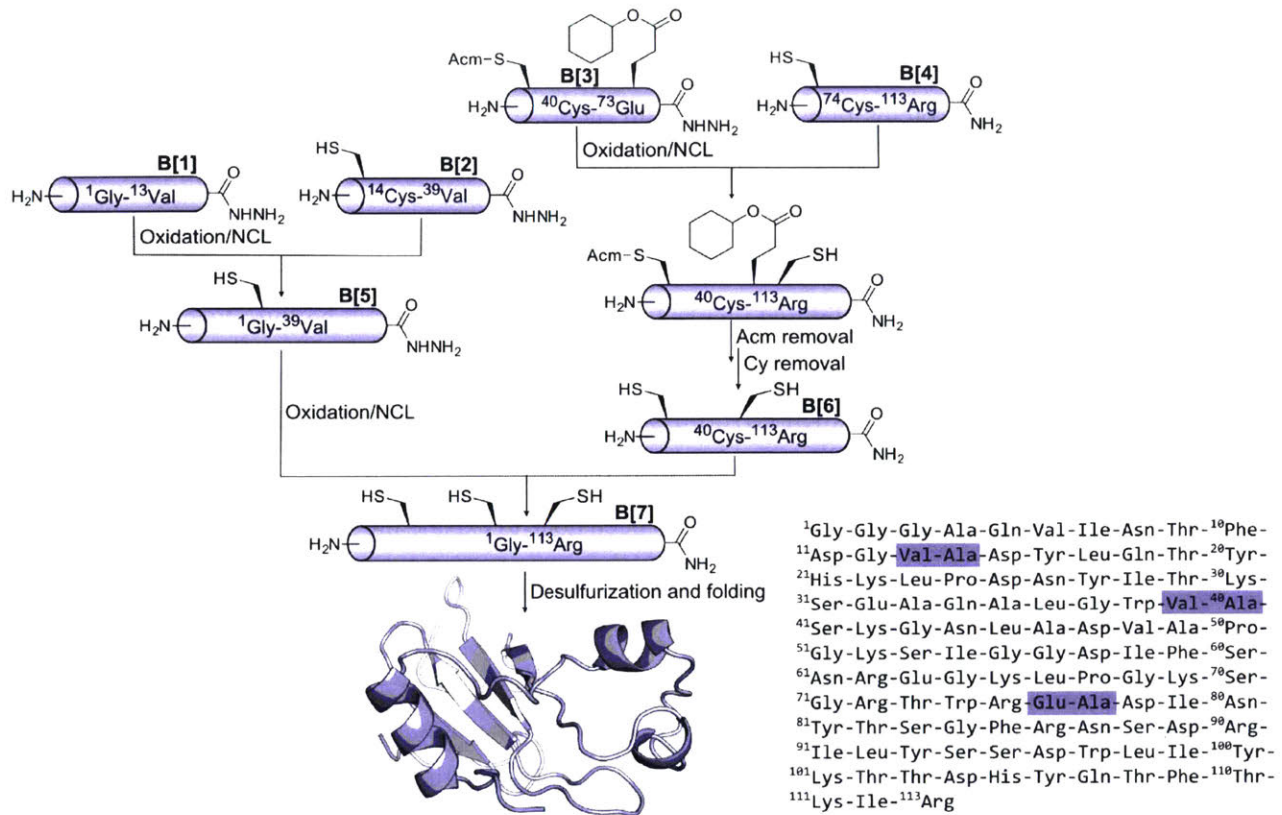
## 1.1. Introduction

Chemically modified proteins play a vital role in modern science, especially in the pharmaceutical sector and chemical biology research.<sup>1-11</sup> However, our ability to selectively modify the properties of proteins is often restricted by the use of biological expression systems. Although impressive advances have been made towards the routine incorporation of select non-proteogenic amino acids with *in vitro*<sup>12</sup> and *in vivo*<sup>13,14</sup> expression systems, only a limited number of unnatural moieties can be introduced. In contrast, total synthesis enables complete chemical control and the ability to tailor a protein's properties as desired. Using solid phase peptide synthesis (SPPS) to assemble smaller fragments, it is possible to construct moderately sized proteins through the use of native chemical ligation (NCL).<sup>15,16</sup>

Since Kent's first report of NCL, much work has been done to expand the scope of the reaction through auxiliary thiols,<sup>17-21</sup> desulfurization,<sup>22,23</sup> and other strategies.<sup>24-29</sup> These efforts have resulted in robust, flexible ligation chemistry. However, obtaining the high quality peptide fragments required for ligation can prove to be difficult and time consuming. Fmoc-based SPPS, often preferred for its safety and the low cost of reagents, typically takes 60 minutes or more to incorporate each amino acid residue.<sup>30</sup> Furthermore, sequence-specific optimization is commonly required, thus necessitating repeated synthesis of fragments under different conditions.

In order to accelerate this process, we have developed a flow based peptide synthesizer<sup>55</sup> that incorporates an amino acid residue every 3 minutes. With our fast flow peptide synthesis methodology well validated for shorter peptides, we undertook the total synthesis of two proteins, a 130-residue Designed Ankyrin Repeat Protein (DARPin) and the 113-residue RNase *B. a.* (barnase, EC 3.1.27). The syntheses of fragments for these two proteins serve as case studies for the application of our fast flow synthesizer methodology towards longer and difficult peptides. This chapter will primarily focus on the synthesis of barnase.

Barnase, a small 12 kDa protein produced and secreted by *Bacillus amyloliquefaciens*,<sup>31,32</sup> is a potent RNase with endonuclease activity and a prime requirement for GpN nucleotides with the selectivity A > G >> C > U for N.<sup>33</sup> Barnase is composed of one polypeptide chain, free of disulfide bonds, and does not require cofactors or metal ions for its folding and catalytic activity.<sup>34</sup> These factors make barnase an appealing candidate for future *in vivo* model studies. Our initial strategy was to assemble barnase from three peptide fragments of equal length, but we ultimately opted for convergent synthesis from four fragments, as shown in Figure 1.1. A triglycine motif was appended to the N-terminus in order to facilitate future enzymatic ligation by the transpeptidase, Sortase A. <sup>76</sup>Glu was protected with the non-standard cyclohexyl ester in order to prevent possible isomerization during NCL.<sup>32,35</sup>



**Figure 1.1. Synthetic scheme of barnase**  
 Protein sequence is shown on the right. Ligation sites are highlighted in blue.

## 1.2. Results

### 1.2.1. Initial Peptide Synthesis

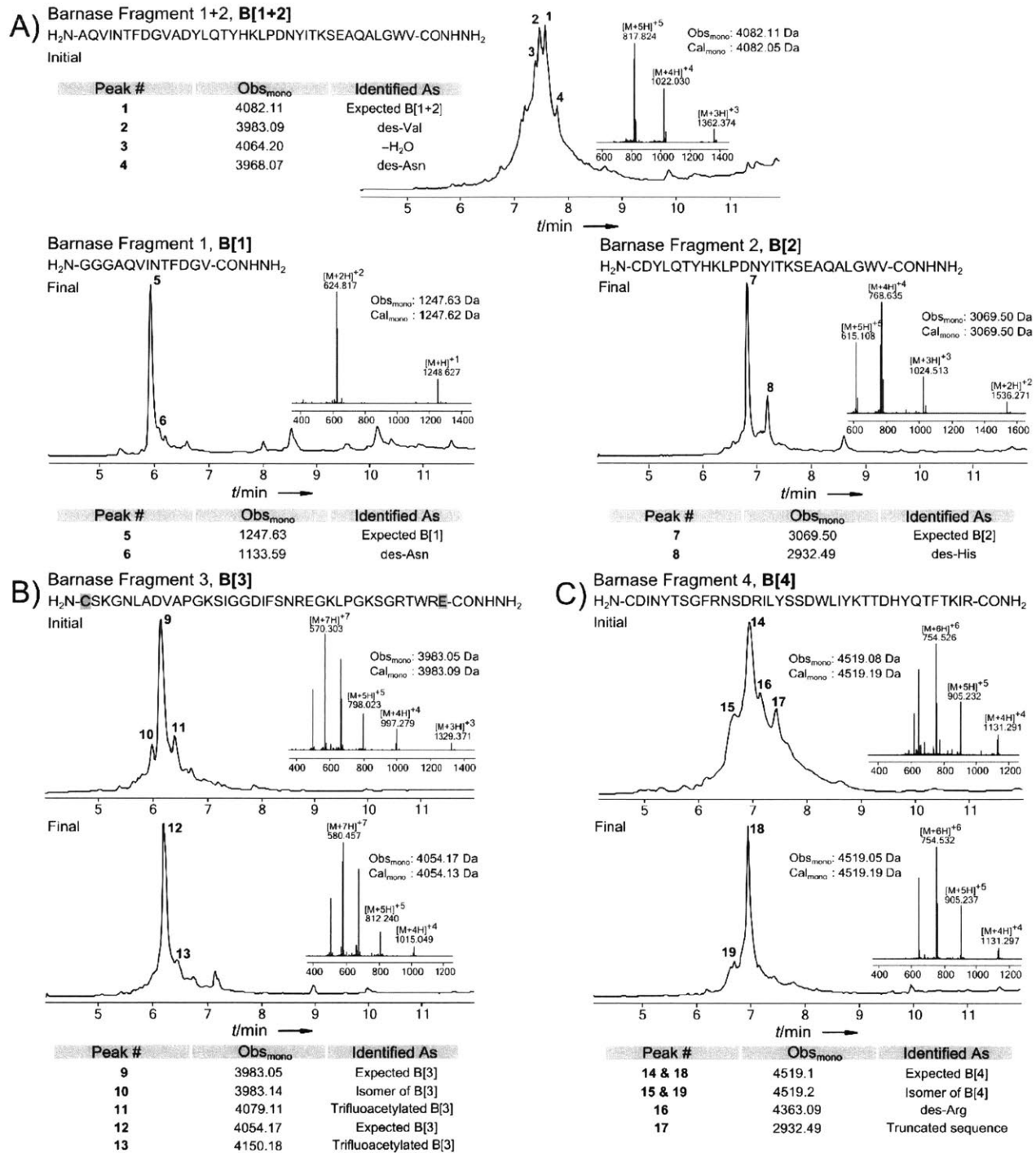
Barnase was initially split into three fragments: B[1+2] ( $\text{H}_2\text{N}-[{}^4\text{Ala}-{}^{39}\text{Val}]-\text{CONHNH}_2$ ), B[3] ( $\text{H}_2\text{N}-[{}^{40}\text{Cys}-{}^{76}\text{Glu}]-\text{CONHNH}_2$ ) and B[4] ( $\text{H}_2\text{N}-[{}^{74}\text{Cys}-{}^{113}\text{Arg}]-\text{CONH}_2$ ). Ultimately, B[1+2] was divided into B[1] ( $\text{NH}_2-[{}^1\text{Gly}-{}^{13}\text{Val}]-\text{CONHNH}_2$ ) and B[2] ( $\text{NH}_2-[{}^{14}\text{Cys}-{}^{39}\text{Val}]-\text{CONHNH}_2$ ).

The initial crude barnase peptides synthesized on a standard three minute cycle (section 1.4.2) are shown in Figure 1.2. B[1+2], B[1], B[2], and B[3] were synthesized as peptide hydrazides on 2-chlorotrityl hydrazine polystyrene resin. B[4] was synthesized as a C-terminal carboxamide on Rink-aminomethyl polystyrene resin.

While crude B[3] was readily purifiable, B[1+2], and B[4] required improvements in the synthetic conditions prior to RP-HPLC. Additionally, we attempted to refine synthetic conditions B[2] in order to increase the yield and explore factors which contribute to peptide quality with our platform. Because our fast flow synthesis platform allowed us to obtain long polypeptides in hours rather than days, we systematically studied numerous synthetic conditions. Ultimately, we made only a few changes to the standard synthetic strategy to achieve significant improvement in crude peptide quality.

Our initial synthetic strategy for barnase included assembly of fragments B[1] and B[2] as a single polypeptide,  $\text{H}_2\text{N}-[{}^4\text{Ala}-{}^{39}\text{Val}]-\text{CONHNH}_2$  (B[1+2]), by stepwise SPPS. The first synthesis of this peptide using HBTU as an activating agent yielded crude material of unsatisfactory quality. The major impurities were aspartimide-derived compounds, and deletion products, including 50% des-Valine (Figure 1.2a). We hypothesized that the des-Val product may have resulted from incomplete coupling of the C-terminal  ${}^{39}\text{Val}$  to the resin, and, thus, we coupled  ${}^{39}\text{Val}$  for 10 minutes during the next synthesis. Additionally, in an attempt to prevent the formation of other deletion products, we employed HATU activation throughout the sequence. These measures generally increased the quality of the crude peptide (Figure 1.9), including elimination of the des-Val product, but an isomer and several deletion products prevented facile RP-HPLC purification. Impressed with the results of HATU in this synthesis, we used it for all other barnase fragments.

Striving to suppress incomplete couplings throughout the synthesis, we doubled the amount of activated amino acid for every coupling step, while maintaining 30 s coupling times by doubling the flow rate. However, this did not improve synthetic quality. In a separate experiment, reducing the amount of DIEA



**Figure 1.2. HPLC-MS data (total ion current vs. time) of crude barnase fragments**

The initial as well as the final syntheses are shown for every peptide. Each panel also displays peptide sequences, MS insets of the major peak, comparison of monoisotopic calculated and observed molecular masses for the expected product, and a table with identification of the major impurities. The charge state series on the inset mass spectra display the most abundant ions; observed and calculated masses are monoisotopic. a) Barnase fragments 1 and 2. b) Barnase fragment 3. Residues highlighted in grey were incorporated with non-standard side-chain protection; Cys(Trt) is utilized in the initial synthesis and Cys(Acm) in the final synthesis; Glu is protected as cyclohexyl ester in both cases. c) Barnase fragment 4.

used to activate the amino acids did not improve the peptide quality (Figures 1.10 and 1.11). At this point, we placed the synthesis of B[1+2] on hold and turned our attention to fragments B[3] and B[4].

Synthesis of the 37-residue B[3] under standard conditions resulted in the high quality crude product shown in Figure 1.2b. The first synthesis of B[3] did not employ the orthogonally protected cysteine ultimately required by our synthetic strategy, so it was resynthesized with AcM protection. Additionally, the resynthesis used 20% (v/v) piperidine as the deprotection reagent, which somewhat suppressed formation of an early eluting isomer. The late eluting trifluoroacetylation product was also less prevalent, for poorly understood reasons.

Preliminary experiments suggested that B[4] would be difficult to synthesize, and prone to aspartimide formation. For these reasons, we conducted the initial synthesis with 20% (v/v) piperidine and 0.1nM HOBt in DMF as the deblocking solution, and used the Rink-AM resin. Unfortunately, the resulting peptide, shown in Figure 1.2c, contained numerous deletion products, a prominent isomer, and a significant premature chain termination product. Because the resin may affect the quality of synthesized peptides, we next assembled B[4] under the same conditions but on PEG resin, which resulted in a significant improvement in the quality of the peptide. Under these conditions, the two major impurities were a well resolved isomer and a product of premature chain termination at <sup>90</sup>Arg (Figure 1.26). Deletion products were nearly eradicated. Although the peptide was of acceptable quality, we attempted to further improve the yield. In a successful attempt to suppress the persistent sequence truncation which had a mass 42.05 Da greater than H<sub>2</sub>N-[<sup>91</sup>Ile-<sup>113</sup>Arg]-CONH<sub>2</sub>, we performed the synthesis of B[4] again, but used DCC/HOBt activation of <sup>90</sup>Arg, instead of HATU. Additionally, we removed HOBt from the deprotection solution throughout the synthesis. As a result, the formation of truncated peptide was alleviated, while removal of HOBt from the deprotection solution had little to no effect on the formation of the isomer (Figure 1.2c). Based on these findings, we selected 20% (v/v) piperidine in DMF as a universal deprotection agent for barnase fragment assembly.

Finally, we revisited synthesis of B[1+2]. Based on the data obtained earlier, we assembled B[1+2] as a C-terminal carboxamide on Rink-PEG resin using 20% (v/v) piperidine for deprotection. Unfortunately, these measures did not improve the quality of the crude product. To identify regions of incomplete Fmoc removal and slow coupling, we employed UV monitoring of the reactor effluent and intermediate cleavage of the polypeptide chain, as described above. Real time monitoring of the deprotection step identified incomplete Fmoc removal in the region <sup>29</sup>Thr – <sup>27</sup>Tyr, and the data obtained from intermediate cleavage at <sup>20</sup>Tyr supported this finding.

At this point, we divided B[1+2] into two peptides: B[1] (NH<sub>2</sub>-[<sup>1</sup>Gly-<sup>13</sup>Val]-CONHNH<sub>2</sub>) and B[2] (NH<sub>2</sub>-[<sup>14</sup>Cys-<sup>39</sup>Val]-CONHNH<sub>2</sub>), concurrently appending an oligoglycine motif to the N-terminus of B[1]. We assembled B[1] and B[2] as C-terminal peptide-hydrazides on 2-chlorotrityl hydrazine polystyrene

resin under standard conditions, except 20% (v/v) piperidine was used as a deprotection reagent and C-terminal Val residues were coupled for 10 minutes. Synthesis of B[1] produced very high quality crude peptide, as shown in Figure 1.2a, and RP-HPLC afforded pure B[1] in good yield. Although crude B[2] was of acceptable quality, we sought to reduce the proportions of several prominent deletion products (Figure 1.21). We hypothesized that prolonged coupling times may suppress deletion products. During the next synthesis of B[2], we deprotected <sup>27</sup>Tyr – <sup>25</sup>Asp for one minute with 20% (v/v) piperidine and coupled <sup>28</sup>Ile for 10 minutes and <sup>27</sup>Tyr – <sup>25</sup>Asp for 20 minutes. This synthesis failed to improve the crude peptide quality (Figure 1.19), and thus we abandoned the strategy of extended coupling times.

Next, we turned to the 2-hydroxy-4-methoxybenzyl (Hmb) protecting group, a well-known N<sup>α</sup>-protecting group which is supposed to disrupt peptide chain aggregation on resin. We identified <sup>33</sup>Ala, the closest alanine to the problematic region, as a suitable site for incorporation of Fmoc-(Hmb)-Ala-OH. Model coupling studies with Fmoc-(Hmb)-Ala-OH on our synthetic platform led us to activate this and the subsequent residue with DCC/HOBt rather than HATU or HBTU, to avoid premature termination of the peptide chain. Further, at 60 °C, coupling of the Hmb backbone protected amino acid required 25 minutes, presumably due to intramolecular formation of the less active Hmb-ester, and coupling to the resulting secondary amine required half an hour. Otherwise, standard conditions were used. Incorporation of the N<sup>α</sup>-protected residue near the region of incomplete coupling drastically improved the quality of the crude peptide, leaving the His deletion as the only significant impurity (Figure 1.2a). RP-HPLC readily afforded the last barnase fragment, B[2], as a homogeneous peptide.

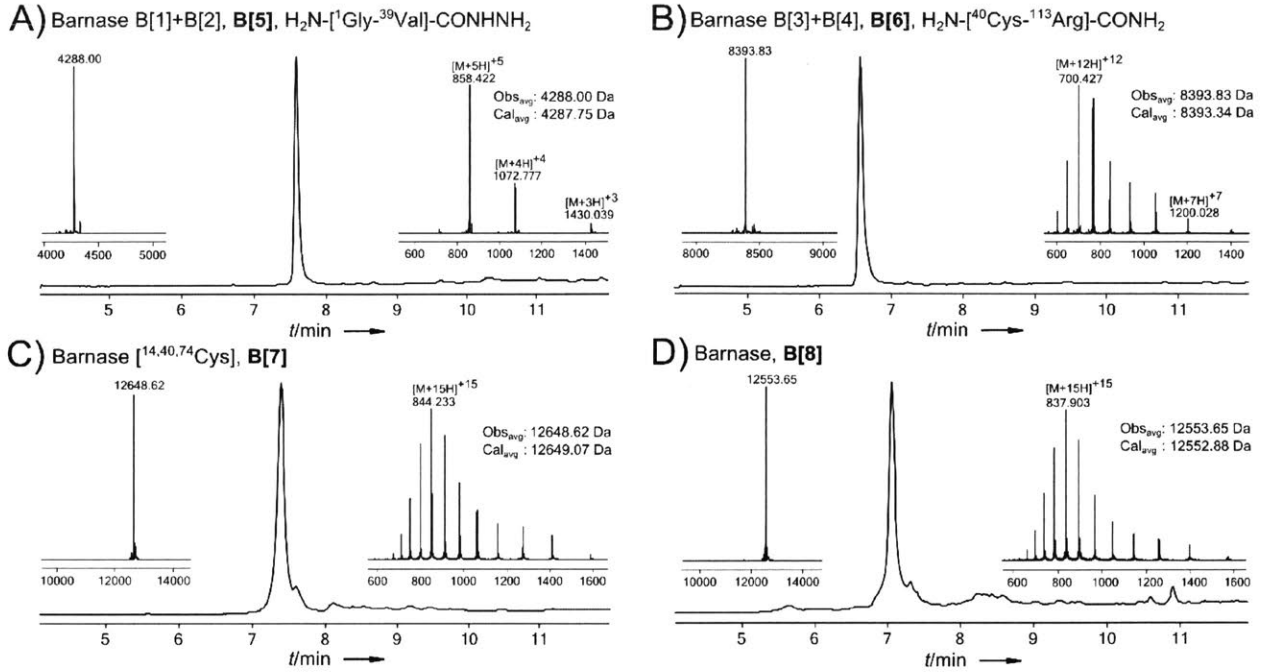
The final crude barnase peptides are shown in Figure 1.2. All final barnase fragments were produced with a standard three minute cycle using HATU activation and 20% piperidine in DMF (v/v) replacing the initial 50% (v/v) piperidine deprotection reagent. B[1], B[2], and B[3] were produced on 2-chlorotrityl hydrazine resin, and B[4] was produced on Rink-PEG resin.

Additionally, several peptide specific modifications were made to the synthesis. For B[1] and B[2], the C-terminal Val were coupled to the resin for 10 minutes. For B[2] <sup>33</sup>Ala, incorporated with Hmb backbone-protection, was coupled for 25 minutes, and the subsequent residue, <sup>32</sup>Glu, was coupled for 30 minutes. Both <sup>33</sup>Ala and <sup>32</sup>Glu were activated with HOBt/DCC, rather than HATU. For B[3], no additional modifications were made. For B[4], <sup>90</sup>Arg was activated with HOBt/DCC. The masses and RP-HPLC purification yields of all the peptides assembled above, as well as overall yield of the synthetic protein, are summarized in Table 1.1.

### 1.2.2. Protein Synthesis via Native Chemical Ligation

Following facile RP-HPLC purification of the crude peptides, we turned our attention to NCL. A one-pot oxidation/NCL procedure reported by Lei Lu and coworkers<sup>36</sup> was employed with only minor modifica-





**Figure 1.3. HPLC-MS data (total ion current vs time) of purified barnase ligation products**  
 Each panel also displays MS insets of the major peak, comparison of average calculated and observed molecular masses for the expected product, and a deconvolution result. The charge state series on the inset mass spectra display the most abundant ions; observed and calculated masses are average. a) Purified N-terminal polypeptide H<sub>2</sub>N-[<sup>1</sup>Gly-<sup>39</sup>Val]-CONH<sub>2</sub>. b) Purified C-terminal polypeptide H<sub>2</sub>N-[<sup>40</sup>Cys-<sup>113</sup>Arg]-CONH<sub>2</sub>. c) Purified full length barnase [<sup>14,40,77</sup>Cys]. d) Purified full length, native, desulfurized barnase.

**Table 1.1. Yields of peptides and overall yield of synthesized barnase**

Fragments	Crude peptide, mg	Purification Yield, %
B[1]	183	39
B[2]	340	30
B[3]	308	36
B[4]	115	28
Barnase	N/A	12 (3) <sup>a</sup>

[a] Yield is calculated from purified starting fragments and excludes the purification yield of the fragment peptides. Yield in parenthesis is calculated from crude peptides and include the purification yield of the peptide.

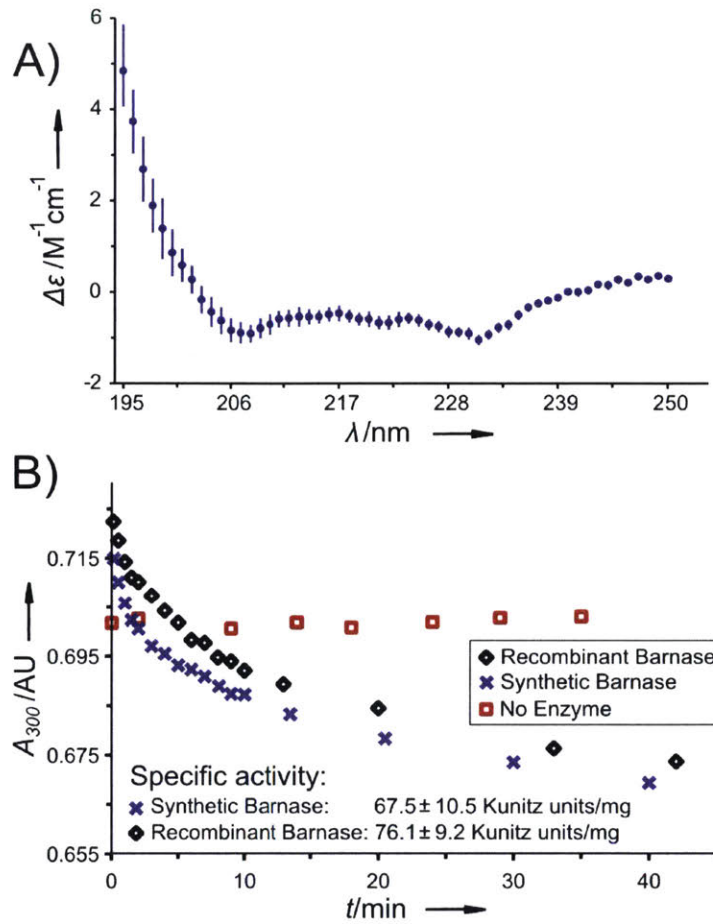
tions. This procedure relies on oxidation of C-terminal hydrazides to the azide, followed by addition of thiol to generate thioesters for NCL. Thus, the N-terminal fragment B[1] was dissolved in the ligation buffer consisting of 0.2 M phosphate buffer (pH 3.1) with 6 M Gn·HCl and oxidized with 20 mM NaNO<sub>2</sub>. 100 mM 4-Mercaptophenyl acetic acid (MPAA) was added to yield the C-terminal thioester.<sup>37</sup> The second peptide, B[2], was added to afford ligated fragment B[5], H<sub>2</sub>N-[<sup>1</sup>Gly-<sup>39</sup>Val]-CONHNH<sub>2</sub> after adjusting the final pH of the reaction mixture to 7.0. The C-terminal segment H<sub>2</sub>N-[<sup>40</sup>(Acm)Cys-<sup>113</sup>Arg]-CONH<sub>2</sub> was prepared in an analogous manner from peptides B[3] and B[4]. A typical ligation reaction was allowed to proceed for 16 hours at 1.5 mM peptide concentrations and at pH 7.0. Ligation yields and purification procedures are additionally summarized in section 1.4.6. All ligated products were isolated in high yield as pure compounds after RP-HPLC (Figure 1.3).

In order to prepare peptides for the final ligation, C-terminal segment H<sub>2</sub>N-[<sup>40</sup>(Acm)Cys-<sup>113</sup>Arg]-CONH<sub>2</sub> was subjected to silver (I) acetate in 50% aqueous acetic acid to remove Acm protection on the N-terminal cysteine,<sup>38</sup> and the cyclohexyl ester of <sup>76</sup>Glu was removed using 10% v/v *p*-cresol in HF to prepare protecting group free peptide B[6] for ligation. NaNO<sub>2</sub> oxidation, MPAA thioesterification, and subsequent ligation of peptide B[5] with B[6] afforded the full length protein.

Finally, to obtain the native sequence, the desulfurization protocol developed by Danishefsky and coworkers<sup>22</sup> was used to globally convert Cys to Ala. Barnase [<sup>14,40,77</sup>Cys] was reacted with excess TCEP and Mesna in the presence of VA-044 radical initiator, affording the native protein (Figure 1.3d). After barnase was successfully refolded from a 6M Gn·HCl solution (characterized by circular dichroism spectroscopy; Figure 1.4a), we characterized the enzymatic activity of the synthetic protein and determined it to be comparable to the recombinant analog. Barnase is a potent RNase, and its activity was assayed by measuring the catalytic hydrolysis of Yeast *Torula* RNA according to an RNase hydrolytic assay described by Kunitz.<sup>39</sup> We found that the specific activity of the synthetic barnase was 67.5 ± 10.5 Kunitz units/mg (Figure 1.4b), which corresponds to the values we obtained for the recombinant analogue (76.1 ± 9.2 Kunitz units/mg).

### 1.3. Discussion

The total convergent synthesis of barnase serves as a case study in the application of our fast flow peptide synthesizer to the rapid and facile assembly of longer and difficult sequences via Fmoc SPPS. Initial syntheses of barnase peptide fragments on our system identified two major classes of impurities: point deletions, and aspartimide-derived products. Additionally, in isolated instances, we observed carboxylation, premature chain termination, and methionine oxidation. First, the two major classes of impurities are addressed, and then miscellaneous problems are discussed. Finally, major side reactions encountered during oxidation and thioesterification of C-terminal peptide hydrazides are described.



**Figure 1.4. Characterization of the synthesized protein**

a) CD spectrum of the synthesized barnase. c) Barnase RNase activity assay; specific activity was calculated using additional data points (section 1.4.4). RNA hydrolysis rate was responsive to the enzyme concentration in all experiments.

### 1.3.1. Deletion Products

Impurities arising from incomplete acylation are the most common side products in peptide synthesis.<sup>40</sup> As the polypeptide grows, these products constitute a greater fraction of the crude material and are less resolved by RP-HPLC. This fact limits the maximum length of a polypeptide prepared by stepwise SPPS, if homogenous material is desired. Many strategies are employed to reduce deletion side products and in the following sections we discuss some common approaches in the context of our fast flow platform.

*Changing the Activating Agent.* As our previous work indicates,<sup>55</sup> the use of HATU over HBTU as an activating agent may improve synthetic outcomes with our flow synthesizer. Because of this, it was often our first option to suppress deletion products. In the synthesis of B[1+2] we observed a substantial improvement in the crude peptide quality with HATU instead of HBTU activation, and HATU was generally adopted for synthesis of barnase fragments. In contrast, during the synthesis of DARPin fragments the use of HATU showed little or no improvement when individual amino acids or the entire peptide was assembled with HATU activation. With our flow synthesizer, HATU activation does not always improve synthetic outcomes, but changing the activating agent is an easy strategy to implement and a good response to the poor synthetic quality of a peptide.

*Increased Coupling Time.* One of the most common protocols to overcome difficult couplings is to increase the coupling time, with or without adding fresh reagents (*i.e.*, double coupling).<sup>30</sup> To examine the efficacy of these strategies in the context of our rapid synthesis platform, we synthesized peptide B[1+2], with extended coupling times on stretches of problematic residues, and on individual residues in B[1], and B[2]. The results were mixed. In B[1+2], B[1], and B[2], coupling of the first residue, a C-terminal Val, to the resin proceeded to completion within 10 minutes, eliminating a 50% valine deletion product. All other deletions were essentially unaffected by increased coupling time. Although extended couplings are time consuming and typically ineffective, they are extremely easy to implement, and a reasonable early response to deletion products with our flow system.

*Resins.* The nature of the solid support may have a drastic effect on the outcome of peptide synthesis.<sup>41</sup> While working with B[1+2] and B[4], we screened several resins, including polystyrene resins with various functionalities and a polyethylene glycol resin. We found that the synthesis of B[4] on PEG resin radically improved the crude quality compared to other resins. However, the use of this resin did not alter the outcome of B[1+2] synthesis. In general, we expect the success of this strategy to be highly sequence dependent, but changing resins can be strikingly effective.

*Backbone Protection.* In one case, B[2], the above measures did not produce the crude peptide of the desired purity. For this reason we explored the use of backbone protection strategy and incorporated N<sup>α</sup>-protected Fmoc-(Hmb)-Ala-OH during the SPPS. N<sup>α</sup>-Hmb residues are known to drastically improve the

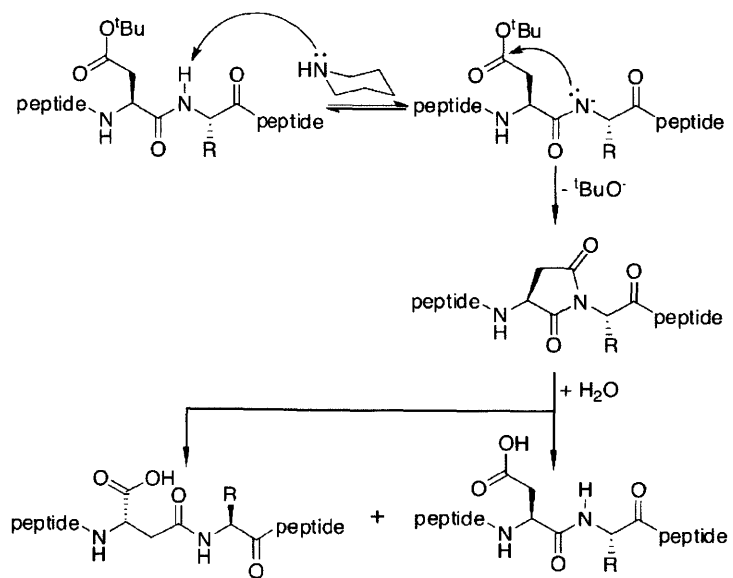
synthesis of difficult sequences,<sup>42,43</sup> presumably by disrupting chain aggregation. Using procedures outlined above, we identified <sup>29</sup>Thr – <sup>27</sup>Tyr as the difficult stretch, and installed Fmoc-(Hmb)-Ala-OH at <sup>33</sup>Ala, which vastly improved the quality of the synthesis. We reserved N<sup>α</sup>-protected amino acids for special cases because they are relatively difficult to prepare, and quite expensive when obtained commercially. When used, however, these amino acid derivatives were highly effective.

### 1.3.2. Aspartimide Formation

Another common side reaction is formation of a five membered imide ring with the aspartic acid side chain and adjacent N<sup>α</sup>-nitrogen as shown in Figure 1.5.<sup>44,45</sup> The 5-membered ring can then be opened by a variety of nucleophiles, such as water, piperidine, or methanol. Ring opening can occur at either of two electrophilic sites, producing two isomers from each reaction. As a result, aspartimide leads to multiple products in the crude peptide. Formation of the imide is reported to be more common at sterically less hindered Asp-Ala, Asp-Gly, and Asp-Asn sites.<sup>44</sup> Below, several known strategies to reduce formation of this side product are evaluated for use in our fast flow system.

*Reduced Piperidine.* To reduce highly undesirable aspartimide formation, we first turned to the most straightforward solutions: lower piperidine concentrations, shorter deprotection steps, and addition of HOBt to the deprotection reagent.<sup>46</sup> In B[3] and B[4], reducing the concentration of piperidine from 50% to 20% (v/v in DMF) significantly reduced the formation of isomers, without otherwise affecting the synthesis. For this reason, 20% piperidine in DMF was adopted as the standard deprotection reagent for synthesis of barnase fragments. Further reduction in aspartimide products was obtained by deprotecting for 20 s with 20% v/v piperidine and 0.1 M HOBt in DMF, a reagent known to suppress aspartimide formation.<sup>47</sup> We expect the use of less piperidine, with or without added HOBt, to be generally effective against aspartimide formation in our flow synthesizer.

*Alternate Deprotection Base.* Another reported solution to the aspartimide problem is the use of secondary amines, such as piperazine and morpholine, which are less basic than piperidine but still sufficiently strong to remove Fmoc.<sup>46,48</sup> Unfortunately, piperazine is only soluble to 5% (w/v) in DMF. The use of volatile DCM is incompatible with our flow system as this solvent cannot be effectively drawn into the pump, and would boil in the heated synthesis vessel. Using 5% piperazine solution in DMF resulted in slow (over 60 sec) Fmoc removal, so we decided that piperazine was unsuitable for use in our rapid flow synthesizer. Next, we turned to morpholine, which is less basic than piperazine, but miscible with DMF. Predictably, Fmoc removal with any proportion of morpholine in DMF was extremely slow and this deprotection strategy was abandoned. We were unable to implement alternate bases for Fmoc removal without substantially increasing cycle times; piperazine and morpholine are unlikely to be viable with our rapid synthesis platform.



**Figure 1.5. Base-promoted formation of aspartimide and its hydrolysis during Fmoc SPPS**  
 Key aspartimide is highlighted in red.

### 1.3.3. Miscellaneous Problems

*Reversible Carboxylation.* In barnase fragments, we observed addition of CO<sub>2</sub> (+43.99 Da relative to the expected product) as an intense, close eluting impurity. This impurity likely arose from the carboxylation of tryptophan and was found to disappear upon incubation in acetonitrile/water with 0.1% TFA.<sup>49</sup>

*Premature Chain Termination.* We identified a truncated impurity in the crude product of B[4] synthesis. The molecular weight of the impurity indicated that the chain was abruptly capped with a 42.05 Da moiety immediately before <sup>90</sup>Arg. Although we could not detect any acetic acid in the reagents upon MS and NMR analysis, changing the activating reagent mitigated the problem. Thus, when we activated <sup>90</sup>Arg with DCC/HOBt instead of HATU, the truncated product disappeared. As we only observed one major truncation product during our studies, we refrain from commenting on the general utility of this strategy to suppress chain termination with our flow system.

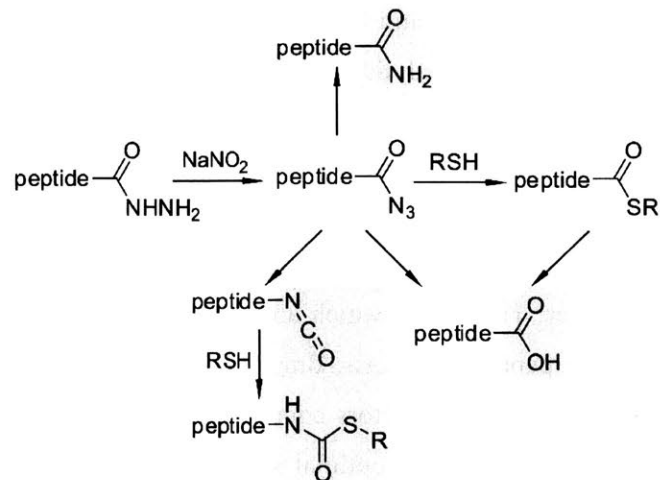
*Trifluoroacetylation.* We observed trifluoroacetylation (+95.98 Da) in several peptide fragments. It is reported to be a prominent side product of Fmoc synthesis, formed during cleavage and global side chain deprotection in TFA.<sup>50</sup> The levels of this side product were highly variable, even amongst almost identical syntheses, and we were unable to eliminate it by incubating the peptide at pH 10 overnight. The impurity elutes very late during RP-HPLC and is easily removed during purification.

### 1.3.4. Oxidation of Peptide Hydrazides and Native Chemical Ligation

During the ligation steps of this work, we identified several side products necessitating minor optimization of Lu's oxidation/thioesterification protocol (Figure 1.6).<sup>36</sup> The most severe was the almost quantitative formation of a side product with molecular mass 15.00 Da greater than the expected thioester upon oxidation/MPAA thioesterification of B[1]. After additional experimentation (Section 1.4.6), we concluded that this side-product resulted from a Curtius rearrangement of the C-terminal azide to an isocyanate, followed by addition of MPAA to yield undesired carbamothioate, as shown in Figure 1.6. We screened the temperature, pH and duration of the reaction to no avail. Fortuitously, we discovered that the formation of carbamothioate can be avoided if the azide is allowed to stand at room temperature with MPAA for 20 minutes before adding the second peptide and adjusting the pH from 3 to 7. This side reaction is explored in detail in chapters 3 and 4.

Additionally, we identified several less prominent side reactions, including formation of C-terminal carboxamide, hydrolysis of the azide to yield a C-terminal carboxylate, and products of incomplete oxidation. Formation of these species is well-documented in acyl azide chemistry.<sup>51-54</sup> Carboxamide and carboxylate products were especially pronounced in the case of B[5] oxidation. Following suggestions from





**Figure 1.6. Transformations of C-terminal peptide hydrazides during oxidation with sodium nitrite**  
Major identified side products are highlighted in grey.

the literature,<sup>53</sup> we reduced the temperature and the pH. While reduction of the pH was unsuccessful, lowering the temperature to  $-30\text{ }^{\circ}\text{C}$  alleviated side reactions and afforded MPAA thioesters of barnase fragments suitable for subsequent ligation.

### 1.3.5 Conclusion

In summary, we have successfully synthesized the native RNase *B. a.* protein by Fmoc chemistry on a novel fast flow peptide synthesizer. The rate at which amino acids could be incorporated allowed us to explore numerous conditions for peptide synthesis, dramatically improve the synthetic quality of our peptides, and systematically study important factors contributing to the successful synthesis of difficult sequences using this platform. Using identified optimal synthetic conditions, high-quality crude peptide fragments were produced in a rapid manner. Convergent NCL via peptide hydrazides afforded the full-length protein, which, after desulfurization and refolding, was characterized and found to be biologically active.

This work highlights the capacity of the fast flow peptide synthesis methodology to quickly and reliably make high quality peptide fragments, and serves as a guide to sequence specific optimization of rapid peptide synthesis. We present an extensive analysis of the utility of various strategies to improve synthetic outcomes with our fast flow platform and report conditions that effectively mitigate or eliminate deletion and aspartimide-derived products, while incorporating most amino acid residues in three minutes.

We believe this work represents a major step towards routine total synthesis of proteins. Our fast flow platform overcomes the primary bottleneck in protein synthesis — the rate at which pure peptides can be obtained. During the course of this study, we were able to routinely assemble  $\sim 30$  residue polypeptides in a matter of hours, rather than days. We coupled approximately 500 amino acids, which translates to about 30 hours of linear synthetic time; using traditional Fmoc SPPS this task would require about 500 hours (60 working days) of continuous effort. Such rapid synthesis allows peptides to be made multiple times, under different conditions, to obtain crude material of exceptional quality, simplifying purification and increasing yields. All barnase fragments were rapidly purified by RP-HPLC yielding pure compounds in greater than 28% isolated yields, which is comparable to the expected outcomes for the state-of-the art methods. Using standard Fmoc chemistry, we were able to assemble peptides of excellent crude quality as long as 37 residues, and it seems that even longer fragments, often necessary for protein synthesis, could be routinely obtained.

We believe this work will make the total chemical synthesis of proteins significantly faster and more appealing to a broad scientific community.

## 1.4. Experimental

### 1.4.1. General

N<sup>α</sup>-Fmoc amino acids, Rink amide linker, 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU), and hydroxybenzotriazole (HOBt) were purchased primarily from Chem-Impex International. Additionally, some of the amino acids used were obtained through Advanced ChemTech and Novabiochem.

2-Chlorotrityl chloride resin (200-400mesh, 1.2mmol/g) was purchased from Chem-Impex International and used to prepare 2-chlorotrityl hydrazine resin. 4-Methylbenzhydramine (MBHA) functionalized cross-linked polystyrene resin (particle size unspecified, 1.0 mmol/g) and TentaGel S-NH<sub>2</sub> polyethylene glycol-polystyrene hybrid resin (140-170 mesh, 0.29 mmol/g) was purchased from Anaspec. Amino methyl resin was prepared from S-X1 Bio-Beads from Bio-Rad (200-400 mesh). Rink amide-ChemMatrix polyethylene glycol resin was purchased from Matrix Innovation (35-100 mesh, 0.49 mmol/g).

*N,N*-Dimethylformamide (DMF), dichloromethane (DCM), methanol (MeOH), diethyl ether, and HPLC-grade acetonitrile were from EMD Millipore. Trifluoroethanol (TFE), trimethylsilyl chloride (TMSCl), sodium borohydride, and 1,2-Ethanedithiol (EDT) were from Alfa Aesar. Methane sulfonic acid was from Fluka. Solvents for LC-MS were purchased from J.T. Baker and Fluka. All other reagents were purchased from Sigma-Aldrich.

Common solvent mixtures used throughout these experiments are: 0.1% (v/v) TFA in water (A), 0.1% (v/v) formic acid in water (A'), 0.1% (v/v) TFA in acetonitrile (B), and 0.1% (v/v) formic acid in acetonitrile (B').

### 1.4.2. Peptide Synthesis Protocols

All peptides were synthesized on the flow-based platform<sup>55</sup> using a 2<sup>nd</sup> generation vessel (2 mL volume). Unless noted, all reagents (coupling, wash, and deprotection) were preheated to 60°C immediately before reaching the synthesis vessel. Final conditions for syntheses using this flow-based platform varied depending on the peptide and are detailed in the following sections. However, the following standard 3-minute cycle served as a starting point for the assembly of RNase *B. a.* fragments:

*Standard 3-minute cycle at 60°C:*

- i) Amide bond formation (coupling) – 30 sec at 6 mL/min
- ii) Removal of coupling reagent (wash) – 60 sec at 20 mL/min
- iii) N<sup>α</sup>-Fmoc removal (deprotection) – 20 sec at 20 mL/min

- iv) Removal of deprotection reagent and products (wash) – 60 sec at 20 mL/min
- v) Manual manipulations, not included above – 10 sec

*Coupling.* Unless noted, coupling was performed by delivering the following coupling solution at 6 mL/min over approximately 30 s. The coupling solution consisted of 1 mmol of N<sup>α</sup>-Fmoc and side chain protected amino acid dissolved in 2.5 mL of 0.4 M HBTU or HATU in DMF and 0.5 mL of DIEA. In order to minimize racemization of Cys, His, and Trp residues, according to our previous study,<sup>55</sup> only 157-190 μL DIEA were added when activating these amino acids. In all cases, this coupling solution contained at least four equivalents of activated amino acid with respect to the resin. Amino acids were dissolved in activating agent solution up to several hours before use, and DIEA was added within 2 min of use.

Unless noted, side chain protection was as follows: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Trt), Gln(Trt), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu).

*Wash.* Unless noted, excess reagent and reaction by-products were washed out from the synthesis vessel with 20 mL of DMF delivered at 20 mL/min over 1 minute.

*Deprotection.* Unless noted, N<sup>α</sup>-Fmoc protecting groups were removed with 6.6 mL of 20-50% (v/v) piperidine in DMF delivered at 20 mL/min over 20 seconds. For syntheses with modified deprotection times, a flow rate of 20 mL/min is maintained, but more or less deprotection reagent is used.

### 1.4.3. Functionalization of Resins

All peptides were synthesized on one of the following solid supports:

- i) PS – Polystyrene based resin, 200-400 mesh
  - 2-chlorotrityl hydrazine, determined loading of 0.70-0.81 mmol/g
  - Aminomethyl (AM), determined loading of 1.2 mmol/g
  - MBHA, determined loading of 0.8 mmol/g
- ii) PEG – Polyethylene glycol based resin, 35-100 mesh, Rink amide-ChemMatrix with a specified loading of 0.49 mmol/g,
- iii) PEG-PS – Polystyrene based resin with a polyethylene glycol spacer, 140-170 mesh, Tentagel-S-NH<sub>2</sub>, stated loading of 0.29 mmol/g

In order to obtain peptides with C-terminal hydrazides for ligations, peptides were synthesized directly on PS resin functionalized with a 2-chlorotrityl hydrazine linker. In order to obtain peptides with a C-terminal carboxamide, a Rink linker was used. If the rink linker was not already installed on the resin, it was coupled as the first residue with a standard three minute cycle.

*Preparation of 2-chlorotrityl hydrazine resin.* 2-Chlorotrityl hydrazine resin was synthesized from 2-chlorotrityl chloride resin using a known procedure.<sup>56</sup> 2-Chlorotrityl chloride resin (25 g) was stirred and allowed to swell in 150 mL of anhydrous DMF under an inert atmosphere in a 500 mL, three-necked, round-bottom flask. In a separate round-bottom flask, 10 mL of anhydrous hydrazine was added to approximately 50 mL of anhydrous DMF under an inert atmosphere. To this, DIEA was added dropwise until phase separation of the mixture began. This entire hydrazine-DIEA-DMF mixture was then added dropwise to the swollen 2-chlorotrityl chloride resin and stirred for 60 min at room temperature. The reaction was quenched by adding 50 mL of methanol and then stirred for 20 min. The resin slurry was then transferred to a glass fritted funnel and washed with 500 mL aliquots of the following in order: DMF, H<sub>2</sub>O, DMF, diethyl ether. The resin was then dried and determined to have a loading of 0.7 mmol/g.

*Preparation of aminomethyl resin.* Bio-Rad, Bio-beads S-X1 (styrene-divinylbenzene copolymer, 1% crosslinkage, 25 g) and 4.9 g (27.6 mmol) of *N*-hydroxymethylphthalimide were added to 450 mL of DCM in a 1 L round bottom flask. To this, 50 mL of methanesulfonic acid was added and the reaction was stirred gently for 5 hours at room temperature. After 5 hours of stirring, the slurry was transferred to a coarse fritted glass funnel and washed with DCM (1500-2000 mL) and ethanol (1500 mL). The resin was then dried under vacuum for 1 hour. After drying, the phthalimidomethyl-resin was transferred to a 500 mL round bottom flask and suspended in a 200 mL solution of hydrazine monohydrate (20% v/v) in absolute ethanol. A reflux condenser was attached and the solution was refluxed gently for at least 8 hours while gently stirring. The resulting gelatinous material was transferred hot to a glass fritted funnel and washed with hot ethanol (1000-2000 mL) and then hot methanol (1000 mL) in order to completely wash away the white phthalhydrazide precipitate. The resin was then washed with DMF (800 mL), DCM (800 mL), 10% (v/v) DIEA in DMF (500 mL), DMF (600 mL), and DCM (1000 mL). The resin was then dried under vacuum and determined to have a loading of 1.2 mmol/g.

*Determination of resin loadings.* Resin loadings were calculated using a known procedure.<sup>30</sup> Using the flow-synthesizer and the resin of interest, a single amino acid was coupled using a standard 3-minute cycle. Immediately after switching to the deprotection step, the waste stream and subsequent wash was collected for quantification of the piperidine-dibenzofulvene (piperidine-DBF) deprotection by-product. This solution was diluted appropriately and its absorbance at  $\lambda_{301}$  recorded against a proper blank. The concentration of piperidine-DBF adduct and the loading of the resin was calculated using a piperidine-DBF  $\epsilon_{301} = 6000 \text{ M}^{-1}\text{cm}^{-1}$ .

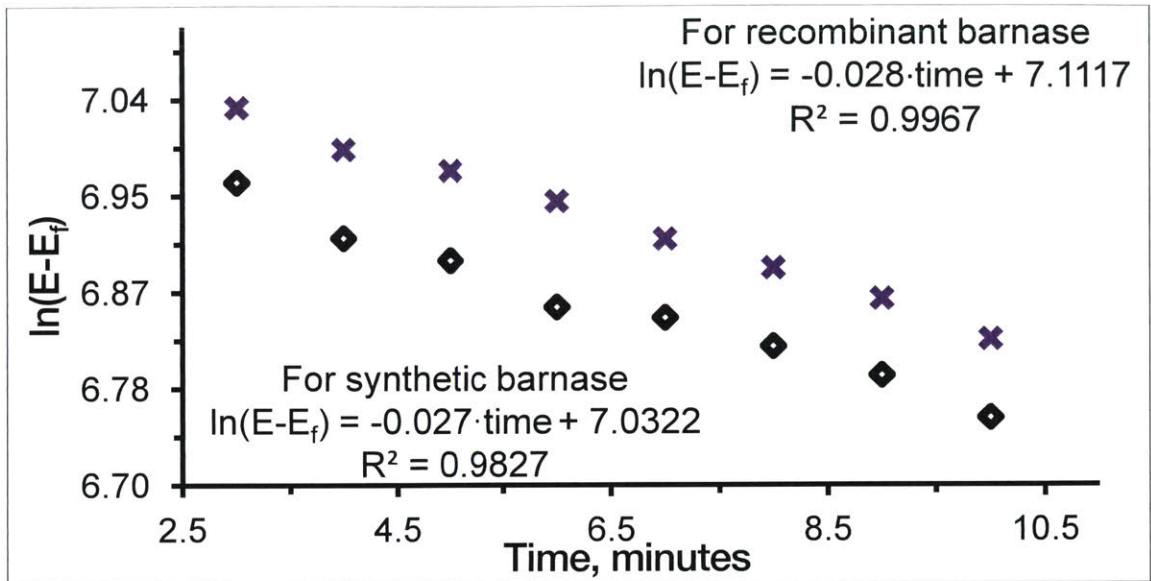
#### 1.4.4. Analytical Methods

*HPLC-MS Analysis.* Unless specifically noted, all peptides and proteins were analyzed on an Agilent 6520 Accurate Mass Q-TOF LC-MS using a narrowbore Agilent Zorbax 300SB C<sub>3</sub> column (300 Å, 5 µm, 2.1 x 150 mm) run with the following method. At 40 °C and a flow rate of 0.8 mL/min, the following gradient was used: 5% B' in A' for 2 min, 5-65% B' in A' ramping linearly over 11 min, 65% B' in A' for 1 minute. MALDI-TOF mass spectra were taken on a PerSeptive Biosystems, Voyager-DE Pro Biospectrometry workstation. Two matrix solutions were used to assist in ionizing peptides and proteins: a saturated solution of  $\alpha$ -cyano-4-hydroxycinnamic acid or 3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid in 50% A / 50% B, respectively.

*Peptide Purification.* All peptides were purified on a Waters 600 HPLC system with a Waters 484 or 486 UV detector using solvents A and B. The following columns were used for purification and are listed in order of decreasing column volume: Agilent Zorbax 300SB preparative C<sub>3</sub> column (300Å, 7µm, 21.2 x 250mm), Agilent Zorbax 300SB C<sub>3</sub> semi-preparative column (300Å, 5µm, 9.4 x 250mm), Phenomenex Jupiter C<sub>5</sub> semi-preparative column (300Å, 5µm, 9.4 x 250mm), Phenomenex Jupiter C<sub>5</sub> analytical column (300Å, 5µm, 4.6 x 250mm), and Agilent Zorbax 300SB C<sub>3</sub> (300Å, 3.5 µm, 4.6 x 150mm) analytical column. Fractions were collected and screened for the desired material using LC-MS and MALDI. Exact flow rates and gradients varied throughout this work, and are detailed individually for each peptide and ligation below.

*Circular Dichroism.* All circular dichroism experiments were performed on an Aviv Model 202 Circular Dichroism Spectrometer. All spectra were obtained at 25 °C scanning from 250 nm down to 190 nm with 1 nm steps and a 15 sec averaging time. The buffer used for CD was 10 mM K<sub>2</sub>HPO<sub>4</sub> (pH 6.4), 50 mM Na<sub>2</sub>SO<sub>4</sub>. An appropriate blank was recorded and subtracted from all spectra.

*Barnase activity assay.* Catalytic hydrolysis of Yeast Torula RNA by Barnase was determined using an assay as previously described.<sup>39</sup> Upon hydrolysis, RNA absorbance at 300 nm decreases, which makes it possible to monitor the hydrolysis. To determine the E<sub>f</sub> values (A<sub>300</sub> for fully hydrolyzed RNA), 0.3 mL of properly diluted (~0.6 µM) enzyme solution were added to a separate 0.7 mL aliquot of assay solution (1.0 mg/mL Yeast Torula RNA in 0.125 M Tris buffer, pH 8.2) and the first kinetic curve was obtained after measuring A<sub>300</sub> on HP 4554A spectrophotometer relative to a proper blank for 2-3 hours. In the second set of experiments 0.3 mL of ~30 nM solution of the enzyme in an appropriate buffer was mixed with 0.7 mL aliquot of assay solution and the rate of the hydrolysis was measured for 10 minutes. Obtained data (Figure 1.7) were used to calculate specific activities and standard errors of both recombinant and synthetic enzymes. The enzyme concentration was determined by the A<sub>280</sub> using the published value:  $A_{0.1\%}^{280} = 2.01$  and was additionally confirmed by CD spectroscopy and from LC-MS quantification.<sup>33</sup> Yeast RNA concentration was verified using UV spectroscopy monitoring the A<sub>260</sub>/A<sub>280</sub> ratio.



**Figure 1.7. The kinetic data for the RNA hydrolysis assay**

Purple crosses are data points obtained for the recombinant enzyme; black squares — for the synthetic. The data was used to calculate specific activity of the enzyme.

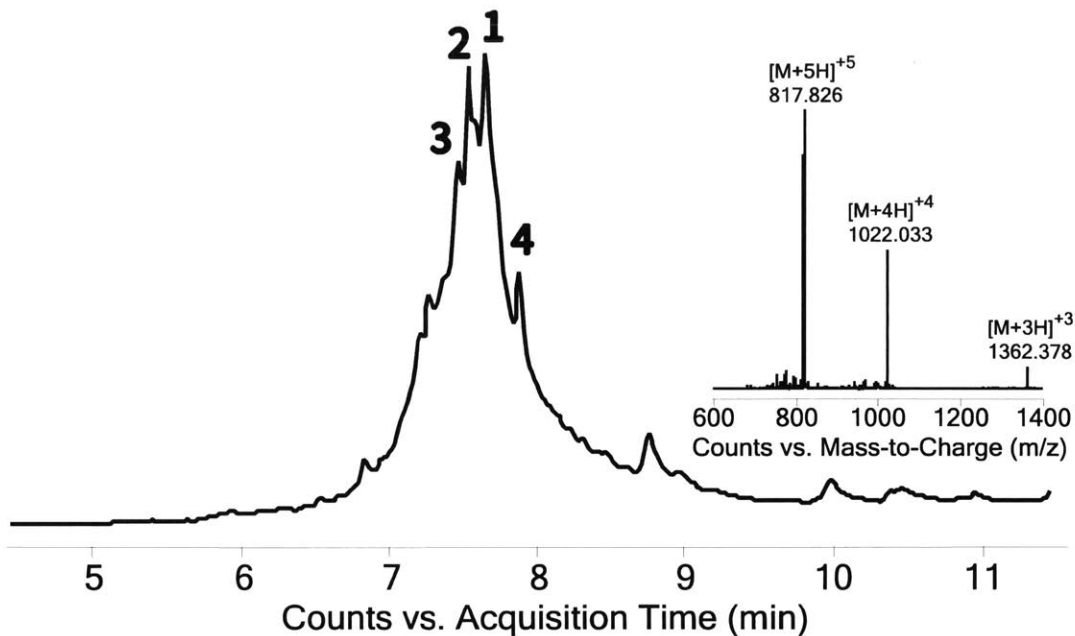
#### 1.4.5. Synthesis of and Purification of Barnase Fragments

*Original synthesis of B[1+2].* The 36-residue N-terminal fragment of barnase was initially synthesized on 185 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HBTU activation, 50% (v/v) piperidine deprotection). We did not attempt purification of this synthesized fragment due to the poor quality of the product (Figure 1.8).

*Synthesis of B[1+2]: HATU activation.* In an attempt to suppress the incomplete coupling of amino acids to the growing peptide chain, B[1+2] was resynthesized on 185 mg of 2-chlorotrityl hydrazine resin using HATU activation instead of HBTU. Additionally, the C-terminal Val was allowed to couple to the resin for 10 min (Figure 1.9). Although the assembled peptide improved relative to that in Figure 1.8, it was still deemed unsatisfactory. Therefore, purification was not attempted.

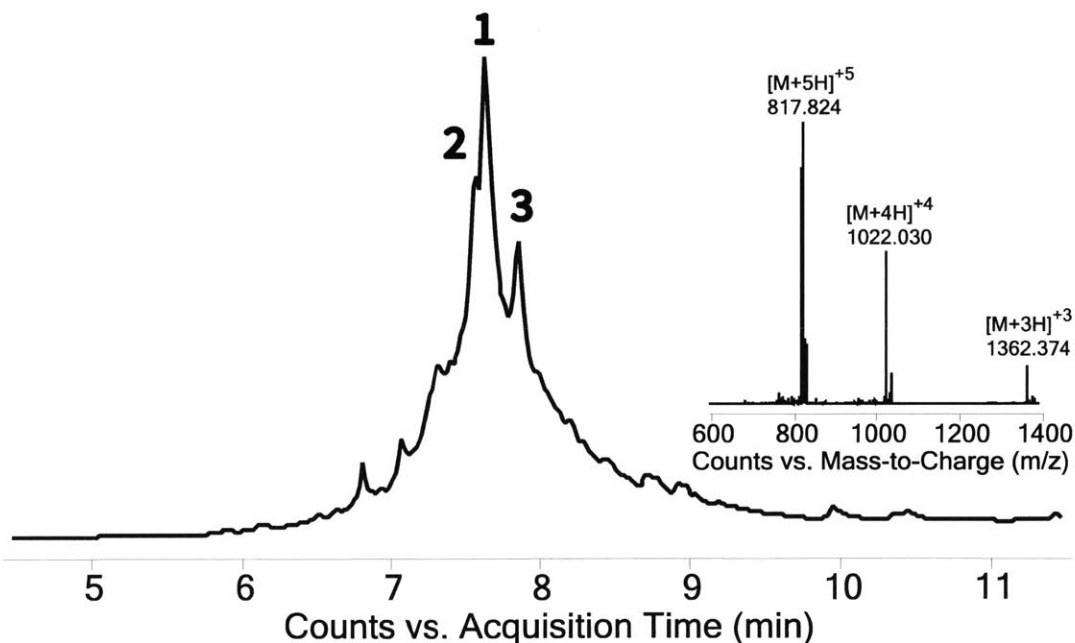
*Synthesis of B[1+2]: 10 equivalents of amino acid.* This synthesis was performed to test a hypothesis that larger quantities of coupling solutions may help with incomplete coupling of particular residues. B[1+2] was resynthesized on 195 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 50% (v/v) piperidine). The amino acid coupling solution consisted of 2 mmol of N<sup>α</sup>-Fmoc and side chain protected amino acid in 5.0 mL of 0.4 M HATU solution in DMF. 1.0 mL of DIEA was added within two minutes of use. Trp and His were activated with 314 μL of DIEA. Amino acids were delivered at double the normal flow rate (12 mL/min) to maintain the 3-minute cycle time. The C-terminal Val was coupled to the resin for 10 minutes. The outcome of this experiment did not support our initial hypothesis as the quality of the crude material did not improve (Figure 1.10).





**Figure 1.8. LC-MS total ion current vs. time for the synthesis of B[1+2] using a standard 3-minute cycle (HBTU, 50% (v/v) piperidine)**

1: Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4082.1 Da, obs. = 4082.0 Da). 2: Val deletion relative to the desired product. 3: Co-eluting isomer of 2 and -18 Da relative to the desired product. 4: Asn deletion relative to the desired product.



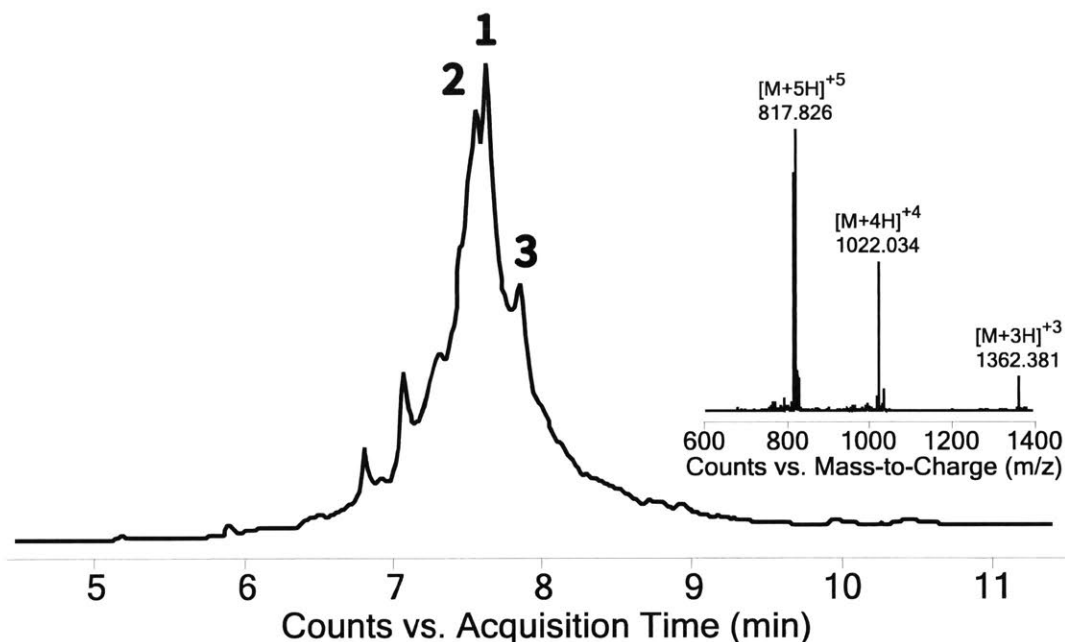
**Figure 1.9. LC-MS total ion current vs. time for the synthesis of B[1+2] using a standard 3-minute cycle (HATU, 50% (v/v) piperidine)**

C-terminal Val was coupled for 10 min. 1: Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4082.1 Da, obs. = 4082.0 Da). 2: Co-eluting isomer of 1 and -18 Da relative to the desired product. 3: Asn deletion relative to the desired product.

*Synthesis of B[1+2]: using less DIEA.* B[1+2] was resynthesized on 150 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 50% (v/v) piperidine), and a lowered amount of activating DIEA for all amino acids. With the exception of glycine, only 157  $\mu\text{L}$  DIEA was added to amino acids being coupled. The C-terminal Val was coupled to the resin for 10 min. Poor quality of the crude peptide and high levels of aspartimide suggest that the use of smaller quantities of DIEA was ineffective and potentially detrimental to peptide assembly (Figure 1.11).

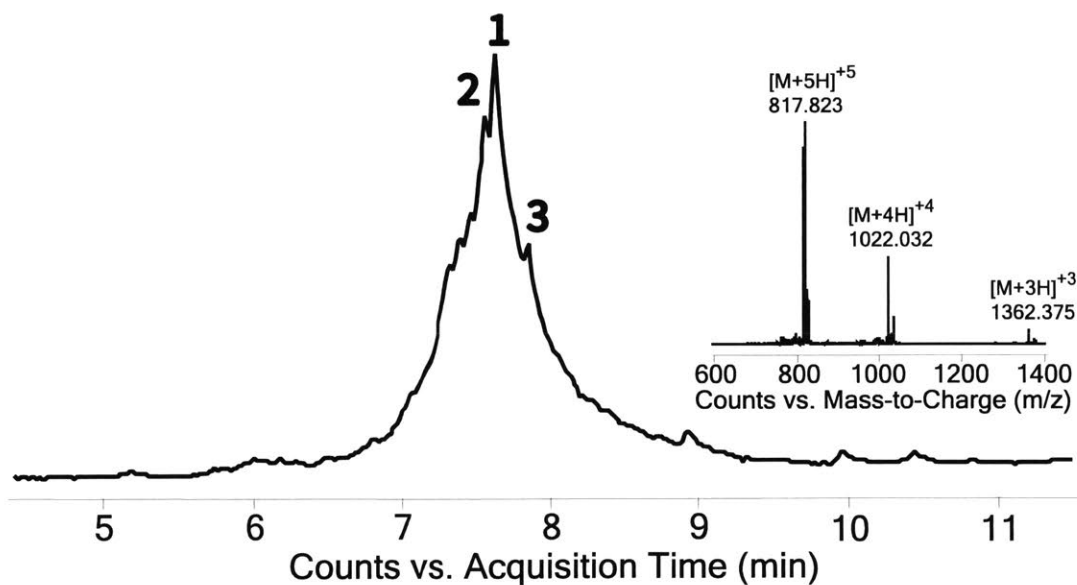
*Synthesis of B[1+2]: PEG resin.* Encouraged by the results of synthesizing some other difficult sequences on ChemMatrix PEG resin, we decided to further investigate the effect of the resin on peptide quality in the case of Barnase fragment B[1+2]. The peptide was resynthesized as a carboxamide on 120 mg of PEG resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine). The C-terminal Val was coupled to the resin for 10 min and Trp and His residues were activated with 165  $\mu\text{L}$  of DIEA. The quality of the resulting crude product (Figure 1.12) is comparable to the one assembled on the polystyrene based 2-chlorotrityl hydrazine resin. In contrast to DARPin fragment D[4], we did not pursue this path for this particular peptide.

*Synthesis of B[1+2]: UV monitoring.* In an attempt to identify stretches of the B[1+2] peptide chain that might be aggregating and causing poor peptide quality, the synthesis of this peptide was repeated while monitoring the reactor effluent in real time using a UV detector as described.<sup>55</sup> The peptide was resynthesized on 130 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 50% (v/v) piperidine). An aliquot of resin was also taken out for a test cleavage after deprotecting the 21<sup>th</sup> residue coupled, <sup>16</sup>Tyr (Figure 1.13). The real time monitoring of piperidine-DBF adduct removal indicated incomplete Fmoc removal under standard conditions in the region of the 11<sup>th</sup> through 13<sup>th</sup> residues coupled, Thr-Ile-Tyr. Exhaustive deprotection was performed on this stretch until no piperidine-DBF adduct was observed via the UV detector. However, after completing the synthesis there was no significant improvement in peptide quality relative to previous experiments (Figure 1.14). This experiment suggests that by themselves, prolonged/multiple deprotection times on troublesome regions are inadequate to improve peptide quality.



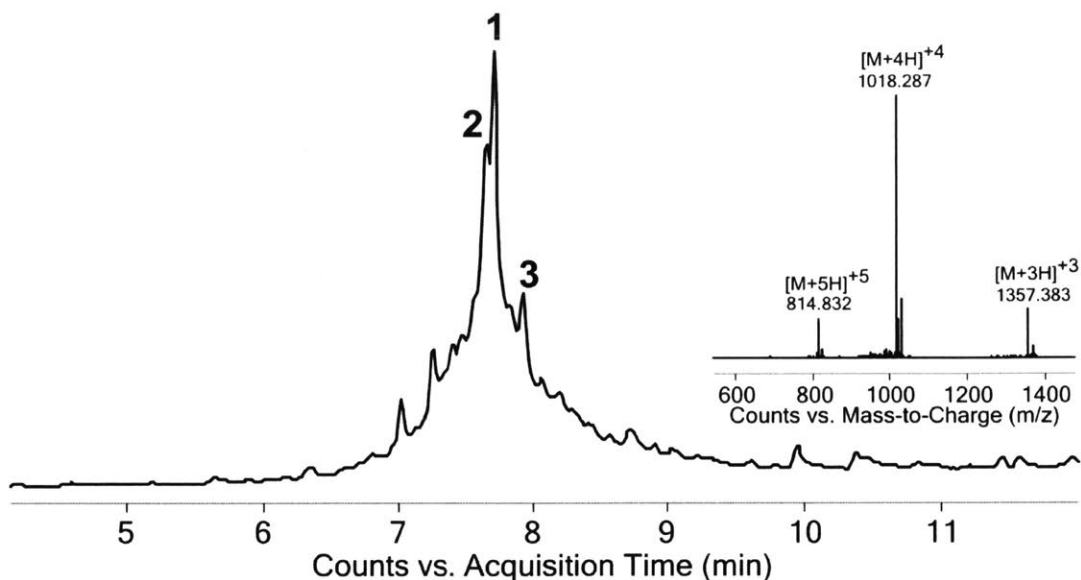
**Figure 1.10. LC-MS total ion current vs. time for the synthesis of B[1+2] using 10 equivalents of amino acids (HATU, 50% (v/v) piperidine)**

C-terminal Val was coupled for 10 min. **1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4082.1 Da, obs. = 4082.0 Da). **2:** Co-eluting isomer of 1 and -18 Da relative to the desired product. **3:** Asn deletion relative to the desired product.



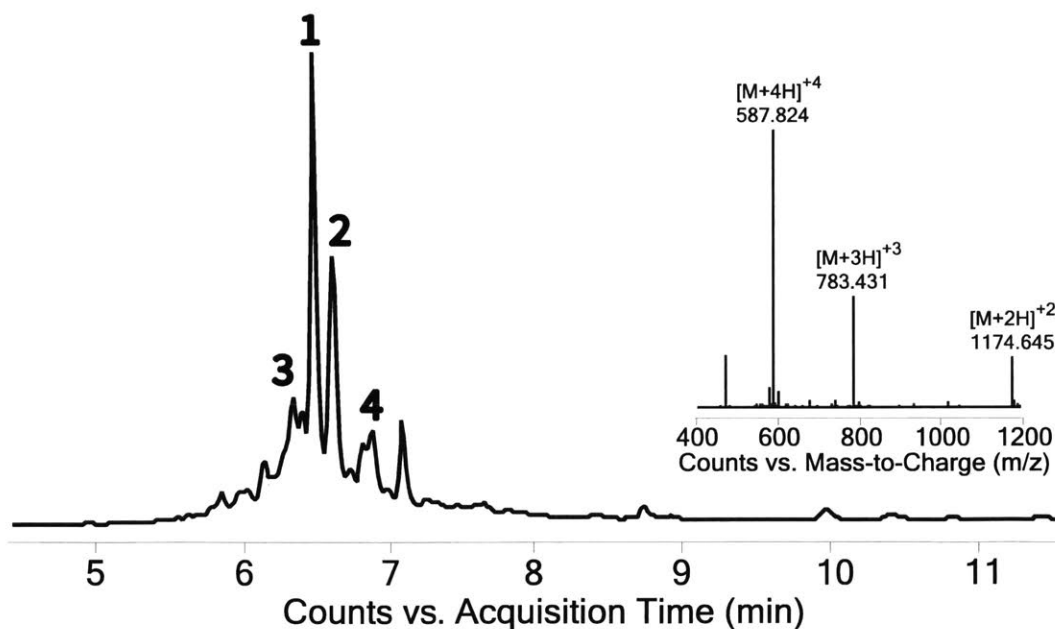
**Figure 1.11. LC-MS total ion current vs. time for the synthesis of B[1+2] using a standard 3-minute cycle (HATU, 50% (v/v) piperidine) and less DIEA**

C-terminal Val was coupled for 10 min. **1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4082.1 Da, obs. = 4082.1 Da). **2:** Co-eluting isomer of 1 and -18 Da relative to the desired product. **3:** Asn deletion relative to the desired product.



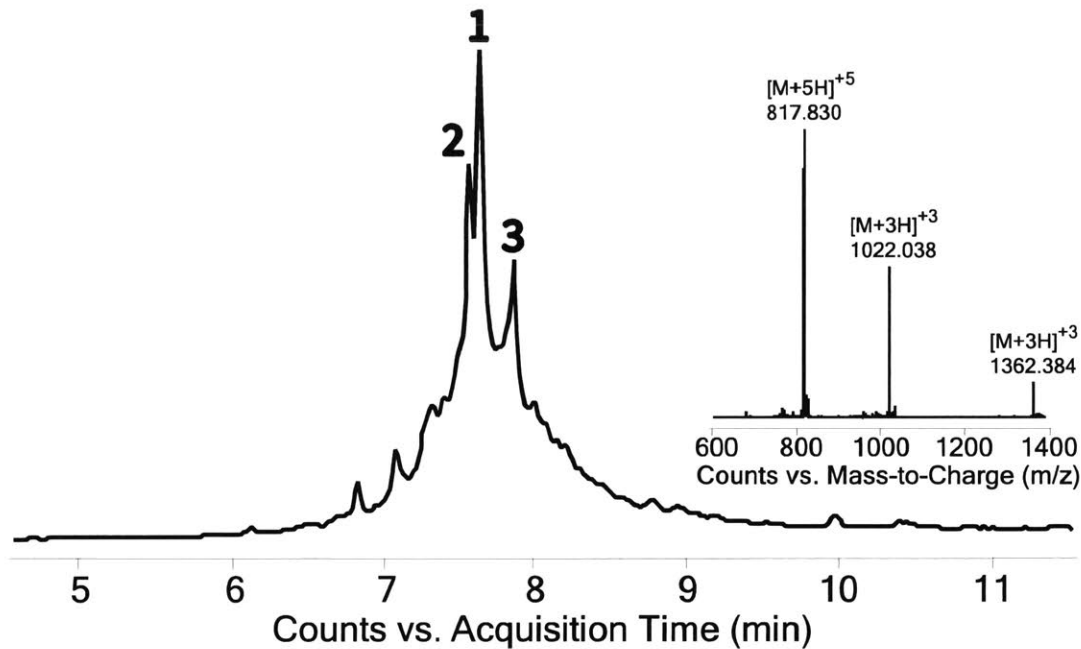
**Figure 1.12. LC-MS total ion current vs. time for the synthesis of B[1+2] on PEG resin using a standard 3-minute cycle (HATU, 50% (v/v) piperidine)**

C-terminal Val was coupled for 10 min. **1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4067.1 Da, obs. = 4067.2 Da). **2:** Co-eluting isomer of **1** and -18 Da relative to the desired product. **3:** Asn deletion relative to the desired product.

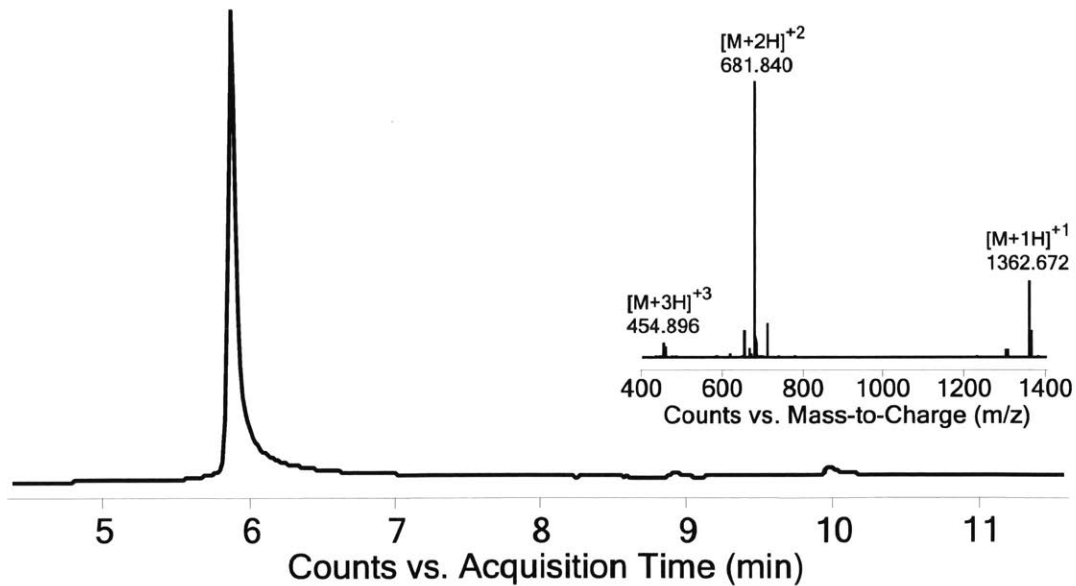


**Figure 1.13. LC-MS total ion current vs. time for the first twenty residues of B[1+2]**

**1:** Desired product with inset of charge state series of peak apex. **2:** Asn deletion. **3:** Co-eluting isomer of **1**, Ile deletion, and Tyr deletion **4:** His deletion



**Figure 1.14. LC-MS total ion current vs. time for the synthesis of B[1+2] (HATU, 50% (v/v) piperidine) using exhaustive deprotection on the 11<sup>th</sup>-13<sup>th</sup> residues coupled C-terminal Val was coupled for 10 min. 1: Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4082.1 Da, obs. = 4082.1 Da). 2: Co-eluting isomer of 1 and -18 Da relative to the desired product. 3: Asn deletion.**



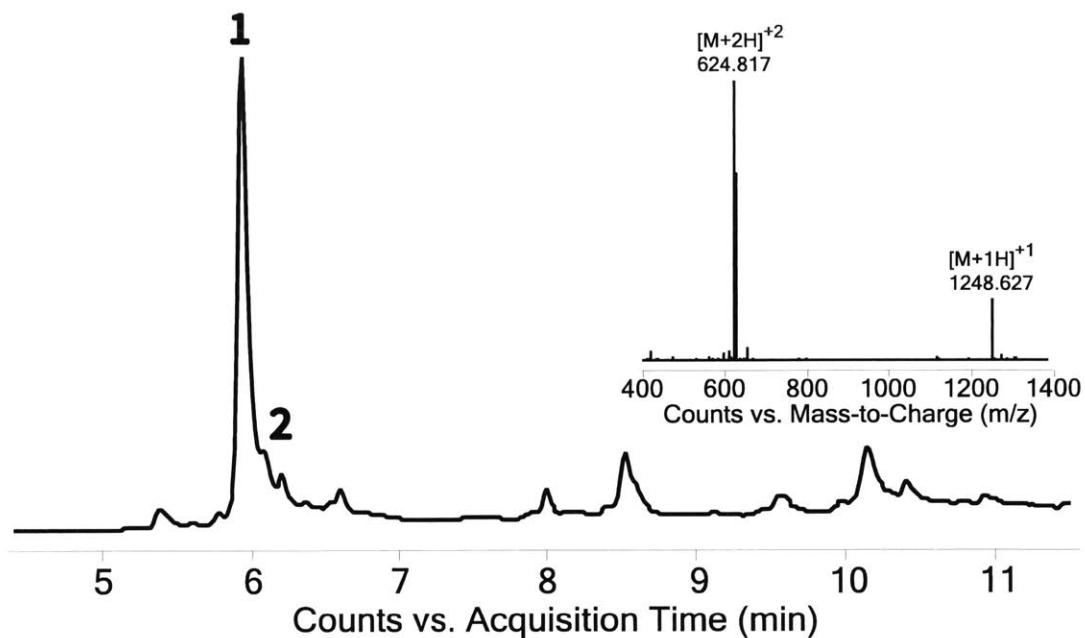
**Figure 1.15. LC-MS total ion current vs. time of purified G<sub>5</sub>-B[1]**  
 Inset of charge state series over the entire peak on the right (calc. monoisotopic mass = 1361.7 Da, obs. = 1361.7 Da) shows co-eluting Gly deletion and additions.

*Synthesis of G<sub>5</sub>-B[1]*. Because the synthesis B[1+2] proved to be persistently difficult, we divided this peptide into two separate fragments, barnase fragments B[1] and B[2]. Further, we appended pentaglycine to the N-terminus of B[1] to facilitate enzymatic ligation of barnase in future studies. This peptide was synthesized on 200 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine). The C-terminal Val was coupled to the resin for 10 min. Standard cleavage and lyophilization yielded 93 mg of crude material, which was purified over an Agilent Zorbax 300SB preparative C3 column (300Å, 7µm, 21.2 x 250mm). The method used for purification was: 20 mL/min flow rate, 5% B in A for 10 minutes, 5-35% B in A ramping linearly over 60 min (0.5% B/min), 70 % B in A for 5 min. Purification yielded 38 mg of pure B[1] (40%). Figure 1.15 shows the resulting pooled fractions with Gly deletion and addition impurities unresolved in the chromatography.

*Synthesis of G<sub>3</sub>-B[1]*. Unresolved Gly deletion and addition impurities in the synthesis G<sub>5</sub>-B[1] prompted the synthesis of B[1] with a shorter triglycine sequence appended to the N-terminus. G<sub>3</sub>-B[1] was synthesized on 190 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine). The C-terminal Val was coupled to the resin for 10 min. Standard cleavage and lyophilization yielded 183 mg of crude material (Figure 1.16), which was then purified in the same manner as G<sub>5</sub>-B[1]. Purification yielded 72 mg of pure G<sub>3</sub>-B[1] (39%). Figure 1.17 shows the results of purification.

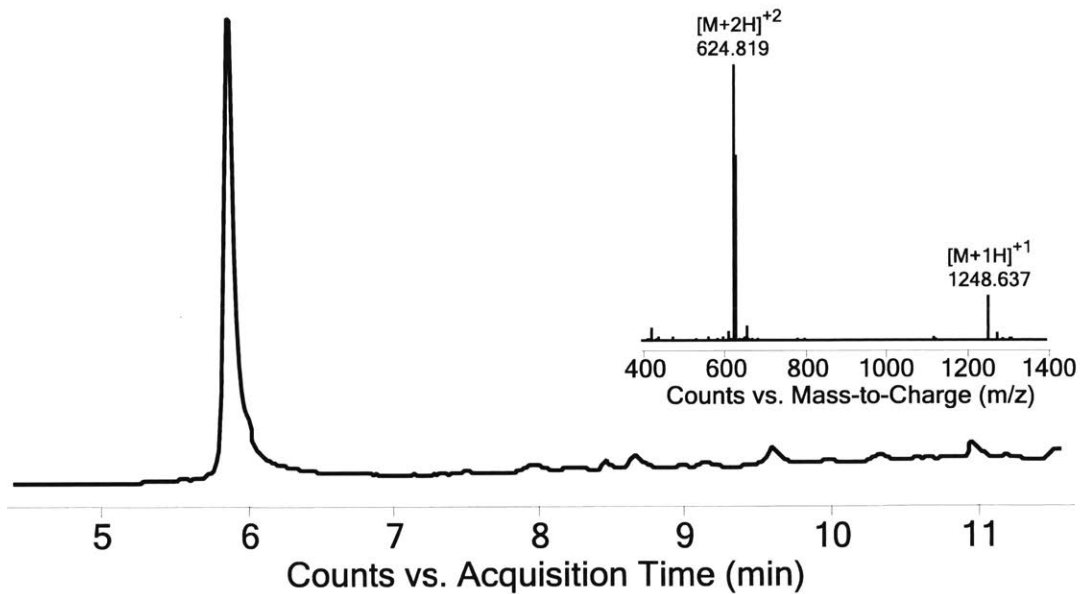
*Original synthesis of B[2]*. B[2] was originally synthesized on 193 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine deprotection) and 2.5 mL of 0.38 M HATU in DMF as an activating agent. Trp, Cys, and His, residues were activated with 190 µL of DIEA. The C-terminal Val was coupled to the resin for 10 minutes. Unsurprisingly, peptide quality was poor (Figure 1.18) likely owing to the stretches of peptide aggregation.

*Synthesis of B[2]: extended deprotection and coupling*. From the UV monitoring experiments on B[1+2], it was known that this peptide was prone to aggregation. To combat this problem, extended deprotection and coupling was employed in an attempt to reduce the severity of deletion products. B[2] was resynthesized on 160 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine deprotection) and 2.5 mL of 0.38 M HATU in DMF as an activating agent. Trp, Cys, and His were activated with 190 µL of DIEA and the C-terminal Val was coupled for 10 min. Deprotection times were prolonged to 60 sec on the 12<sup>th</sup> through 16<sup>th</sup> residues coupled (<sup>29</sup>Thr through<sup>25</sup>Asp). Additionally, the 12<sup>th</sup> through 14<sup>th</sup> residues coupled, <sup>29</sup>Thr through <sup>27</sup>Tyr, were coupled for 20 min, while the 15<sup>th</sup> residue, <sup>26</sup>Ile, was coupled for 10 min. The results of this experiment are shown in Figure 1.19. This strategy for combatting peptide aggregation during the synthesis of B[2] resulted in virtually no improvement in peptide quality.



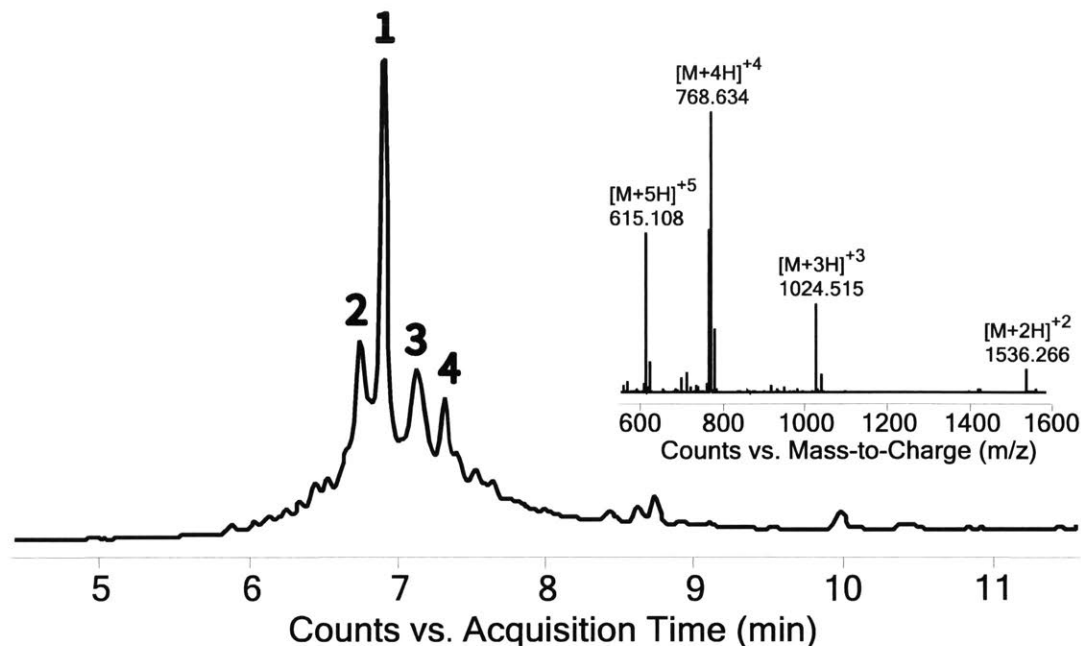
**Figure 1.16. LC-MS total ion current vs. time the synthesis G<sub>3</sub>-B[1] on a standard 3-minute cycle (HATU, 20% (v/v) piperidine)**

1: Desired product with inset of charge state series for peak apex (calc. monoisotopic mass = 1247.6 Da, obs. = 1247.6 Da). 2: Asn deletion.



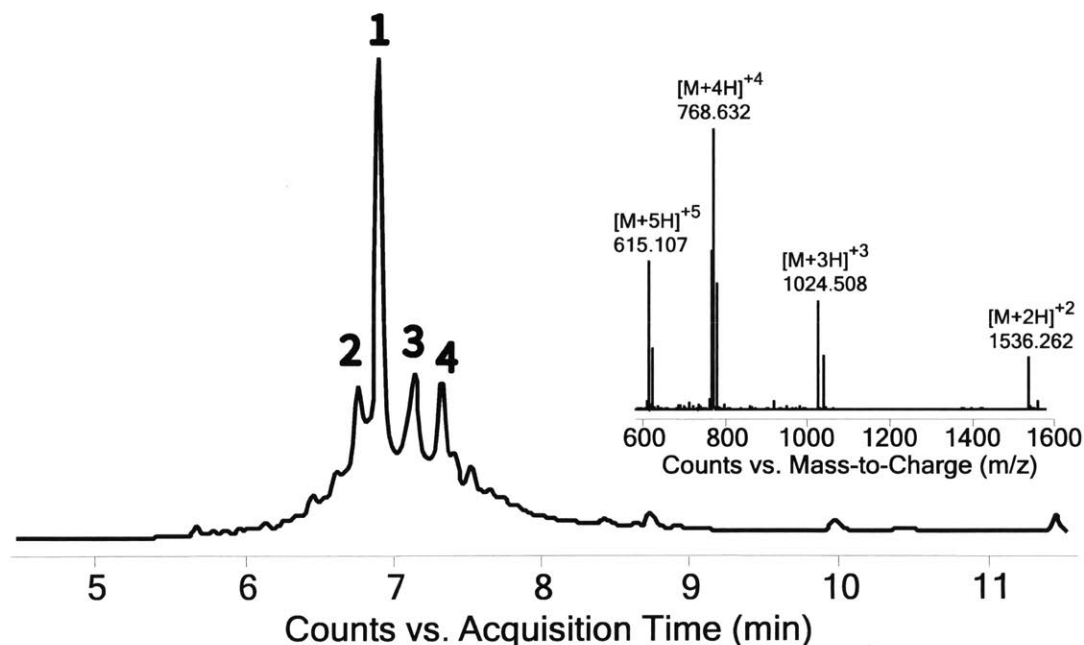
**Figure 1.17. LC-MS total ion current vs. time of purified G<sub>3</sub>-B[1]**

Inset of charge state series over the entire peak (calc. monoisotopic mass = 1247.6 Da, obs. = 1247.6 Da) is shown on the right.



**Figure 1.18. LC-MS total ion current vs. time for the original synthesis of B[2] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine)**

**1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 3069.5 Da, obs. = 3069.5 Da). Co-eluting +44 Da peptide. **2:** Co-elution of Tyr deletion and unidentified low molecular weight compound. **3:** Asn deletion. **4:** His deletion.



**Figure 1.19. LC-MS total ion current vs. time for the resynthesis of B[2] (HATU, 20% (v/v) piperidine) using extended deprotection and couplings**

**1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 3069.5 Da, obs. = 3069.5 Da). Co-eluting +44 Da peptide. **2:** Co-elution of Tyr deletion and unidentified low molecular weight compound. **3:** Asn deletion. **4:** His deletion.

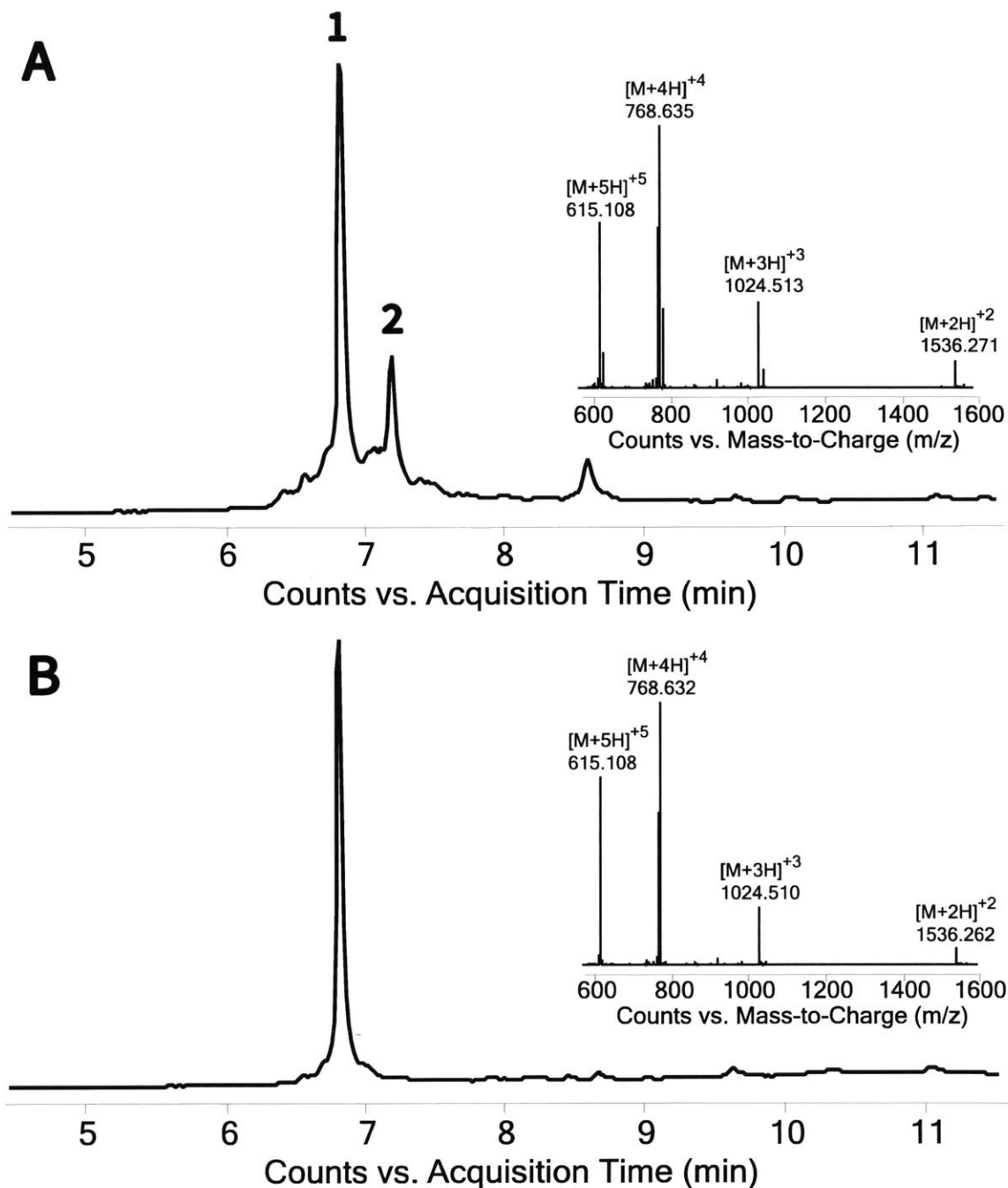


*Synthesis of B[2]: Hmb-backbone protection.* In another attempt to combat peptide aggregation on resin, a strategy of amide backbone protection via Fmoc-(Hmb)Ala-OH was employed. B[2] was assembled on 187 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine deprotection). Trp, His, and Cys were activated with 190  $\mu$ L of DIEA and the C-terminal Val was coupled for 10 min. The 7<sup>th</sup> residue coupled, <sup>33</sup>Ala, was incorporated as Fmoc-(Hmb)Ala-OH and coupled for 25 min. The next residue (Glu) was coupled for 30 minutes. 1 mmol of each of these amino acids was activated using a solution of 0.95 equivalent DCC and 1.00 equivalent HOBt in 3.0 mL DMF. The amino acid was dissolved in the activating agent solution within 60 min of use and the *N,N'*-dicyclohexylurea was removed by filtration through a 0.22- $\mu$ m filter immediately before the coupling step. Completion of the synthesis resulted in the peptide shown in Figure 1.20a. This experiment demonstrated the utility of Hmb amide backbone protection in the synthesis of a difficult sequence.

The resin was cleaved in a standard cleavage cocktail for 7 min at 60 °C. After work up of the peptide, lyophilization yielded 340 mg of crude material. 170 mg of the peptide were loaded onto an Agilent Zorbax 300SB preparative C3 column (300 $\text{\AA}$ , 7 $\mu$ m, 21.2x250mm) as a solution in 95% A / 5%B and purified at a flow rate of 20 ml/min with the following gradient: 5% B in A for 10 min, 5-18% B in A ramping linearly over 13 min, 18-40% B in A ramping linearly over 110 min, and 70% B in A for 5 min. The purification yielded 51 mg of pure compound (30%, Figure 1.20b).

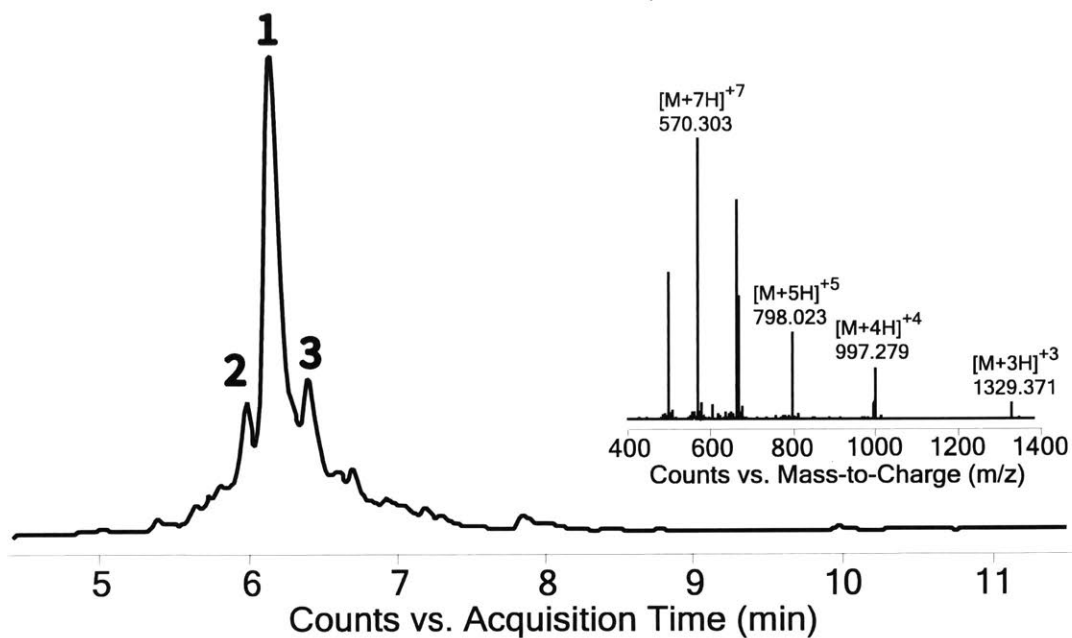
*Original synthesis of B[3], no Cys(Acm).* Before dividing B[1+2] into B[1] and B[2], barnase fragment B[3] was originally synthesized without orthogonal protection on Cys. Using a standard 3-minute cycle (HATU, 50% (v/v) piperidine), B[3] was synthesized on 190 mg of 2-chlorotrityl hydrazine resin. Trp and Cys residues were activated with 157  $\mu$ L of DIEA (Figure 1.21). The main impurity identified was +96 Da relative to the expected product. This was presumed to be trifluoroacetylation of the desired product. Purification of the peptide was not attempted because of the change in synthetic strategy, which required orthogonally protected Cys.

*Synthesis of B[3] with Cys(Acm) and improved conditions.* Once the synthetic approach to barnase was revised to include four fragments, B[3] was remade with Acm side chain protected cysteine. The peptide was synthesized on 190 mg of 2-chlorotrityl hydrazine resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine) with 2.5 mL of 0.38 M HATU in DMF as an activating agent. Trp and Cys were activated with 190 $\mu$ L of DIEA. Upon standard cleavage, work up, and lyophilization, 308 mg of crude peptide was obtained (Figure 1.22). 150 mg of the peptide were loaded onto an Agilent Zorbax 300SB preparative C3 column (300 $\text{\AA}$ , 7 $\mu$ m, 21.2 x 250mm) as a solution in 95% A / 5% B and purified at a flow rate of 20 mL/min using the following method: 5% B in A for 10 min, 5-18% B in A ramping linearly over 13 min, 18-40% B in A ramping linearly over 110 min, and 70%B in A for 5 min. 54 mg of a pure compound was isolated (36% purification yield, Figure 1.23).



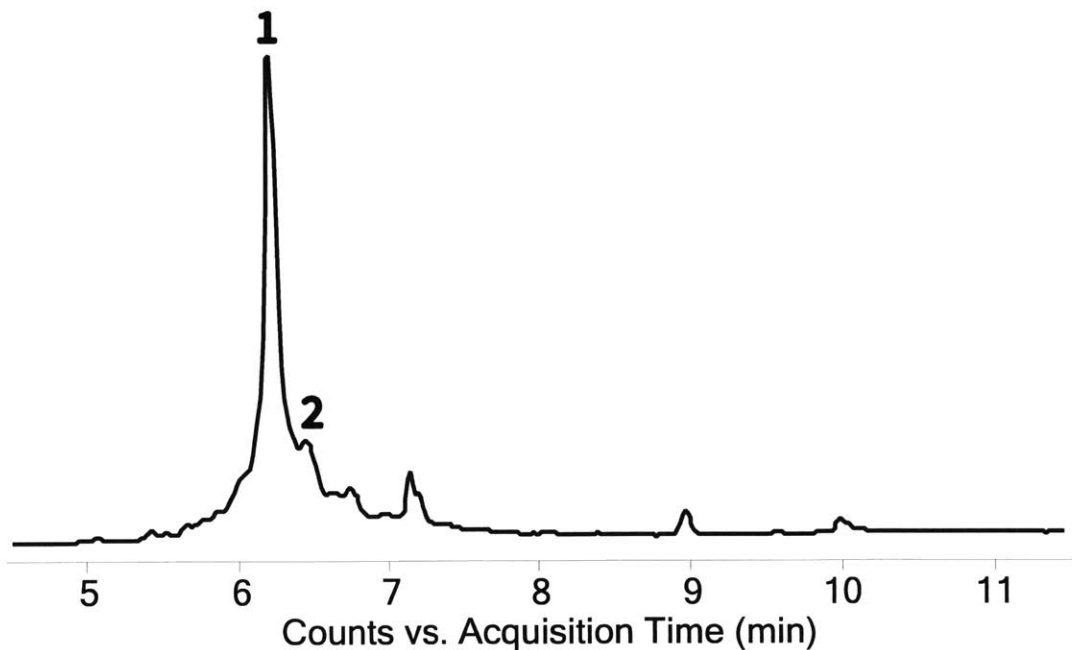
**Figure 1.20. LC-MS analysis of B[2]**

(A) LC-MS total ion current vs. time for the resynthesis of RNase *B. a.* fragment B[2] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine) and incorporating Fmoc-(Hmb)Ala-OH as described in the text. **1**: Desired product with inset charge state series of peak apex (calc. monoisotopic mass = 3069.5 Da, obs. = 3069.5 Da). Co-eluting +44 Da peptide. **2**: His deletion relative to desired product. (B) LC-MS total ion current vs. time of purified B[2]. Inset of charge state series over the entire peak (calc. monoisotopic mass = 3069.5 Da, obs. = 3069.5 Da).



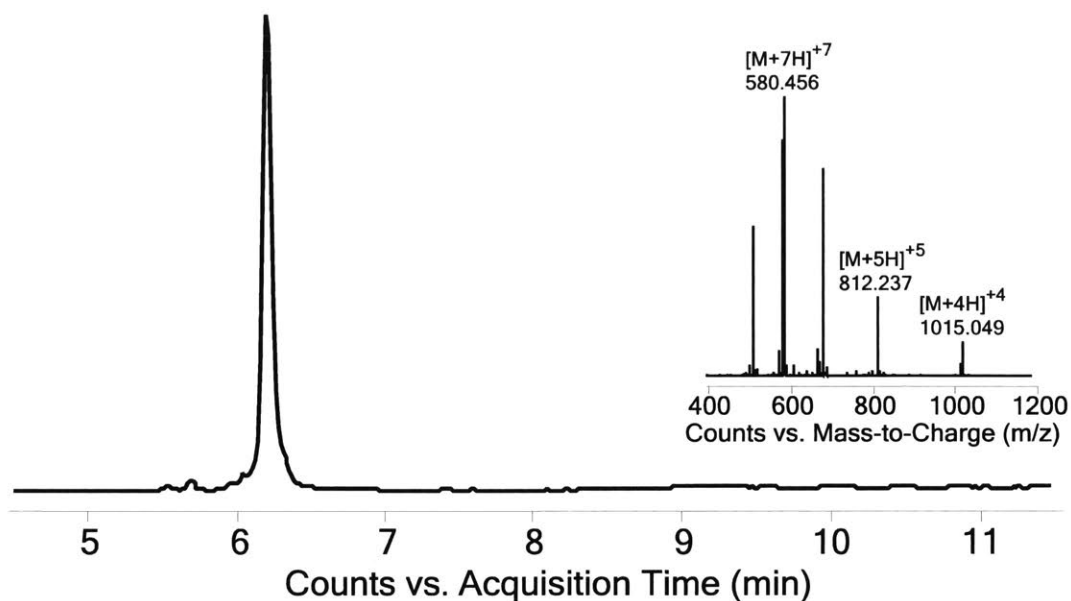
**Figure 1.21. LC-MS total ion current vs. time for the synthesis B[3] using a standard 3-minute cycle (HATU, 50% (v/v) piperidine)**

1: Desired product with inset charge state series of peak apex (calc. monoisotopic mass = 3983.1 Da, obs. = 3983.0 Da). 2: Isomer of desired product. 3: +96 Da relative to desired peptide.

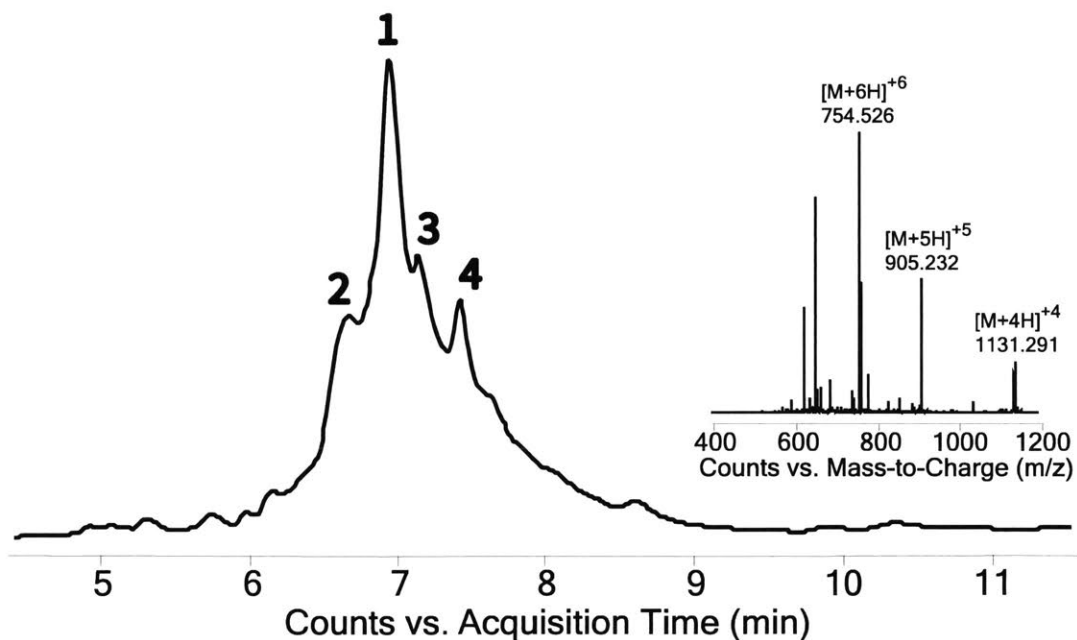


**Figure 1.22. LC-MS total ion current vs. time for the synthesis B[3] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine)**

1: Desired product with inset charge state series of peak apex (calc. monoisotopic mass = 4054.1 Da, obs. = 4054.2 Da). 2: +96 Da relative to desired peptide.



**Figure 1.23. LC-MS total ion current vs. time of purified B[3]**  
 Inset of charge state series over entire peak is shown on the right. Calc. monoisotopic mass = 4054.1 Da, obs. = 4054.2 Da).



**Figure 1.24. LC-MS total ion current vs. time for the synthesis of B[4] on AM resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine with 0.1M HOBt)**  
**1:** Desired product with inset charge state series of peak apex (calc. monoisotopic mass = 4519.2 Da, obs. = 4519.0 Da). **2:** Co-eluting isomer of 1 and Trp deletion. **3:** Arg deletion; **4:** Truncated peptide with the sequence H-ILYSSDWLIYKTTDHYQTFTKIR-NH<sub>2</sub> +42 Da.

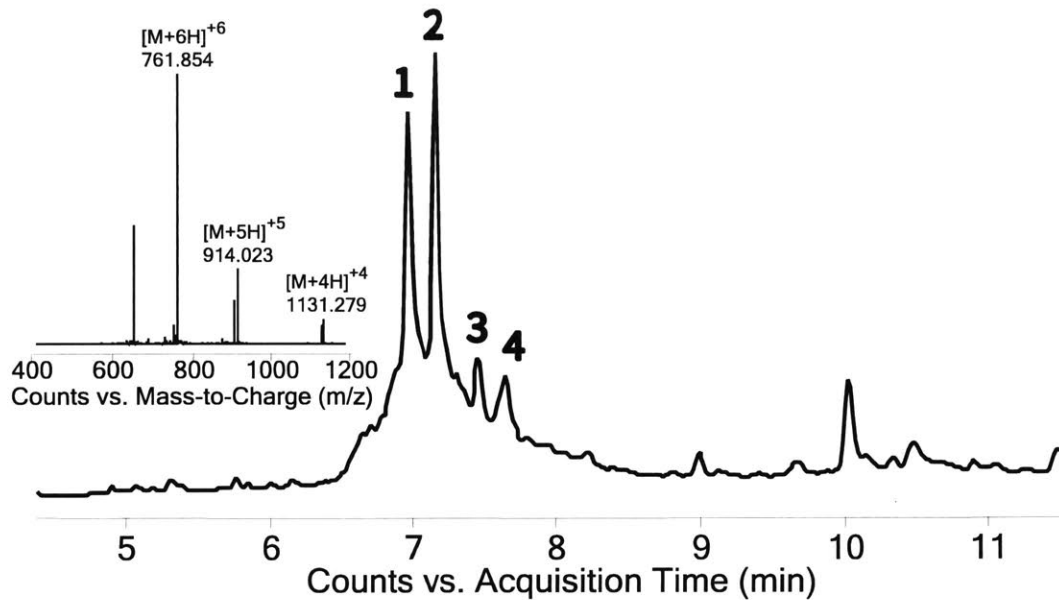
*Original synthesis of B[4].* B[4] was originally synthesized as a carboxamide on 180mg of aminomethyl resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine with 0.1M HOBt). Cys, His, and Trp were activated with 165 $\mu$ L of DIEA. 0.1M HOBt was added to the deprotection mixture in an attempt to suppress aspartimide formation. This quality of this peptide was too poor to attempt purification (Figure 1.24).

*Synthesis of B[4]: PEG resin.* In order to explore the influence of the solid support on the quality of this peptide, B[4] was resynthesized as a carboxamide on 100 mg of PEG resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine with 0.1M HOBt). Cys, His, and Trp were activated with 165 $\mu$ L of DIEA. For this peptide, switching from a polystyrene based resin to one that was PEG based dramatically improved the quality of the peptide (Figure 1.25). When synthesized on PEG resin, this peptide consistently appeared as mixture of desired product and +44 Da relative to desired product in an approximately 1:1 ratio. Assuming that +44 Da might be addition of carbon dioxide, the crude peptide was incubated in 50% A / 50% B for approximately 5 hours to reverse the reaction. Indeed, the +44Da product disappeared and the desired product grew in intensity (Figure 1.26). Purification of this peptide was not attempted.

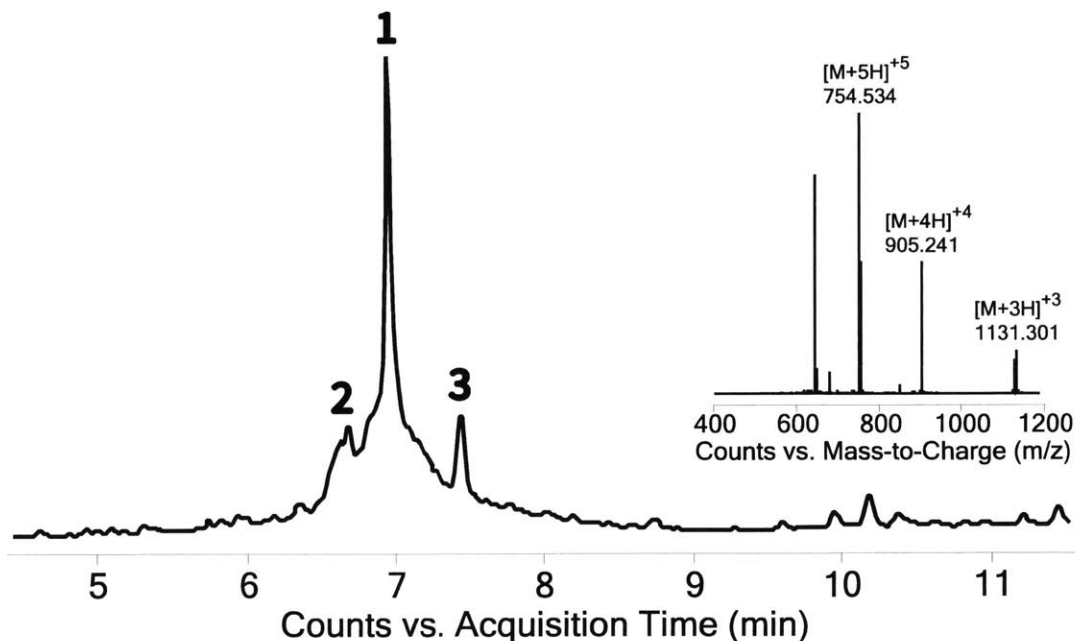
*Synthesis of B[4]: DCC/HOBt coupling of Arg.* In order to alleviate the formation of the truncated peptide observed in Figures 1.24-1.26, we synthesized B[4] using DCC/HOBt activation of the 24<sup>th</sup> residue coupled, <sup>90</sup>Arg. The peptide was synthesized as a carboxamide on 120 mg of PEG resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine). Cys, His, and Trp residues were activated with 165  $\mu$ L of DIEA. The 24<sup>th</sup> residue, <sup>90</sup>Arg, was coupled using a solution of 0.95 equivalents *N,N'*-dicyclohexylcarbodiimide and HOBt in 3.0 mL of DMF as an activation agent. The amino acid was dissolved in the activating agent within 60 min of use and filtered immediately before the coupling step. After a standard cleavage and workup, the crude peptide was dissolved in 50% A / 50% B and lyophilized 4 hours later to allow for the decarboxylation to occur. Lyophilization yielded 115 mg of crude material (Figure 1.27). DCC/HOBt activation of this particular Arg significantly reduced the intensity of the truncated sequence; in addition, removal of HOBt from the deprotection mixture had no effect on the amount of isomer observed. Purification of the fragment was not attempted.

*Synthesis of B[4] using improved conditions.* B[4] was resynthesized as a carboxamide on 182 mg of PEG resin using a standard 3-minute cycle (HATU, 20% piperidine), and 2.5 mL of 0.38 M solution of HATU in DMF as an activating agent throughout the synthesis. Cys, His, and Trp residues were activated with 190  $\mu$ L of DIEA. The 24<sup>th</sup> residue (<sup>90</sup>Arg) was coupled using a solution of 0.95 equivalents of DCC and 1.00 equivalent of HOBt in 3.0 mL of DMF as an activating agent. This amino acid was dissolved in the activating agent solution within 60 min of use and filtered immediately before the coupling step. After a standard cleavage, crude material was dissolved in 50% A / 50% B, analyzed on LC-MS and lyophilized 4 hours later. Lyophilization typically yielded 85–120 mg of crude material (Figure 1.28). 74 mg of the

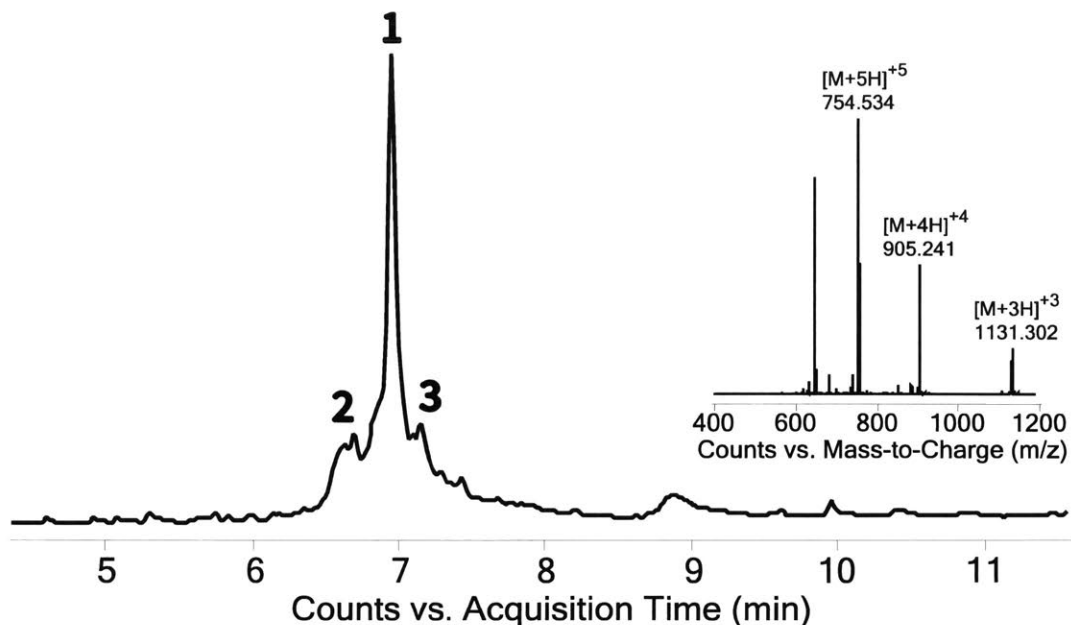
peptide were loaded onto an Agilent Zorbax 300SB preparative C3 column (300Å, 7µm, 21.2 x 250mm) as a solution in 85% A / 15% B and purified with a 20 mL/min flow rate of the following gradient: 15% B in A for 10 min, 15-20% B in A ramping linearly over 10 min, 20-25% B in A ramping linearly over 50 min, 25-35% B in A ramping linearly over 100 min, and 70% B in A for 5 minutes. The purification afforded 21 mg of pure B[4] (28% yield, Figure 1.29).



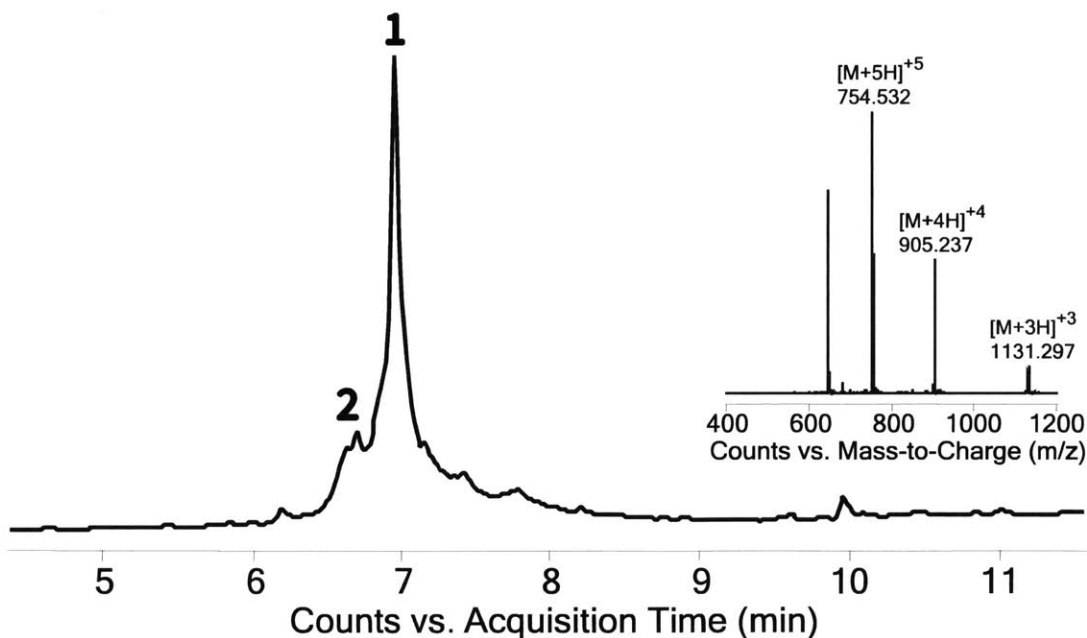
**Figure 1.25. LC-MS total ion current vs. time for the synthesis of B[4] on PEG resin using a standard 3-minute cycle (HATU, 20% (v/v) piperidine with 0.1M HOBt)**  
 1: Desired product. 2: +44 Da relative to the expected product with charge state series of peak apex on the left (calc. monoisotopic mass = 4519.2 + 44.0 Da, obs. = 4519.0 + 44.0 Da). 3: Truncated peptide with the sequence H-ILYSSDWLIYKTTDHYQTFTKIR-NH<sub>2</sub> +42 Da. 4: +44 Da relative to 3.



**Figure 1.26. B[4] after incubation in 50% A / 50% B for 5 hours**  
 1: Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4519.2 Da, obs. = 4519.0 Da). 2: An isomer of 1; 3: Truncated peptide with the sequence ILYSSDWLIYKTTDHYQTFTKIR-NH<sub>2</sub> +42 Da.

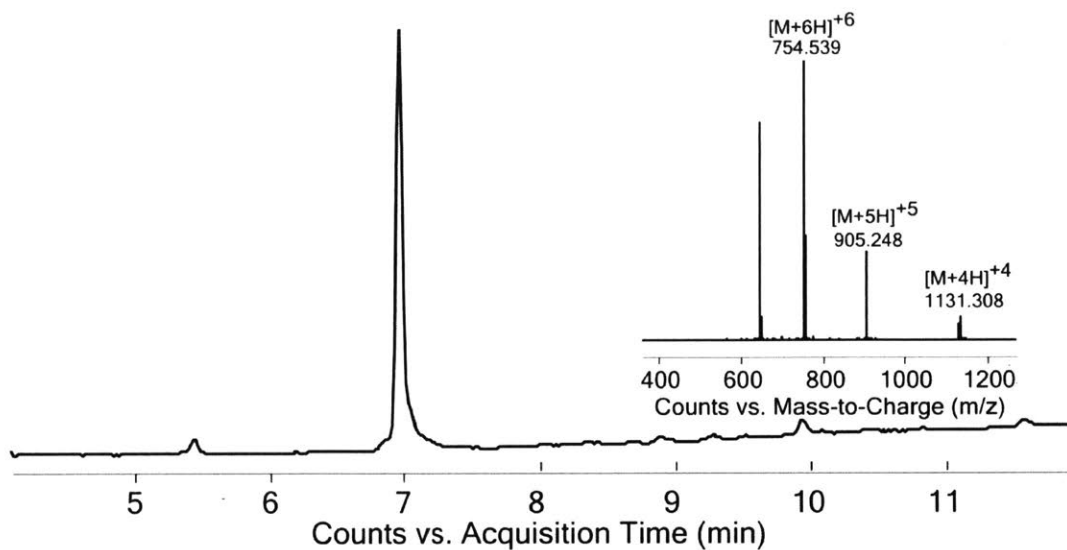


**Figure 1.27. LC-MS total ion current vs. time for the synthesis B[4] using PEG resin, a standard 3-minute cycle (HATU, 20% (v/v) piperidine), and DCC/HOBt activation of <sup>90</sup>Arg**  
**1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4519.2 Da, obs. = 4519.0 Da). **2:** +44 Da relative to desired product. **3:** His deletion



**Figure 1.28. LC-MS total ion current vs. time for the synthesis of B[4] using a standard 3-minute cycle (HATU, 20% (v/v) piperidine and slightly modified coupling solutions**  
**1:** Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 4519.2 Da, obs. = 4519.0 Da). **2:** +44 Da relative to desired product.





**Figure 1.29. LC-MS total ion current vs. time for purified B[4]**

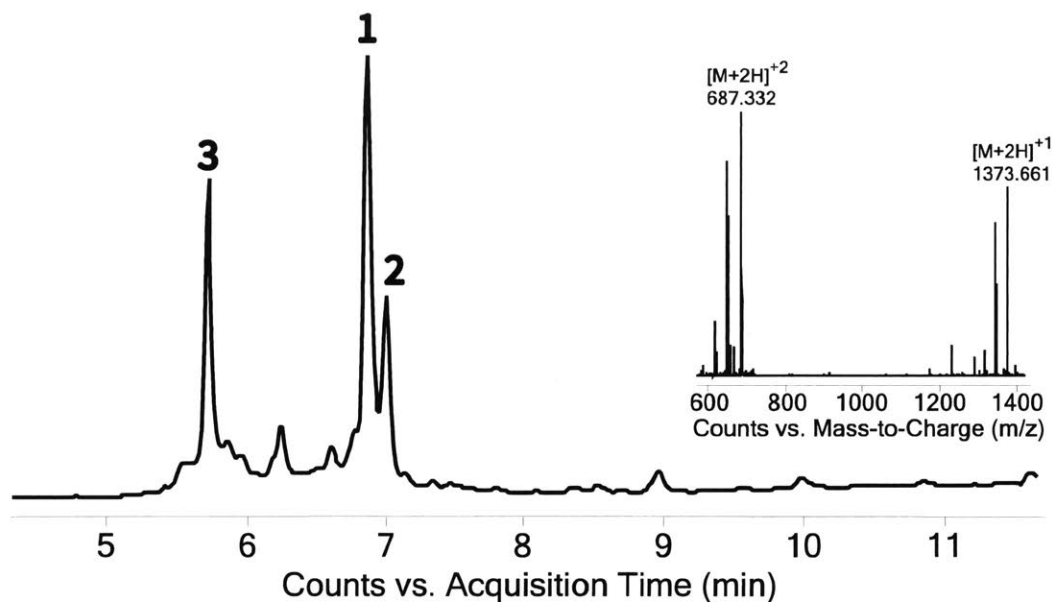
Inset of charge state series over entire peak is shown on the right (Calc. monoisotopic mass = 4519.2 Da, obs. = 4519.0 Da).

#### 1.4.6. Native Chemical Ligation of Barnase Peptides

*Challenges of oxidizing B[1].* Initially, there were problems oxidizing B[1] for subsequent thioesterification and ligation. In particular, the standard oxidation and thioesterification procedure described by Liu and co-workers<sup>36</sup> and outlined above yielded less than 15% of the expected product. Most of the following studies were performed with G<sub>5</sub>-B[1] but the G<sub>3</sub> analog behaved similarly. 2.00 mg (1.47 μmol) of G<sub>5</sub>-B[1] were dissolved in 0.48 mL of ligation buffer, and the pH of the resulting solution was adjusted to 3.0. The solution was incubated in an ice-salt bath at -18°C for 10 min and then 48 μL of oxidative solution was added to it dropwise while stirring. After 20 min, an analytical sample for LC-MS was taken (Figure 1.30) and then 0.48 mL of 0.2 M MPAA in ligation buffer at pH 6.8 were added to the reaction to the final pH of 4.5. Upon complete addition of the MPAA solution, another analytical sample for the LC-MS was taken (Figure 1.31).

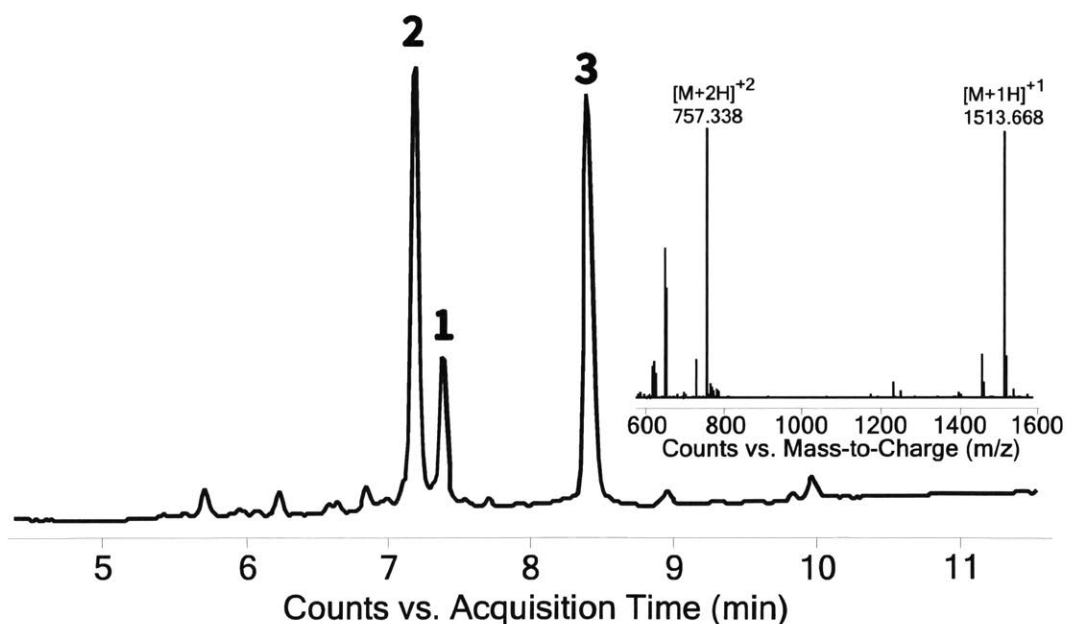
Figures 1.30 and 1.31 suggest that the azide that is formed during oxidation may lose a molecule of nitrogen (-28 Da). This intermediate may then add to MPAA resulting in +15 Da relative to expected product. This observation is consistent with the hypothesis that acyl azides undergo Curtius rearrangement to release nitrogen and form isocyanate, which can be trapped by excess nucleophile such as MPAA, eventually forming carbamothioate. This side reaction is explored in detail in chapters 3 and 4. In part by accident, it was found that carbamothioate formation could be circumvented if MPAA was allowed to react with the peptide-azide for approximately 20 min at room temperature without adjusting the pH to 7. These conditions yielded the desired MPAA thioester almost quantitatively (Figure 1.32). This adjustment was later implemented in synthesis of all other relevant thioesters.

*Synthesis of B[5].* Fragments G<sub>3</sub>-B[1] and B[2] were ligated using a one-pot oxidation-ligation procedure. 10.0 mg (8.01 μmol) fragment B[1] were dissolved in 2.61 mL of ligation buffer, and the pH of the resulting solution was adjusted to 3.0. The solution was incubated in an ice-salt bath at 0 °C for 10 min, and then 261 μL (52.2 μmol) of sodium nitrite in water was added to it dropwise. After 15 min, 2.61 mL of 0.2 M MPAA in ligation buffer at pH 6.8 were added to the reaction mixture, and the reaction was allowed to proceed at pH 4.5 for another 20 min at room temperature. Then, 25 mg (8.14 μmol) of B[2] was added and the pH of the solution was adjusted to 6.95. The reaction mixture was stirred for 19 hours and then quenched with 8 mL of 1 M dithiothreitol and 6 M Gn·HCl in water. The resulting mixture was allowed to stand for 20 min (Figure 1.33). The crude product was then loaded on the Agilent Zorbax 300SB preparative C3 column (300 Å, 7 μm, 21.2 x 250 mm) and purified with 20 mL/min of the following gradient: 15% B in A for 10 min, 15-25%B in A ramping linearly over 20 min, 25-45% B in A ramping linearly over 100 min, and 70% B in A for 5 min. Purification afforded 23.0 mg (67% yield) of pure B[5] (Figure 1.34).



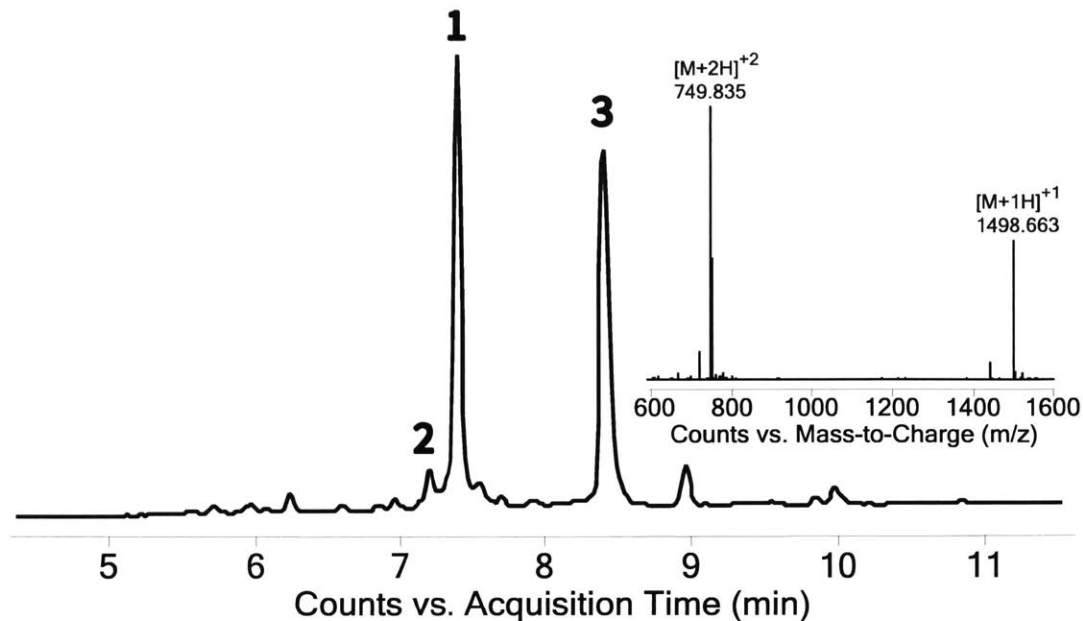
**Figure 1.30. LC-MS total ion current vs. time for G<sub>5</sub>-B[1] oxidation reaction**

1: Desired product with inset of charge state series of peak apex (calc. monoisotopic mass = 1372.7 Da, obs. = 1372.7 Da). Second major set of peaks is fragmentation to a 1345 Da ion. 2: -28 Da relative to desired azide. 3: Co-eluting -54 Da, -71 Da, and -142 Da relative to the desired azide.



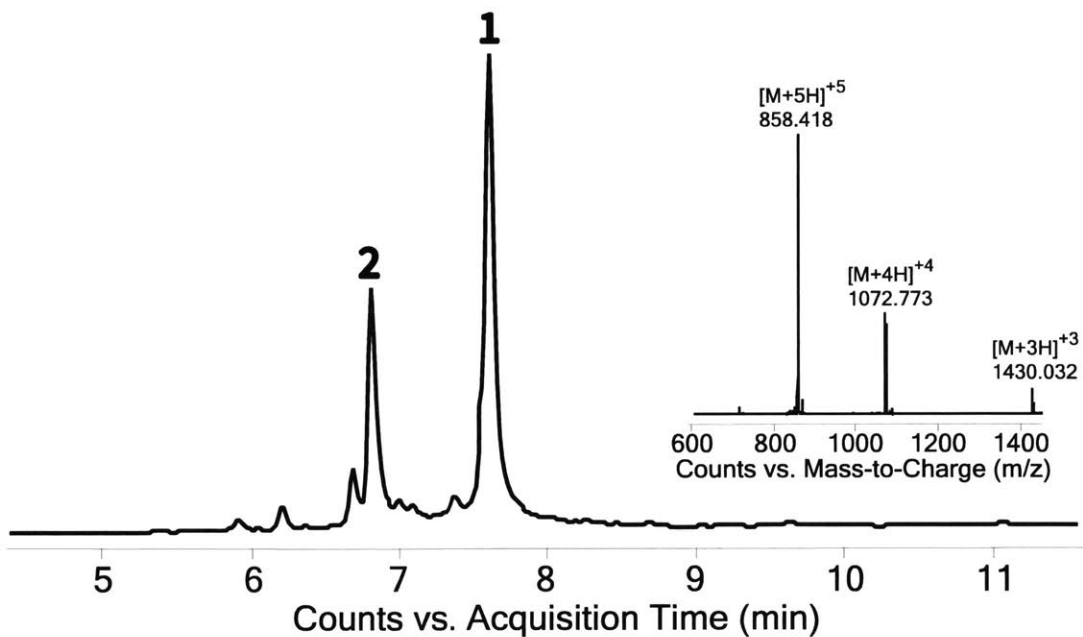
**Figure 1.31. LC-MS total ion current vs. time for G<sub>5</sub>-B[1] thioesterification reaction**

1: Desired product. 2: +15 Da relative to expected MPAA thioester with inset of charge state series of peak apex (obs. monoisotopic mass of product = 1512.7 Da). 3: MPAA oligomers.



**Figure 1.32. LC-MS total ion current vs. time for improved G<sub>5</sub>-B[1] MPAA formation reaction**

**1:** Desired product with inset of charge state series of peak apex (calc. average mass = 1497.7 Da, obs. = 1497.7 Da). **2:** Carbamothioate, +15 Da relative to desired peptide MPAA-thioester. **3:** MPAA oligomers.



**Figure 1.33. LC-MS total ion current vs. time for the ligation of B[1] to B[2] after 19 hours**

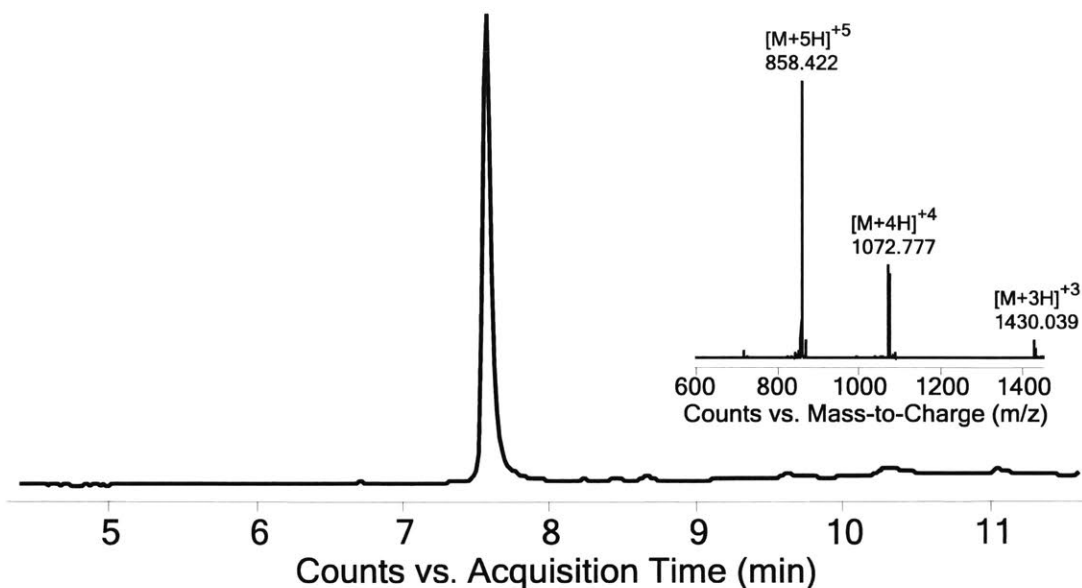
The ligation was quenched and reduced with DTT. **1:** Desired product with inset of charge state series of peak apex on the right (calc. average mass = 4287.8 Da, obs. = 4288.0 Da). **2:** Excess B[2].

*Synthesis of (Acm,Cy)-B[6].* B[3] and B[4] were ligated using a one-pot oxidation-ligation procedure. 13.3 mg (2.94  $\mu\text{mol}$ ) of B[3] were dissolved in 1.07 mL of ligation buffer and the pH of the resulting solution was adjusted to 3.0. The solution was incubated in an ice-salt bath at  $-18\text{ }^{\circ}\text{C}$  for 10 minutes, then 107  $\mu\text{L}$  (21.4  $\mu\text{mol}$ ) of sodium nitrite in water was added to it dropwise. After waiting 20 min, 1.07 mL of 0.2 M MPAA in ligation buffer at pH 6.8 was added to the reaction mixture and the reaction was allowed to proceed for another 20 minutes at room temperature. Then 11.8 mg (2.61  $\mu\text{mol}$ ) of B[4] were added and the pH of the solution was adjusted to 6.9. The reaction mixture was stirred for 10 hours, then it was quenched with 2 mL of 1 M DTT and 6 M  $\text{Gn}\cdot\text{HCl}$  in water. The resulting mixture was allowed to stand for 20 minutes (Figure 1.35). The ligated material was loaded on the semi-preparative Zorbax C3, 5  $\mu\text{m}$  column and purified with 5 mL/min of the following gradient: 5% B in A for 10 min, 5-20% B in A ramping linearly over 30 min, 20-40% B in A ramping linearly over 100 min, and 70% B in A for 5 minutes. Purification afforded 16.7 mg (77% yield) of pure (Acm,Cy)-B[6] (Figure 1.36).

*Synthesis of (Cy)-B[6].* Acm protection was removed from (Acm,Cy)-B[6] as follows: 10.7 mg (1.25  $\mu\text{mol}$ ) of (Acm,Cy)-B[6] was dissolved in 4.4 mL of 1:1 acetic acid/water solution. Silver (I) acetate (43.1 mg, 258  $\mu\text{mol}$ ) were then added to the solution and the resulting mixture was stirred gently at room temperature. After 105 min, the reaction was quenched with 9 mL 0.77 M DTT and 4.5 M  $\text{Gn}\cdot\text{HCl}$  in water. Yellowish precipitate formed upon DTT addition. The mixture was additionally stirred for 20 minutes, and then centrifuged at 4000 rpm for 5 minutes. The supernatant was desalted by solid-phase extraction on an Alltech C18 SPE cartridge and lyophilized (Figure 1.37) to yield 8.3 mg of (Cy)-B[6] (0.97  $\mu\text{mol}$ , 78% yield).

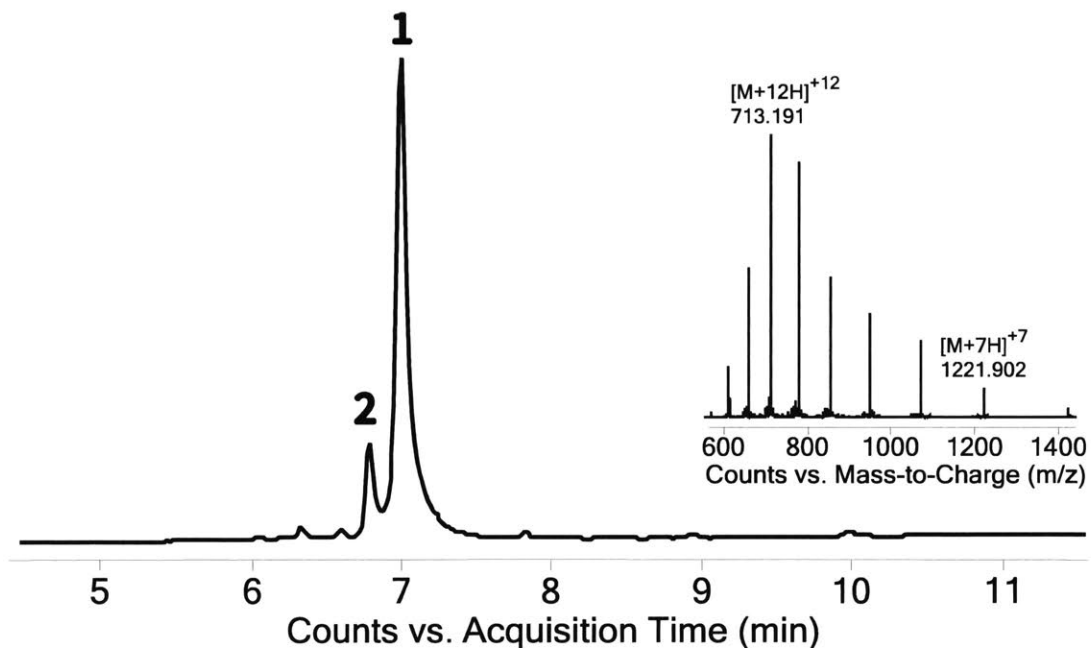
Before arriving at these conditions, several other experiments failed to yield the expected product. It was found that Acm could not be removed in 50% A / 50% B with either  $\text{AgOAc}$  or  $\text{AgOTf}$ . In contrast, 6 hours in 50% AcOH with  $\text{AgOAc}$  led to over-oxidation and significant degradation of the peptide. A 90-120 min reaction time ultimately proved to be quite general for Acm removal under these conditions, leading to a clean desired product.

*Synthesis of B[6].* To remove cyclohexyl protecting group the cleavage with anhydrous hydrofluoric acid in a way essentially identical to the classic procedure employed in the Boc SPPS strategy was performed. 10% (v/v) of *p*-Cresol was added to 8.2 mg (0.97  $\mu\text{mol}$ ) of (Cy)-B[6] and the resulting mixture was cooled in an ethanol-dry ice bath. Anhydrous hydrofluoric acid (HF) was added. The reaction was warmed to  $0\text{ }^{\circ}\text{C}$  and allowed to proceed for 60 min. Next, HF was evaporated under reduced pressure, and the residual solid was dissolved in 40 mL of 40% A / 60% B. After lyophilization, residual cresol was found together with peptide (Figure 1.38). This solid was re-dissolved in 85% A / 15% B, loaded on to a semi-preparative Zorbax C3 5 $\mu\text{m}$  column, and purified at flow rate of 5 mL/min with the following gradient: 15%B in A for 10 min, 15-25% B in A ramping linearly over 20 min, 25-45%B in A ramping linearly over



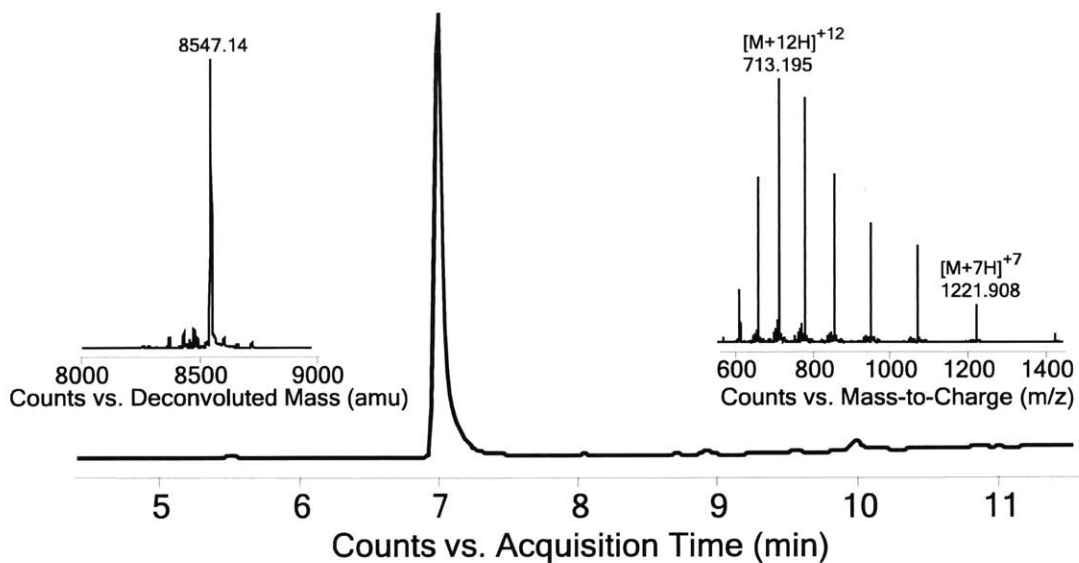
**Figure 1.34. LC-MS total ion current vs. time for purified fragment B[5]**

Inset of charge state series over entire peak is shown on the right (calc. average mass = 4288.2 Da, obs. = 4288.0 Da).



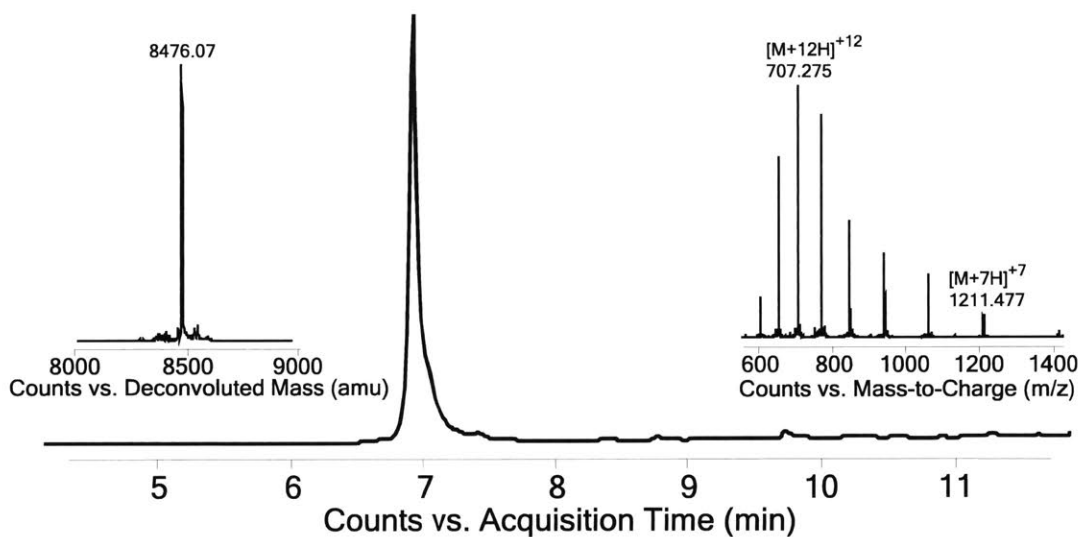
**Figure 1.35. LC-MS total ion current vs. time for the ligation of B[3] to B[4] after 10 hours**

The ligation was quenched and reduced with DTT. 1: Desired product with inset of charge state series of peak apex on the right (calc, average mass = 8546.6 Da, obs, = 8547.1 Da). 2: Excess of B[3] MPAA-thioester.



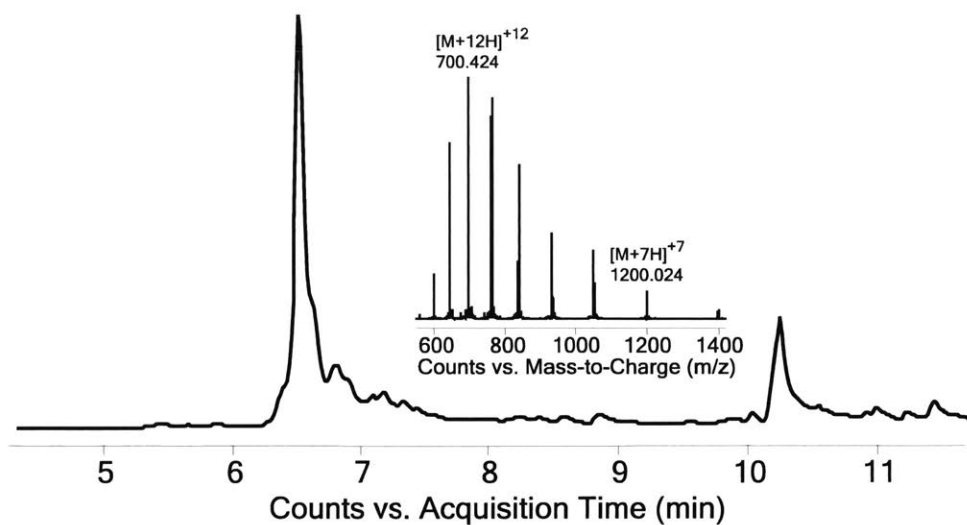
**Figure 1.36. LC-MS total ion current vs. time of purified (Acm,Cy)-B[6]**

Inset of charge state series and mass deconvolution over the entire peak is shown on the right (calc. average mass = 8546.6 Da, obs. = 8547.1 Da).



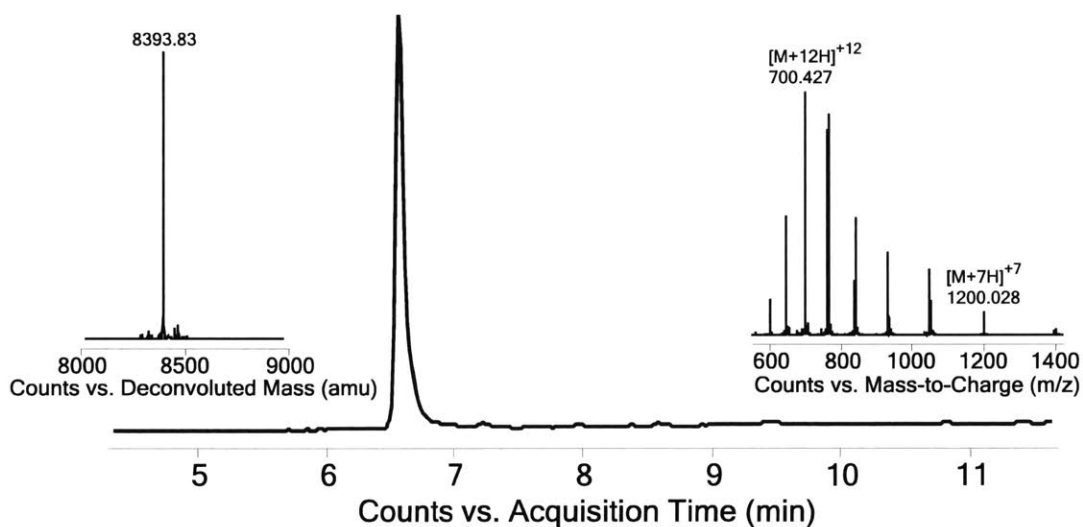
**Figure 1.37. LC-MS total ion current vs. time for Acm removal of (Acm,Cy)-B[6]**

Inset of charge state series and mass deconvolution over entire peak is shown on the right (calc. average mass = 8475.5 Da, obs. = 8475.0 Da).



**Figure 1.38. LC-MS total ion current vs. time of crude B[6] with residual *p*-cresol**

Inset of charge state series for peak apex is shown on the right (calc. average mass = 8393.3 Da, obs. = 8393.8 Da).



**Figure 1.39. LC-MS total ion current vs. time of purified B[6]**

Inset of charge state series for peak apex is shown on the right (calc. average mass = 8393.3 Da, obs. = 8393.7 Da).

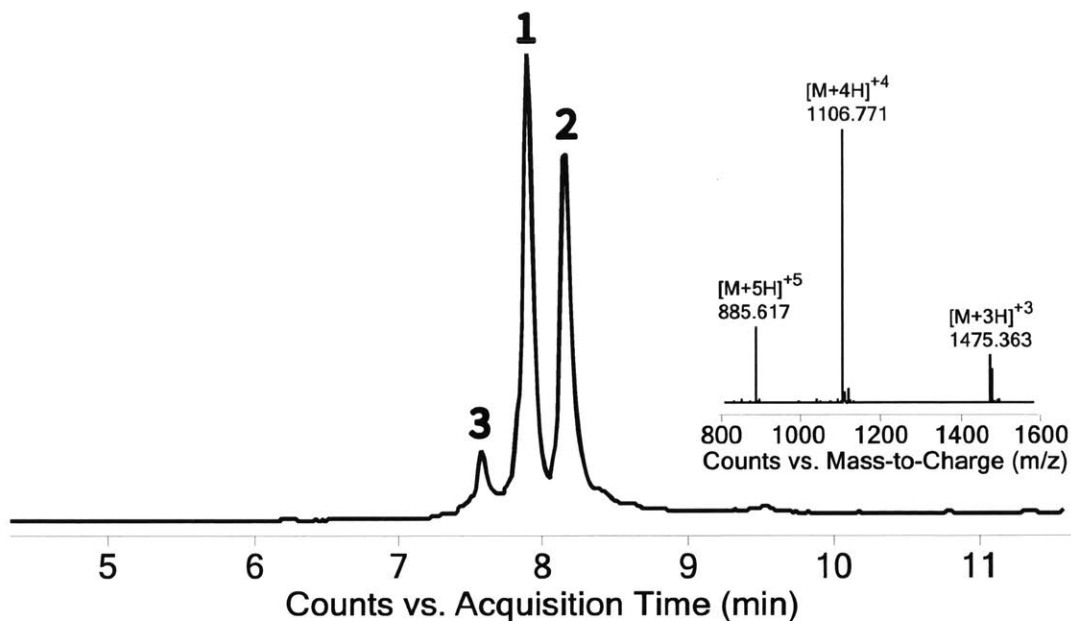


100 min, and 70% B in A for 5 min. Purification afforded 5.3 mg (0.63  $\mu\text{mol}$ , 65% yield) of B[6] (Figure 1.39).

Some efforts were made to cleave cyclohexyl ester of Glu without using HF, all of which failed. The following conditions were screened: (i) 95% methanesulfonic acid (MsOH) with 5% water, (ii) 95% MsOH with 1% water, 3% thioanisole and 1% EDT, (iii) Trifluoromethanesulfonic acid (TfOH) “high” or TfOH “low” cleavages.<sup>57,58</sup> All reactions were attempted at room temperature and at 0 °C, and monitored by LC-MS at 30 minute intervals. All reactions listed above led to full or partial decomposition of the material with less than 5% product observed.

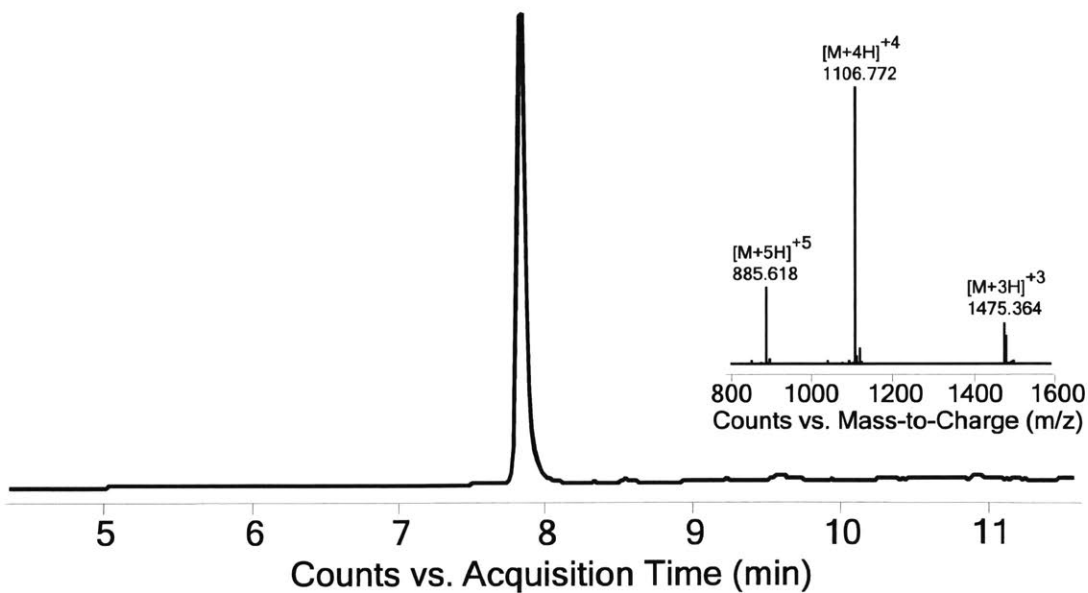
*Synthesis of MPAA thioester of B[5]*. Due to poor oxidation of B[5], the final ligation was performed in two steps. First, the MPAA thioester of B[5] was formed and isolated as follows: 8.7 mg (2.02  $\mu\text{mol}$ ) of B[5] were dissolved in 0.98 mL of ligation buffer and the pH of the resulting solution was adjusted to 3.0. The solution was incubated in a water-methanol-dry ice bath at  $-25$  °C for 10 min, then 98  $\mu\text{L}$  of oxidative solution were added dropwise. After 20 min, 0.98 mL of 0.2 M MPAA in ligation buffer at pH 6.8 were added to the reaction mixture, and reaction was allowed to proceed for another 25 min at room temperature (Figure 1.40). Then, 1.00 mL of 1M DTT and 6M Gn·HCl in water was added and the reaction mixture was allowed to stand for another 20 min. It was then loaded on a semi-preparative Zorbax C3, 5  $\mu\text{m}$  column and purified at a flow rate of 5 mL/min with the following gradient: 15% B in A for 10 min, 15-25% B in A ramping linearly over 20 min, 25-45% B in A ramping linearly over 100 min, and 70% B in A for 5 minutes. Purification afforded 5.9 mg (1.33  $\mu\text{mol}$ , 66% yield) of pure MPAA thioester B[5] (Figure 1.41).

*Synthesis of B[7]*. The MPAA thioester of fragment B[5] and fragment B[6] were ligated as follows: 0.62 mg (0.18  $\mu\text{mol}$ ) of the MPAA thioester of B[5] and 1.16 mg (0.14  $\mu\text{mol}$ ) of B[6] were dissolved in 0.10 mL of neutral, degassed ligation buffer. The reaction was allowed to proceed for 36 hours under oxygen free conditions at room temperature. Then, the reaction was quenched with 0.3mL of 1 M DTT and 6M Gn·HCl in water. The resulting mixture was allowed to stand for 20 min (Figure 1.42). The crude product was then loaded on an analytical Zorbax C3, 5 $\mu\text{m}$  column and purified with 1.5 mL/min of the following gradient: 15% B in A for 10 min, 15-25% B in A ramping linearly over 20 min, 25-45% B in A ramping linearly over 100 min, and 70% B in A for 5 min. Purification afforded the desired product contaminated with unreacted N-terminal fragment B[5]. From LC-MS analysis of the ligation reaction it was known that retention times of these peptides are significantly different when A'/B' solvent mixture is used. Therefore, the mixture was loaded on to the sample column and purified with 1.5 ml/min of the following gradient: A' with 15% B' for 5 min, 15-25% B' in A' ramping linearly over 20 min, 25-45% B' in A' ramping linearly over 60 min, and 70% B' in A' for 5 minutes. Purification afforded 0.60 mg (34% yield) of pure B[7] (Figure 1.43).



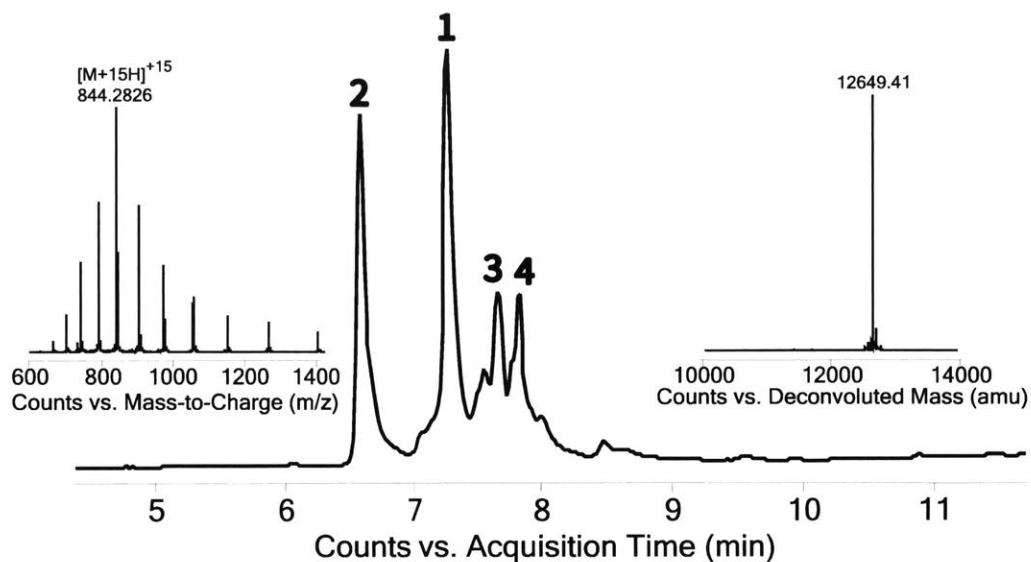
**Figure 1.40. LC-MS total ion current vs. time for the oxidation and thioesterification of B[5]**

**1:** Desired product with inset of charge state series of peak apex (calc. average mass = 4423.9 Da, obs. = 4424.3 Da) **2:** Co-eluting MPAA oligomers and mixed disulfide of MPAA and B[5]. **3:** Co-eluting, C-terminal carboxylate and amide of B[5] side products.



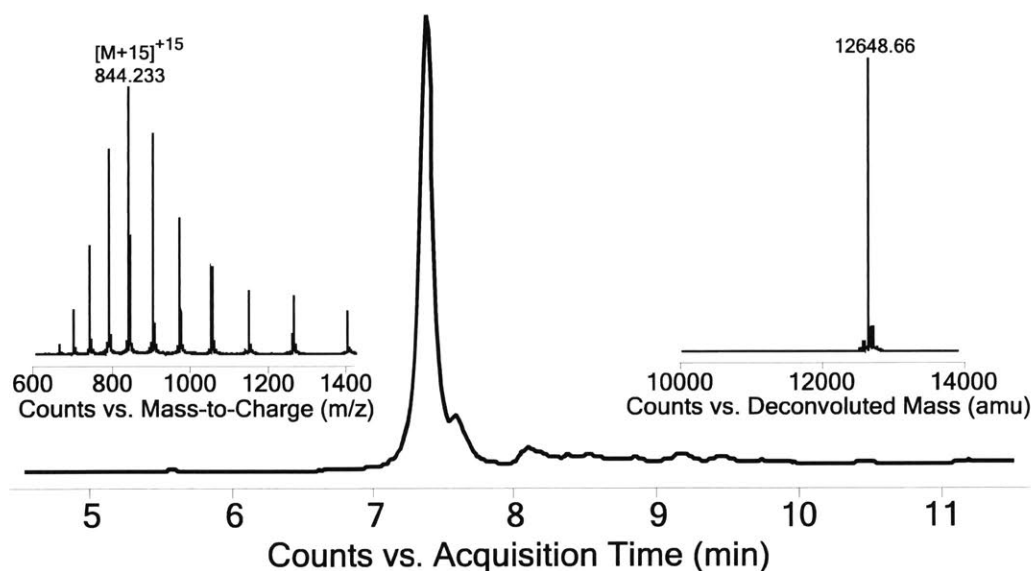
**Figure 1.41. LC-MS total ion current vs. time of purified MPAA thioester of B[5]**

Inset of charge state series over the entire peak is shown on the right (calc. average mass = 4423.9 Da, obs. = 4424.3 Da).



**Figure 1.42. LC-MS total ion current vs. time for the ligation of B[5]-MPAA thioester to B[6] after 36 hours**

The ligation was quenched and reduced with DTT. **1:** Desired product with inset of charge state series and mass deconvolution of peak apex (calc. average mass = 12649.1 Da, obs. = 12649.4 Da). **2:** Unreacted B[6]. **3:** Cyclic thioester of B[5]. **4:** Unreacted MPAA-thioester of B[5].

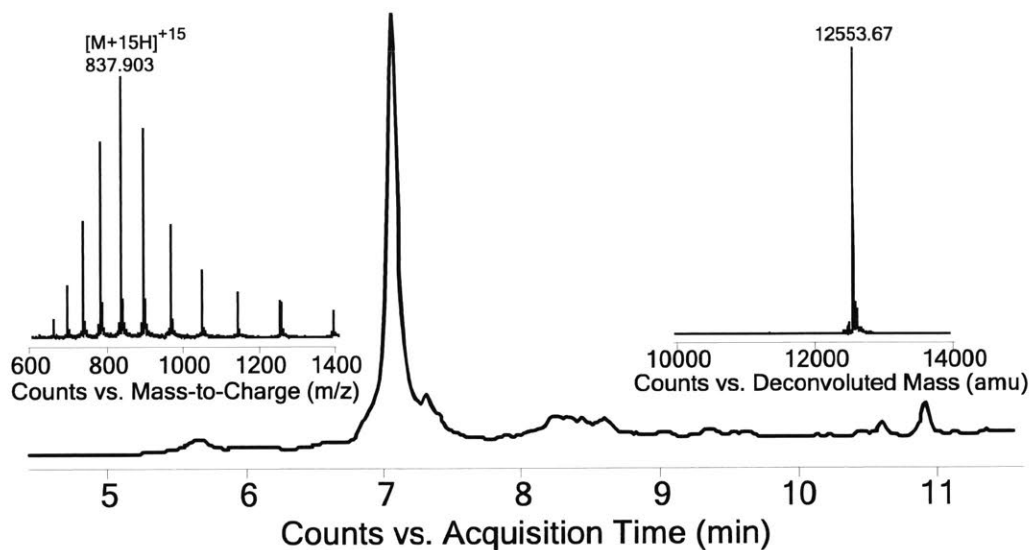


**Figure 1.43. LC-MS total ion current vs. time of purified B[7]**

Inset of charge state series of peak apex and mass deconvolution of entire peak is shown on the right (calc. monoisotopic mass = 12649.1 Da, obs. = 12648.7 Da).

*Desulfurization and synthesis of RNase B. a.* Desulfurization of B[7] to afford RNase *B. a.* was achieved as follows: 0.60 mg (0.47  $\mu\text{mol}$ ) of B[7], the full length protein, were dissolved in 190  $\mu\text{L}$  of neutral ligation buffer (0.2 M  $\text{Na}_2\text{HPO}_4$ , 6 M  $\text{Gn}\cdot\text{HCl}$ ). After 25.0 mg of TCEP $\cdot\text{HCl}$ , 9.5 mg of MesNa, and 0.3 mg of radical initiator VA-044 (2,2'-Azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride) were added to the reaction mixture, the pH of the resulting solution was adjusted to 7.0 and the reaction was allowed to proceed at 37  $^\circ\text{C}$ . The pH was monitored and maintained at 7.0 for 8 hours. The reaction mixture was then loaded on an SPE cartridge and eluted from it using 50% A / 50% B. Lyophilization afforded 0.55 mg (91% yield) of full length barnase (Figure 1.44).

*Folding.* Full length, desulfurized barnase was first dissolved in a minimal amount of 6 M  $\text{Gn}\cdot\text{HCl}$ . This solution was then serially diluted in two or three steps to 0.5 M or less  $\text{Gn}\cdot\text{HCl}$  using the desired buffer. In order to remove remnant guanidinium, this solution was passed over a 5mL, GE Healthcare HiTrap Desalting column and eluted with 5.0 mL/min of the desired buffer. Fractions containing the protein were flash frozen using liquid nitrogen and stored at -80  $^\circ\text{C}$ .



**Figure 1.44. LC-MS total ion current vs. time for isolated, desulfurized full-length barnase**  
 Inset of charge state series of the peak apex and mass deconvolution of the entire peak are also displayed  
 (calc. average mass = 12552.9 Da, obs. = 12553.7 Da).

## 1.5. Acknowledgements

This research was supported by the MIT startup funds for B.L.P., MIT Reed Fund, MIT Deshpande Center, Damon Runyon-Rachleff Innovation Award, Sontag Foundation Distinguished Scientist Award, AstraZeneca Distinguished Graduate Student Fellowship for A.A.V., and Daniel S. Kemp Summer Fellowship and NIGMS/MIT Biotechnology Training Program (5T32GM008334-25) for S.K.M. We thank R. John Collier for providing some of the laboratory equipment used to carry out the experiments. We also thank Dr. Xiaoli Liao, Dr. Alexander Spokoyny, Dr. Amy Rabideau, Richard Chang, and Tatiana Berger for expert technical assistance. We acknowledge the Biological Instrument Facility of MIT for providing access to CD spectrometry (NSF-0070319).

## 1.6. References

- (1) Kochendoerfer, G.; Chen, S.-Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C.; Low, D.; Cagle, E.; Carnevali, M.; Gueriguian, V.; Keogh, P.; Porter, H.; Stratton, S.; Wiedeke, M.; Wilken, J.; Tang, J.; Levy, J.; Miranda, L.; Crnogorac, M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J.; Kung, A.; Kent, S.; Bradburne, J. *Science*. **2003**, *299*, 884–887.
- (2) Chalker, J.; Lercher, L.; Rose, N.; Schofield, C.; Davis, B. *Angew. Chem., Int. Ed.* **2012**, *51*, 1835–1839.
- (3) Roberts, M.; Bentley, M.; Harris, J. *Adv. Drug Deliv. Rev.* **2012**, *64*, 116–127.
- (4) Wu, A.; Senter, P. *Nat. Biotechnol.* **2005**, *23*, 1137–1146.
- (5) Alley, S.; Okeley, N.; Senter, P. *Curr. Opin. Chem. Biol.* **2010**, *14*, 529–537.
- (6) Prescher, J.; Bertozzi, C. *Nat. Chem. Biol.* **2005**, *1*, 13–21.
- (7) Macmillan, D.; Bill, R.; Sage, K.; Fern, D.; Flitsch, S. *Chem. Biol.* **2001**, *8*, 131–145.
- (8) Hackenberger, C.; Friel, C.; Radford, S.; Imperiali, B. *J. Am. Chem. Soc.* **2005**, *127*, 12882–12889.
- (9) Armishaw, C.; Daly, N.; Nevin, S.; Adams, D.; Craik, D.; Alewood, P. *J. Biol. Chem.* **2006**, *281*, 14136–14143.
- (10) Kumar, K.; Bavikar, S.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A. *Angew. Chem., Int. Ed.* **2011**, *50*, 6137–6141.
- (11) Kotch, F.; Raines, R. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 3028–3033.
- (12) Goto, Y.; Katoh, T.; Suga, H. *Nat. Protoc.* **2011**, *6*, 779–790.
- (13) Xie, J.; Schultz, P. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548–554.
- (14) Chin, J.; Cropp, T.; Anderson, J.; Mukherji, M.; Zhang, Z.; Schultz, P. *Science*. **2003**, *301*, 964–967.
- (15) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science*. **1994**, *266*, 776–778.
- (16) Dawson, P.; Kent, S. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
- (17) Ollivier, N.; Behr, J.-B.; El-Mahdi, O.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2005**, *7*, 2647–2650.
- (18) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 8521–8524.
- (19) Crich, D.; Banerjee, A. *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.
- (20) Haase, C.; Rohde, H.; Seitz, O. *Angew. Chem., Int. Ed.* **2008**, *47*, 6807–6810.
- (21) Yan, L.; Dawson, P. *J. Am. Chem. Soc.* **2001**, *123*, 526–533.
- (22) Wan, Q.; Danishefsky, S. *Angew. Chem., Int. Ed. Engl.* **2007**, *46*, 9248–9252.
- (23) Pentelute, B.; Kent, S. *Org. Lett.* **2007**, *9*, 687–690.
- (24) Bang, D.; Kent, S. *Angew. Chem., Int. Ed.* **2004**, *43*, 2534–2538.

- (25) Bang, D.; Pentelute, B.; Kent, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 3985–3988.
- (26) Okamoto, R.; Izumi, M.; Kajihara, Y. *Int. J. Pep. Res. Ther.* **2010**, *16*, 191–198.
- (27) Mende, F.; Seitz, O. *Angew. Chem., Int. Ed.* **2007**, *46*, 4577–4580.
- (28) Li, X.; Kawakami, T.; Aimoto, S. *Tetrahedron Lett.* **1998**, *39*, 8669–8672.
- (29) Muir, T.; Sondhi, D.; Cole, P. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6705–6710.
- (30) Coin, I.; Beyermann, M.; Bienert, M. *Nat. Protoc.* **2007**, *2*, 3247–3256.
- (31) Hartley, R.; Barker, E. *Nat. New Biol.* **1972**, *235*, 15–16.
- (32) Dawson, P.; Churchill, M.; Ghadiri, M.; Kent, S. *J. Am. Chem. Soc.* **1997**, *119*, 4325–4329.
- (33) Mossakowska, D.; Nyberg, K.; Fersht, A. *Biochemistry* **1989**, *28*, 3843–3850.
- (34) Hartley, R. *Trends Biochem. Sci.* **1989**, *14*, 450–454.
- (35) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, 3267–3272.
- (36) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chem., Int. Ed.* **2011**, *50*, 7645–7649.
- (37) Johnson, E.; Kent, S. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.
- (38) Liu, S.; Pentelute, B.; Kent, S. *Angew. Chem., Int. Ed.* **2012**, *51*, 993–999.
- (39) Kunitz, M. *J. Biol. Chem.* **1946**, *164*, 563–569.
- (40) Hyde, C.; Johnson, T.; Owen, D.; Quibell, M.; Sheppard, R. *Int. J. Pept. Protein Res.* **1994**, *43*, 431–440.
- (41) Garcia-Martin, F.; Quintanar-Audelo, M.; Garcia-Ramos, Y.; Cruz, L.; Gravel, C.; Furic, R.; Cote, S.; Tulla-Puche, J.; Albericio, F. *J. Comb. Chem.* **2006**, *8*, 213–220.
- (42) Johnson, T.; Quibell, M.; Owen, D.; Sheppard, R. *J. Chem. Soc. Chem. Commun.* **1993**, 369–372.
- (43) Nicolris, E.; Pujades, M.; Bacardit, J.; Giralt, E.; Albericio, F. *Tetrahedron Lett.* **1997**, *38*, 2317–2320.
- (44) Mergler, M.; Dick, F.; Sax, B.; Weiler, P.; Vorherr, T. *J. Pept. Sci.* **2003**, *46*, 36–46.
- (45) Mergler, M.; Dick, F. *J. Pept. Sci.* **2005**, *11*, 650–657.
- (46) Palasek, S.; Cox, Z.; Collins, J. *J. Pept. Sci.* **2007**, *13*, 143–148.
- (47) Dölling, R.; Beyermann, M.; Haenel, J.; Kernchen, F.; Peter, K.; Brudelb, M.; Bienerta, M. *J. Chem. Soc. Chem. Commun.* **1994**, 853–854.
- (48) Chang, C.-D.; Waki, M.; Ahmad, M.; Meienhofer, J.; Lundell, E.; Haug, J. *Int. J. Pept. Protein Res.* **1980**, *15*, 59–66.
- (49) Franzen, H.; Grehn, L.; Ragnarsson, U. *J. Chem. Soc. Chem. Commun.* **1984**, 1699–1700.
- (50) Haase, C.; Burton, M.; Agten, S.; Brunsveld, L. *Tetrahedron Lett.* **2012**, *53*, 4763–4765.
- (51) Prelog, V.; Wieland, P. *Helv. Chim. Acta* **1946**, *29*, 1128–1132.
- (52) Hofmann, K.; Thompson, T.; Yajima, H.; Schwartz, E.; Inouye, H. *J. Am. Chem. Soc.* **1960**, *82*, 3715–3721.
- (53) Honzl, J.; Rudinger, J. *Collect. Czech. Chem. Commun.* **1961**, *26*, 2333–2344.
- (54) Schwyzer, R.; Kappeler, H. *Helv. Chim. Acta* **1961**, *44*, 1991–2002.
- (55) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. a; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoiny, A.; Jensen, K.; Pentelute, B. *Chembiochem* **2014**, *15*, 713–720.
- (56) Stavropoulos, G.; Gatos, D.; Magafa, V.; Barlos, K. *Lett. Pept. Sci.* **1995**, *2*, 315–318.
- (57) Tam, J.; Heath, W.; Merrifield, R. *J. Am. Chem. Soc.* **1986**, *108*, 5242–5251.
- (58) Nomizu, M.; Inagaki, Y.; Yamashita, T.; Ohkubo, A.; Otaka, A.; Fujii, N.; Roller, P.; Yajima, H. *Int. J. Pept. Protein Res.* **1991**, *37*, 145–152.





## Chapter 2. Total Synthesis and Biochemical Characterization of Mirror Image Barnase

The work presented in this chapter was published in the following manuscript and is open access from Royal Society of Chemistry:

Vinogradov, A.; Evans, E; Pentelute, B. Total Synthesis and Biochemical Characterization of Mirror Image Barnase. *Chem. Sci.* **2015**, *6*, 2997–3002. DOI: 10.1039/C4SC03877K

## 2.1. Introduction

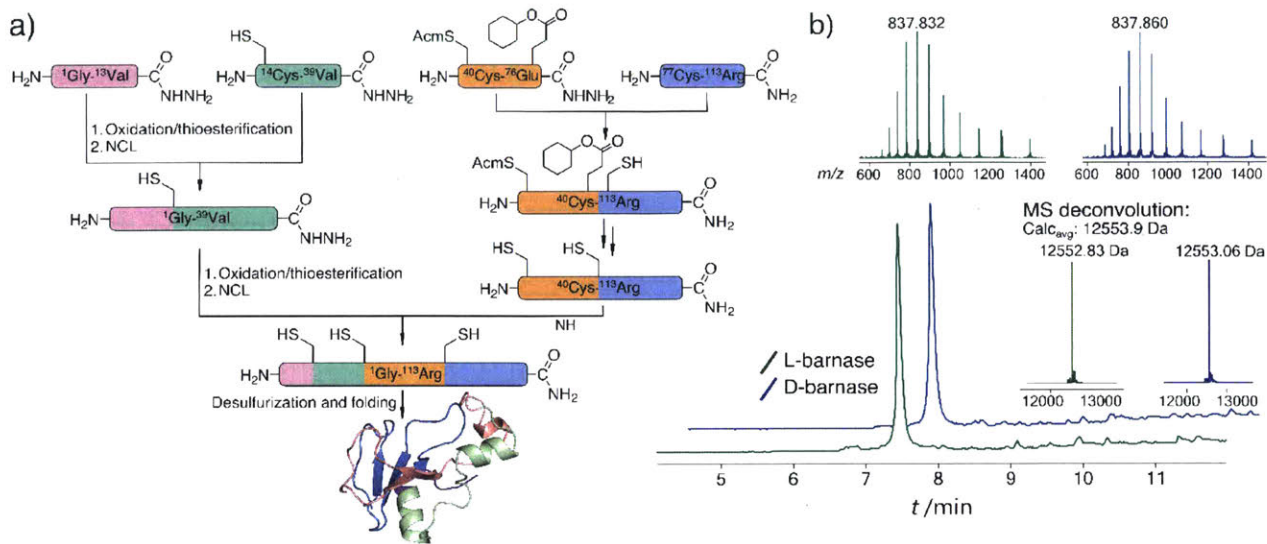
Mirror image enzymes (MIEs), enantiomers of naturally occurring enzymes, are a promising therapeutically relevant class of biomolecules. These enzymes are thought to be more proteolytically stable and less immunogenic than their native counterparts, while possessing catalytic activity with reciprocal chiral specificity.<sup>1</sup> Enhanced proteolytic stability and low immunogenicity of mirror image proteins were demonstrated with D-rubredoxin, which was stable to chymotrypsin<sup>2</sup> and did not elicit an immune response in mice in contrast to its native enantiomer.<sup>3</sup> The catalytic function of MIEs was studied on the examples of D-HIV-1 protease, which cleaved the D-substrate and not its L-form,<sup>4</sup> and D-4-oxalocrotonate tautomerase (4-OT), which acted on the same achiral substrate as its L-4-OT, but produced the enantiomeric product.<sup>5</sup> These studies confirmed the reciprocal chiral specificity of MIEs. In a more recent study,<sup>6</sup> the GroEL/ES-assisted folding of mirror image DapA revealed that MIEs may be folded by the native chaperones. Unfortunately, the field of MIEs is still largely unexplored, as only three enzymes were synthesized to date: the mentioned reports represent all published data regarding properties of MIEs.

We undertook this study to systematically investigate properties of an MIE in greater detail. To this end, we synthesized and characterized the enantiomers of *B. amyloliquefaciens* ribonuclease (barnase). Barnase is a potent guanyl-specific,<sup>7</sup> single-strand RNA-specific<sup>8</sup> endonuclease that operates via the classical mechanism of RNA hydrolysis, producing a 2',3'-cyclic phosphate as an intermediate.<sup>9</sup> The enzyme is more active towards long RNA molecules with the optimum pH at 8.5, but it also hydrolyzes substrates as short as dinucleotides.<sup>10</sup> We identified barnase an ideal target for this study due to its structural simplicity (the protein is comprised of a single 110 amino acid residue polypeptide chain with no cysteines),<sup>11</sup> reversible folding-unfolding transition,<sup>12</sup> and straightforward catalytic activity with a fairly simple readout. Additionally, barnase may be relevant biologically; as bacterial ribonucleases are not inhibited by human Ribonuclease Inhibitor, barnase exhibits strong cytotoxicity on mammalian cells, and shows promising antitumor activity when conjugated to humanized HER-2 antibody.<sup>13</sup>

## 2.2. Results and Discussion

### 2.2.1. Protein Synthesis

To synthesize both enantiomers of barnase, we used a previously established strategy with minor revisions (Chapter 1, Figure 2.1a).<sup>14</sup> In short, four peptide fragments comprising the protein were rapidly synthesized on the fast flow peptide synthesis platform<sup>15</sup> and purified by RP-HPLC. To increase the yields of Liu's oxidation/native chemical ligation (NCL) protocol<sup>16</sup> we performed all ligations in two steps, isolating intermediate thioesters by RP-HPLC. Thus, the C-terminal hydrazide  $\text{H}_2\text{N}-[\text{Gly}^1-\text{Val}^{13}]-\text{CON}_2\text{H}_3$



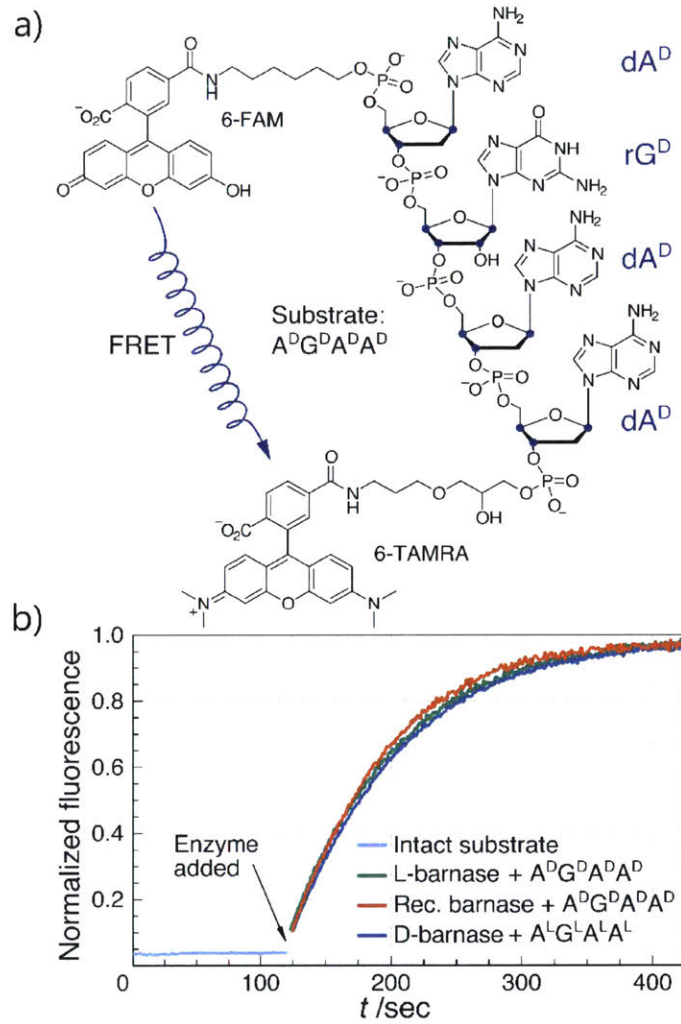
### Figure 2.1. Protein synthesis

a) Synthetic strategy for L- and D-barnase. b) HPLC-MS (TIC) chromatograms for the synthesized proteins with mass spectra insets for the main peaks. Maximum entropy deconvolution spectra of the MS spectra on the top are displayed on the bottom. For both proteins main identified contaminants were +32 Da (Ala  $\rightarrow$  Cys, incomplete desulfurization), and -57 Da (Gly deletion).

was subject to  $\text{NaNO}_2$  oxidation and 4-mercaptophenylacetic acid (MPAA) transesterification, which afforded the C-terminal thioester of the N-terminal fragment. In the second step, the thioester was ligated with  $\text{H}_2\text{N}-[\text{Cys}^{14}\text{-Val}^{39}]\text{-CON}_2\text{H}_3$  to obtain  $\text{H}_2\text{N}-[\text{Gly}^1\text{-Val}^{39}]\text{-CON}_2\text{H}_3$  with 74% yield over two steps (the one-pot procedure used in the original synthesis of L-barnase yielded 67% of the product). In an analogous manner,  $\text{H}_2\text{N}-[\text{Cys}^{40}(\text{Acm})\text{-Arg}^{113}]\text{-CONH}_2$  was synthesized from  $\text{H}_2\text{N}-[\text{Cys}^{40}(\text{Acm})\text{-Glu}^{76}(\text{Cy})]\text{-CON}_2\text{H}_3$  and  $\text{H}_2\text{N}-[\text{Cys}^{77}\text{-Arg}^{113}]\text{-CONH}_2$  with 85% yield over two steps, up from 77% yield obtained originally. After the acetamidomethyl and cyclohexyl protecting groups were removed from the C-terminal segment, the final NCL reaction between  $\text{H}_2\text{N}-[\text{Gly}^1\text{-Val}^{39}]\text{-CON}_2\text{H}_3$  and  $\text{H}_2\text{N}-[\text{Cys}^{40}\text{-Arg}^{113}]\text{-CONH}_2$  afforded full length  $^{14,40,77}\text{Cys}$ -barnase tri-mutant. This step proceeded rather slowly and inefficiently during the original synthesis (>36 hours to go to completion with 34% yield after purification). To accelerate it, we used twofold excess of the N-terminal fragment and increased the concentrations of both peptides, up to 4.5 mM and 9.0 mM. These changes increased the yield of the reaction up to 65% and allowed to run it overnight. Finally, mild desulfurization<sup>17</sup> with TCEP, MESNa, and VA-044 radical initiator yielded the desired protein, which was refolded from 6M  $\text{Gn}\cdot\text{HCl}$  in 50 mM Tris, 100 mM NaCl (pH 7.4) buffer. The improved NCL conditions increased the overall yield of the protein to 19% up from original 12% (calculated from the purified starting fragments). Both synthesized proteins contained a  $\text{Gly}_3$  N-terminal tag to facilitate Sortase A-mediated ligation for future studies. L- and D-Barnase prepared in this way were characterized by HPLC-MS (Figure 2.1b) and found identical to each other by liquid chromatography and mass spectrometry. Additionally, synthetic proteins were similar to recombinant barnase, which lacked the  $\text{Gly}_3$  tag: the mass difference in the deconvoluted MS spectrum was 171 Da, consistent with extra  $\text{Gly}_3$  for synthetic variants (Section 2.3.3).

### 2.2.2. Enzymatic Characterization of D-Barnase

With both L- and D-barnase proteins in hand, we turned to characterizing the catalytic activity of these enzymes. As the RNase activity assay we utilized a modified version of the fast, supersensitive fluorogenic assay developed by Raines and colleagues.<sup>18</sup> The substrates for the assay are DNA/RNA hybrids with a single cleavage site, which provides for a homogeneous substrate needed to establish kinetic parameters for enzyme catalyzed hydrolysis (Figure 2.2a). During the cleavage of the substrate, fluorescence resonance energy transfer between 6-carboxyfluorescein (6-FAM) and 6-carboxytetramethyl-rhodamine (6-TAMRA) fluorophores, installed on the 5' and 3'-end respectively, is perturbed. Thus, the increase in fluorescence of 6-FAM at 515 nm upon excitation at 495 nm can be monitored as a function of time to measure kinetics of the substrate hydrolysis. Enzyme kinetic parameters (primarily,  $k_{cat}/K_M$ ) can then be obtained by the non-linear regression of experimental data to the enzyme catalyzed pseudo-first order rate equation (Section 2.3.2). In this study we investigated several different tetraoligonucleotide substrates of the common structu-



**Figure 2.2. Enzymatic characterization of D-barnase**

a) Chemical structure of the A<sup>D</sup>G<sup>D</sup>A<sup>D</sup>A<sup>D</sup> fluorogenic substrate. Stereogenic centers are highlighted in blue.  
 b) Biochemical characterization of synthetic L- and D-barnase using the fluorogenic assay. One representative kinetic curve is shown for each enzyme. Barnase concentration was 1.0 nM in all cases.

re 6-FAM-dA<sup>X</sup>-rN<sup>X</sup>-dA<sup>X</sup>-dA<sup>X</sup>-6-TAMRA, henceforth A<sup>X</sup>N<sup>X</sup>A<sup>X</sup>A<sup>X</sup>, where N is a certain nucleotide, and the superscript X annotates the chirality of the sugar (D-sugars constitute native RNA and L-sugars — its enantiomer). In a typical assay, enzyme (100 pM to 100 nM) was added to 50-200 nM substrate in MES buffer (100 mM MES, 100 mM NaCl, pH 6.0), and the fluorescence emission was monitored. In cases where enzyme was unable to hydrolyze the substrate completely in under 60 minutes, an additional aliquot of enzyme was added to promote hydrolysis and measure the final fluorescence of the fully hydrolyzed material.

We first compared the catalytic efficiency of synthetic L-barnase to its recombinant analogue. Because barnase is known as a guanyl-specific endonuclease, we studied the hydrolysis of A<sup>D</sup>G<sup>D</sup>A<sup>D</sup>A<sup>D</sup>. We found that synthetic L-barnase had a  $k_{cat}/K_M$  value of  $(1.2 \pm 0.1) \cdot 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$  (Table 2.1), in line with the activity of the recombinant enzyme ( $(1.3 \pm 0.4) \cdot 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ , Figure 2.2b). For D-barnase, we expected the reciprocal catalytic activity (i.e., hydrolysis of the mirror image substrate A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup>). Indeed, D-barnase hydrolyzed this substrate efficiently with  $k_{cat}/K_M = (1.1 \pm 0.2) \cdot 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ , thus confirming our hypothesis. For all studied proteins, the observed hydrolysis was responsive towards different enzyme concentrations. These data allowed us to conclude that both synthetic enzymes had full catalytic activity.

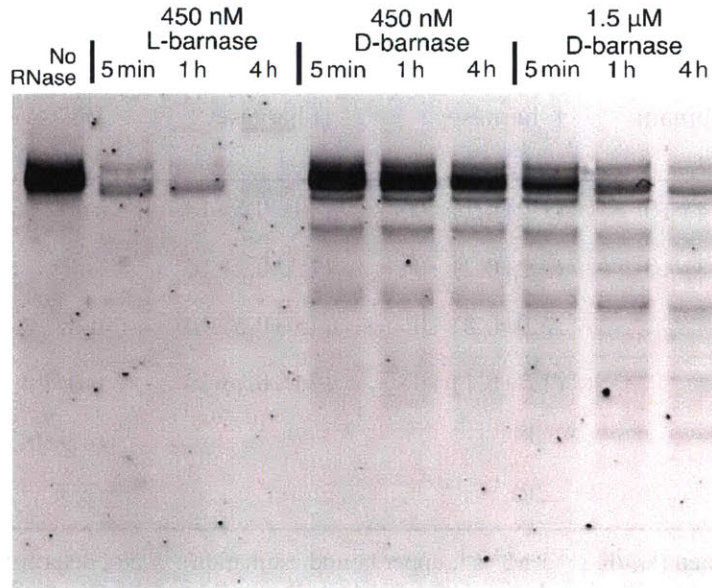
Next, we sought to study the substrate stereospecificity of the enzymes, i.e., to evaluate the hydrolysis of A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup> by L-barnase and of A<sup>D</sup>G<sup>D</sup>A<sup>D</sup>A<sup>D</sup> by D-barnase. Unexpectedly, we found significant remaining activity in both cases: the  $k_{cat}/K_M$  for L-barnase was  $(3.2 \pm 0.2) \cdot 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$ , and  $(3.0 \pm 0.6) \cdot 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$  for D-barnase. Although these values are ~4000 times lower than the corresponding ones for the native substrates,  $k_{cat}/K_M$  of  $3 \cdot 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$  still represents a fairly potent enzyme<sup>19</sup> with the rate acceleration of  $\sim 10^{10}$  over the uncatalyzed RNA hydrolysis.<sup>20</sup> The similarity of  $k_{cat}/K_M$  values suggest the observation is not due to an artifact or RNase contamination. However, we performed additional experiments to exclude these possibilities. We used barstar, a well-known barnase-specific inhibitor,<sup>21</sup> to probe its efficiency in the assays. We found that addition of two equivalents of barstar completely abolished the catalytic activity of L-barnase for both A<sup>D</sup>G<sup>D</sup>A<sup>D</sup>A<sup>D</sup> and A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup>, confirming that L-barnase is responsible for the cleavage of the substrates. Additionally, recombinant L-barnase, obtained independently from synthetic enzymes, had  $k_{cat}/K_M$  of  $(3.3 \pm 0.3) \cdot 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$  towards A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup>. Finally, a common source of RNase contamination are RNase A family enzymes, which are pyrimidine rather than purine specific,<sup>22</sup> and thus are not expected to cleave the studied substrates. Accordingly, we did not detect hydrolysis of either A<sup>D</sup>G<sup>D</sup>A<sup>D</sup>A<sup>D</sup> or A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup> substrates by RNase A of up to 50 nM. Taken together, these data suggested that barnase may accommodate substrates of the opposite chirality.

To further investigate this phenomenon, we studied the hydrolysis of “mixed chirality” substrates, A<sup>L</sup>G<sup>D</sup>A<sup>L</sup>A<sup>L</sup> and its enantiomer A<sup>D</sup>G<sup>L</sup>A<sup>D</sup>A<sup>D</sup>, by L- and D-barnase. We found that both substrates were hydrolyzed by the enzymes less efficiently than the native substrates, but significantly faster than tetranuc-

**Table 2.1. Catalytic activities values ( $k_{cat}/K_M$  in  $M^{-1}\cdot s^{-1}$ ) of select RNases towards different fluorogenic substrates**

	Recombinant barnase,	L-barnase	D-barnase	RNase A
$A^D G^D A^D A^D$	$(1.3 \pm 0.4) \cdot 10^7$	$(1.2 \pm 0.1) \cdot 10^7$	$(3.0 \pm 0.6) \cdot 10^3$	— <sup>[a]</sup>
$A^L G^L A^L A^L$	$(3.3 \pm 0.3) \cdot 10^3$	$(3.2 \pm 0.2) \cdot 10^3$	$(1.1 \pm 0.2) \cdot 10^7$	— <sup>[a]</sup>
$A^D G^L A^D A^D$	n.d. <sup>[b]</sup>	$(6.9 \pm 0.8) \cdot 10^4$	$(1.0 \pm 0.5) \cdot 10^5$	n.d. <sup>[b]</sup>
$A^L G^D A^L A^L$	n.d. <sup>[b]</sup>	$(1.7 \pm 0.1) \cdot 10^5$	$(3.0 \pm 0.7) \cdot 10^4$	n.d. <sup>[b]</sup>
$A^D C^D A^D A^D$	n.d. <sup>[b]</sup>	— <sup>[a]</sup>	— <sup>[a]</sup>	$(4.9 \pm 0.3) \cdot 10^7$
$A^L C^L A^L A^L$	n.d. <sup>[b]</sup>	— <sup>[a]</sup>	— <sup>[a]</sup>	— <sup>[a]</sup>

<sup>[a]</sup>Hydrolysis was not detected ( $k_{cat}/K_M < 1 M^{-1}\cdot s^{-1}$ , upper bound estimation). <sup>[b]</sup>Not determined.  $k_{cat}/K_M$  values in  $M^{-1}\cdot s^{-1} \pm$  one standard deviation are displayed.



**Figure 2.3. Digestion of native RNA by L- and D-barnase**

The RNA gel showing the digest of 112 nucleotide-long D-RNA by 450 nM L-barnase, 450 nM D-barnase, and 1.5  $\mu$ M D-barnase over the course of four hours. The negative control (no RNase added) is shown on the left. The image is digitally modified by inverting the color scheme and adjusting the contrast.



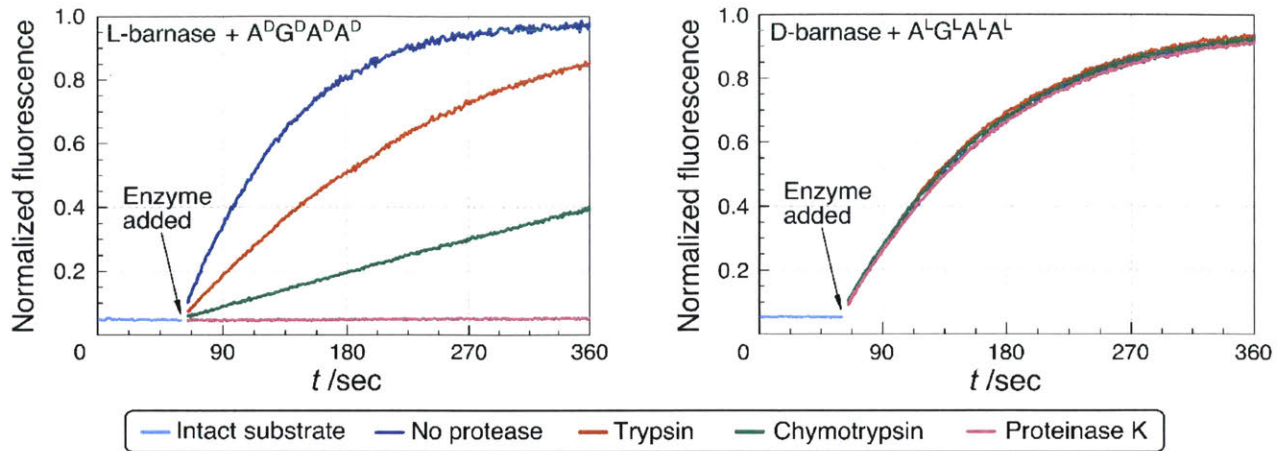
leotides with the fully inverted stereochemistry (Table 2.1). Thus, D-barnase hydrolyzed A<sup>D</sup>G<sup>L</sup>A<sup>D</sup>A<sup>D</sup>, (the recognition guanosine had the correct chirality, while the rest was inverted) only ~100 times less efficiently than A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup> with  $k_{cat}/K_M$  as high as  $(1.0 \pm 0.5) \cdot 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$ . The second substrate, A<sup>D</sup>G<sup>L</sup>A<sup>D</sup>A<sup>D</sup>, which had only the guanosine chirality inverted, was hydrolyzed by D-barnase ~350 times slower than its native substrate. These results were corroborated by the data for L-barnase. At the same time, we could not detect hydrolysis of either A<sup>D</sup>C<sup>D</sup>A<sup>D</sup>A<sup>D</sup> or its enantiomer, A<sup>L</sup>C<sup>L</sup>A<sup>L</sup>A<sup>L</sup>, by L- or D-barnase. As a positive control for this experiment, we demonstrated the efficient hydrolysis of A<sup>D</sup>C<sup>D</sup>A<sup>D</sup>A<sup>D</sup> by RNase A, which was consistent with previous reports. Interestingly, we could not detect the cleavage of A<sup>L</sup>C<sup>L</sup>A<sup>L</sup>A<sup>L</sup> by RNase A.

Collectively, these results confirmed that barnase allows variations in the chirality of its substrates. The chirality of the main recognition nucleoside, guanosine, appears to be more important than the chirality of the rest of the substrate. Moreover, it seems barnase is not simply promiscuous because it did not hydrolyze ACAA substrates, where the key guanosine was replaced by a pyrimidine-based nucleoside. We also found that RNase A did not hydrolyze an enantiomer of its native substrate, which implies that the low substrate stereospecificity is not a universal phenomenon amongst digestive ribonucleases.

To expand our findings beyond the fluorogenic assay we sought to study the hydrolysis of native RNA by D-barnase. Towards this end, we incubated a 70 µg/mL solution of a native 112-mer RNA in 10 mM Tris, 50 mM NaCl buffer (pH 7.4) with various concentrations of D-barnase or 450 nM L-barnase for up to 4 hours and analyzed the RNA digest products by performing 10% denaturing PAGE (Section 2.3.4). As demonstrated in Figure 2.3, the presence of the low molecular weight bands in cases where D-barnase was added, but not in the negative control lane, indicates that D-barnase cleaved native RNA, albeit slower than L-barnase. The latter observation is evident from digests by 450 nM D-barnase versus L-barnase. As such, we confirmed that the results of the fluorogenic assay translate into more complex systems, involving native substrates, and thus, that D-barnase is active towards D-RNA.

### 2.2.3. Proteolytic Stability of D-Barnase

In another part of the study, we compared the stability of L- and D-barnase towards common digestive proteases *in vitro*. Proteolysis of mirror image proteins was investigated before with metal-bound D-rubredoxin, which was completely stable to chymotrypsin in contrast to its enantiomer.<sup>2</sup> Although there is evidence for the enhanced proteolytic stability of short, mostly unfolded D-peptides,<sup>23,24</sup> the rubredoxin study remains the only published example of such behavior for folded mirror image proteins. We aimed to study the proteolytic stability of an MIE in greater detail, assaying different proteases and digestion conditions.



**Figure 2.4. Comparison of the proteolytic stability of L- and D-barnase.**

Remaining catalytic activity of the enzymes, associated with the extent of the proteolytic digestion, was measured using the fluorogenic assay. Three out of six assayed proteases are not displayed for clarity. Digestion of 1.0 nM L-barnase with different proteases was assayed using A<sup>D</sup>G<sup>D</sup>A<sup>D</sup>A<sup>D</sup> as a substrate (data shown on the left). Digestion of 0.8 nM D-barnase with different proteases was assayed using A<sup>L</sup>G<sup>L</sup>A<sup>L</sup>A<sup>L</sup> as a substrate (on the right).

As proteases for this study we selected bovine trypsin,  $\alpha$ -chymotrypsin, proteinase K, porcine elastase type IV, papain, and *S. griseus* protease (actinase E). These enzymes were chosen for their robust digestive proteolytic activities and a wide range of substrate specificities. Papain represented cysteine superfamily proteases, while other enzymes were serine proteases. Additionally, we wanted to assay enzymes, which are able to recognize and cleave peptide bonds after glycine. Since glycine is achiral, we hypothesized that such proteases may potentially recognize and accommodate glycine residues in mirror image proteins, allowing for the hydrolysis of these substrates. Although glycine-specific digestive proteases are uncommon, both elastase and papain are reported to cleave their substrates after glycine fairly efficiently.<sup>25,26</sup>

First, we performed the non-denaturing, in-solution digestion of L- and D-barnase by the selected proteases. Proteins were incubated in appropriate buffers (Section 2.3.5) at 37 °C for up to 19 hours with a 15 : 1 ratio of barnase to protease. The extent of the digestion was determined by HPLC-MS analysis, and by measuring the remaining ribonucleic activity via the fluorogenic assay. As shown in Table 2.2 and Figure 2.4, we found that after 19 hours of digestion L-barnase demonstrated differential stability towards proteases: trypsin-digested barnase had 36% its native activity, while in the case of proteinase K less than 0.2% activity remained. In all six cases L-barnase lost a significant portion of its catalytic activity. In contrast, D-barnase proved completely stable to all assayed proteases; it retained full catalytic activity, and no digestion products could be detected by HPLC-MS.

Additionally, we performed a more forcing in-solution denaturing digestion of L- and D-barnase using the most potent protease, proteinase K. To denature the protein, barnase was incubated in 6M Gn·HCl, 50 mM Tris buffer (pH 7.4) at 95 °C for 20 minutes, and then digested with proteinase K (barnase : protease = 2 : 1) in 2M Gn·HCl at 37 °C. We found that D-barnase was completely stable to proteolysis even under such forcing conditions, in stark contrast to L-barnase, which was digested completely (Section 2.3.5). Finally, we attempted to digest D-barnase with proteinase K by increasing the digestion time. Using HPLC-MS analysis we did not detect any digestion products after 168 hours (1 week) of incubating D-barnase with proteinase K. The digest was indistinguishable from a negative control experiment, where no protease was added to the enzyme (Section 2.3.5).

**Table 2.2. Remaining RNase catalytic activity of L- and D-barnase after the proteolytic digestion with select proteases**

Protease	L-barnase	D-barnase
No protease	100.0 ± 6.4	100.0 ± 5.1
Trypsin	36.7 ± 2.0	108.0 ± 8.9
Chymotrypsin	9.2 ± 0.5	103.1 ± 5.3
Proteinase K	0.2 ± <0.1	99.8 ± 8.1
Elastase	0.6 ± <0.1	103.5 ± 3.3
Papain	0.8 ± <0.1	96.6 ± 4.3
Actinase E	0.3 ± <0.1	101.4 ± 7.3

Values for the remaining barnase activity are normalized to the negative control experiment, where no protease was added to the enzyme, and are displayed as the percentage of the full ribonuclease activity ± one standard deviation.

## 2.3.4. Conclusions

In summary, we successfully synthesized and characterized mirror image barnase. We found that the enzyme was fully active towards mirror image RNA model substrates and was somewhat promiscuous to its substrate chirality. After the systematic investigation of this phenomenon we used mirror image barnase to demonstrate the cleavage of the native RNA by a mirror image enzyme. Separately, we found that D-barnase appears to be extremely proteolytically stable. Our experiments revealed that neither cysteine nor serine superfamily proteases are able to cleave it even under forcing conditions. Contrary to our initial hypothesis, D-barnase was completely stable towards proteases that are able to cleave peptide bonds after achiral glycine. In short, we were unable to find proteases and/or reaction conditions which would lead to the digestion of D-barnase.

The results of this study pose a number of questions. First, it is unclear how barnase recognizes and cleaves substrates of the opposite chirality. The enzyme is known to have several subsites, which facilitate the substrate binding and its proper orientation for catalysis. Our data are consistent with this model, as we observed a range of  $k_{cat}/K_M$  values by only changing the chirality of AGAA tetranucleotide. It is conceivable that substrates of the mixed chirality, e.g. A<sup>L</sup>G<sup>D</sup>A<sup>L</sup>A<sup>L</sup>, occupy only certain subsites in the enzyme, e.g. the guanosine binding subsite in this case, and thus the catalysis may still proceed. It is also unclear whether this enzymatic activity is merely spontaneous or was subject to the evolutionary selection at some point. At this time we are unaware of any practical implications of such catalysis: to the best of our knowledge, RNA of L-configuration is unknown in nature.

Nevertheless, our study suggests that at least in some cases enzymes may utilize substrates of the opposite chirality. The mirror image form of such an enzyme will then act on the same targets as its native counterpart. Although decreased catalytic efficiency is expected, the enzyme may still achieve a notable rate acceleration. This property of MIEs can be highly desirable from the biotechnology standpoint for, as we confirmed in the case of barnase, MIEs can be extraordinarily resistant to proteolysis and at the same time carry the native biological function. Importantly, this effect may manifest itself without protein engineering and/or evolution of the enzyme. As we also found with the example of RNase A, this effect by no means is universal, and more investigations would be needed to establish the generality of our findings.

## 2.3. Experimental

### 2.3.1. General

All reagents for protein synthesis were as described in Chapter 1. 2-(*N*-morpholino)-ethanesulfonic acid (MES) monohydrate, guanidine hydrochloride (Gn·HCl), sodium hydroxide, hydrochloric acid, Tris base, Tris acid, LB broth, and RNase-free water were from VWR International, Pennsylvania. HPLC-purified

tetraoligonucleotides for fluorogenic assays were purchased from ChemGenes, Massachusetts. RNase A was from Qiagen. All other reagents were from Sigma Aldrich and Life Technologies. All reagents were reagent grade and used as received.

### 2.3.2. Methods

*HPLC-MS Analysis.* All proteins and reaction mixtures were analyzed on an Agilent 6520 Accurate Mass Q-TOF LC-MS using an Agilent Zorbax 300SB C3 column (300 Å, 5 µm, 2.1 x 150 mm). Two methods were used interchangeably. Both methods were operated at 40 °C and a flow rate of 0.8 mL/min. The first method (1-61) used the following gradient: 1% acetonitrile with 0.1% formic acid added (FA, solvent B) in water with 0.1% FA (solvent A) for 2 min, 1-61% B in A ramping linearly over 11 min, 61% B in A for 1 minute. The second method (5-65) used the following gradient: 5% B in A for 2 min, 5-65% B in A ramping linearly over 11 min, 65% B in A for 1 minute. Unless noted, all chromatograms shown in this work are plots of total ion current (TIC) versus time. The data were analyzed using Agilent MassHunter Qualitative analysis software. All MS deconvolution spectra were obtained using the maximum entropy algorithm.

*Fluorogenic RNase activity assay.* The assay buffer (100 mM MES, 100 mM NaCl, pH 6.00) was prepared in RNase-free water. Prior to the experiments oligonucleotides were dissolved in the assay buffer, and enzymes were buffer exchanged into the assay buffer. 50–200 nM oligonucleotides and 100 pM to 100 nM enzymes were used to perform the experiments. In a typical experiment, the fluorescence of the substrate solution at 515 nm upon excitation at 495 nm was monitored for 60-300 seconds to measure the starting fluorescence,  $I_0$ , and to ensure no background cleavage took place prior to the addition of enzymes. Then, an aliquot of enzyme was added to the substrate and the solution was rapidly mixed. The increase in fluorescence was then monitored for 120–1600 seconds. In cases where the hydrolysis was not complete, an additional aliquot of the appropriate enzyme was added to measure the final fluorescence of the fully digested substrate.

Substrate concentrations in these assays were much lower than  $K_M$  values for the enzymes. For example, barnase has  $K_M = 36.3 \mu\text{M}$  towards CGAC tetraoligonucleotide.<sup>10</sup> Thus,  $k_{cat}/K_M$  values can be obtained directly from the first order rate equation without resorting to Michaelis-Menten kinetics. The rate equations in this case can be expressed as such:

$$I = I_f - (I_f - I_0)e^{-\frac{k_{cat}}{K_M}[E]t} \quad (1), \quad \text{and}$$

$$I = I_0 + (I_f - I_0)\frac{k_{cat}}{K_M}[E]t \quad (2),$$

where  $I$  is the fluorescence at time  $t$ ,  $I_0$  is fluorescence of the intact substrate,  $I_f$  is the fluorescence of the hydrolyzed substrate,  $[E]$  is the total enzyme concentration,  $k_{cat}$ ,  $K_M$  are steady state enzyme kinetic

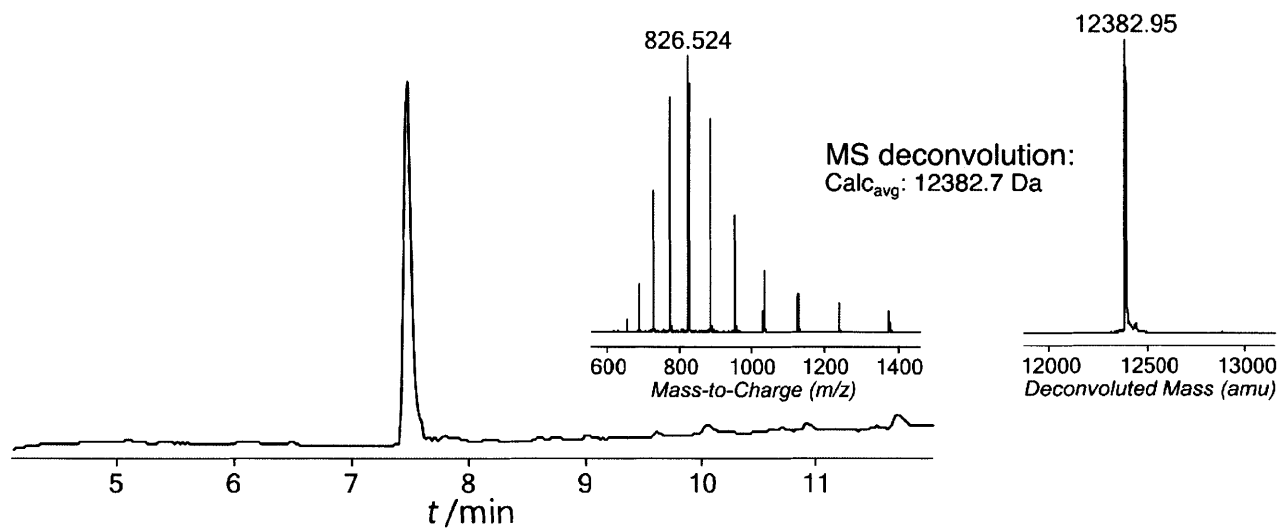
parameters. The values of  $I_f$  were obtained by non-linear least squares regression of (1) with data collected with the addition of sufficient enzyme to hydrolyze all the substrate. The values of  $k_{cat}/K_M$  were obtained by either least squares linear (using eq. (2)) or non-linear regression (using eq. (1)) depending on the efficiency and concentration of the enzymes towards particular substrates. All experiments were performed at least in triplicate.

### 2.3.3. Protein Expression and Purification

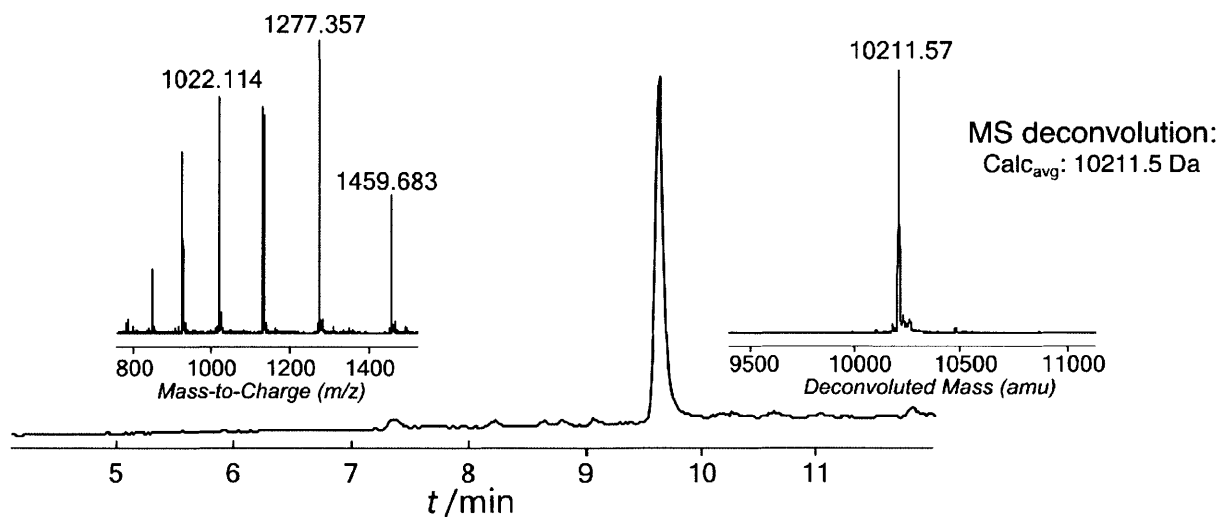
Barnase was co-expressed with its inhibitor, barstar, from pMT1002 (Addgene plasmid 8621) under the control of a strong heat inducible Pr promoter in *E. Coli* strain XL-1 Blue.<sup>27</sup> 1 mL of 6-h culture of *E. Coli* carrying the plasmid grown in LB medium with ampicillin (100 µg/mL), was diluted 1 : 1000 into the same medium and grown overnight at 28 °C, shaking at 220 rpm. Then, (the culture had a density of  $\sim OD_{600} = 0.6$ ) a further 500 mL of hot (85 °C) LB medium was poured into the cell culture, and the shaker was incubated at 42 °C for another 30 minutes. After that, the temperature was adjusted to 37 °C and cells were cultured for another 16 hours. Acetic acid was added to the culture until the final pH of 4.3, and the cells were centrifuged. The pellet was discarded and 7.5 mL SP-Sepharose resin was added to the supernatant. After stirring for 2 hours at 4 °C the resin was allowed to settle, and the supernatant was loaded on the 5 mL HiTrap Capto S column (GE Healthcare, UK). Pure barnase was obtained by cation exchange chromatography and was buffer exchanged into appropriate buffers using a HiPrep 26/10 Desalting column. The protein purity was assessed by HPLC-MS and PAGE.

The barstar gene was cloned from pMT1002 into a pET-SUMO vector and expressed in *E. coli* strain BL21 (DE3). The cells were cultured in LB medium with 30 µg/mL kanamycin at 30 °C until  $OD_{600}$  was 0.6, and then induced with 0.4 mM IPTG and grown at 37 °C overnight. Cells were centrifuged and pellets resuspended in 100 mL of 50 mM Tris, 150 mM NaCl, pH 7.5 buffer containing 200 mg lysozyme, 4 mg Roche DNase I, and 2 tablets of Roche protease inhibitor cocktail. The cells were lysed using sonication and the suspension was centrifuged further. The lysate was loaded onto 5mL GE HisTrap FF crude Ni-NTA column pre-equilibrated with binding buffer (20 mM Tris, 150 mM NaCl, pH 8.5). The column was washed with 50 mL binding buffer and then 50 mL binding buffer containing 40 mM imidazole. Barstar was eluted with 50 mL 500 mM imidazole in the binding buffer and buffer exchanged to remove imidazole using a HiPrep 26/10 Desalting column (GE Healthcare, UK).

Concentrations of proteins were determined spectrophotometrically at 280 nm using appropriate extinction coefficient values and confirmed by HPLC-MS using the standard calibration curves.



**Figure 2.5. HPLC-MS characterization of recombinant barnase**



**Figure 2.6. HPLC-MS characterization of recombinant barstar**



#### 2.3.4. Native RNA Digestion by D-Barnase

The 112mer RNA (5' - GGG ACA AUU ACU AUU UAC AAU UAC AAU GGG UUG CUU CAC GAA CGG GUU GUU GUA CGA GUC CAA GGC GUG CGG CGG UAG CUU AGG CCA CCA UCA CCA UCA CCA CCG GCU AUA G - 3') was prepared via T7 run off transcription from a double stranded DNA template (5' – TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GGT TGC TTC ACG AAC GGG TTG TTG TAC GAG TCC AAG GCG TGC GGC GGT AGC TTA GGC CAC CAT CAC CAT CAC CAC CGG CTA TAG – 3') and purified by 8% denaturing polyacrylamide gel electrophoresis. Following ethanol precipitation, the pellet was dissolved in 10 mM Tris, 50 mM NaCl, pH 7.4 buffer and the concentration was determined spectrophotometrically at 260 nm.

RNA cleavage reactions were initiated by addition of barnase (L- or D-) to a solution of RNA in 10 mM Tris, 50 mM NaCl, pH 7.4 buffer, and left at room temperature. A control RNA solution, lacking barnase, was similarly prepared and left at room temperature. The reactions were quenched at the desired time points by the addition of 4 µl of reaction solution to 16 µl of denaturing gel loading dye (8 M urea, 20 mM EDTA, 0.5% bromophenol blue, 5mM tris pH 7.5) and rapidly frozen on liquid nitrogen. Immediately prior to gel analysis, the aliquots were heated to 95 °C for 5 minutes and then run on a 10% denaturing polyacrylamide gel, stained with ethidium bromide and visualized on the Red Gel Imager (ProteinSimple, USA).

#### 2.3.5. Proteolysis Experiments

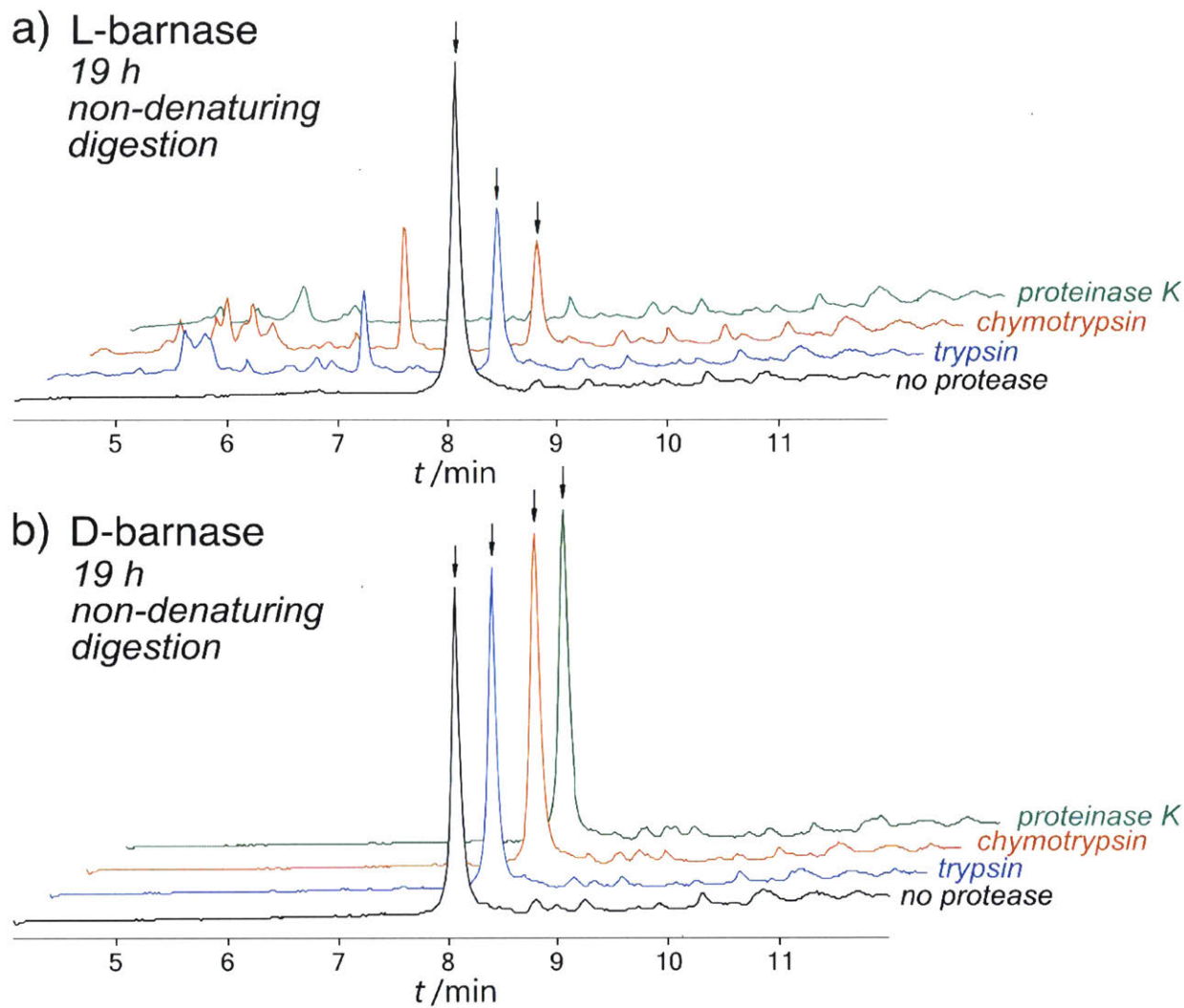
*Non-denaturing in-solution digestion.* The experiments were performed in the following buffers:

- For trypsin, chymotrypsin and proteinase K experiments: 50 mM Tris, 100 mM NaCl, 5 mM CaCl<sub>2</sub> buffer (pH 7.4);
- For elastase and actinase E experiments: 50 mM Tris, 50 mM NaCl, buffer (pH 8.0);
- For papain experiments: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM EDTA, 10 mM L-cysteine buffer (pH 6.3);

Proteins were incubated in appropriate buffers at 37 °C at a 15 : 1 ratio of barnase to protease. HPLC-MS analysis was performed after 19 hours of incubation time. The reactions were quenched by adding excess 50% acetonitrile/50% water with 0.1% TFA added. The results of the analysis are summarized in Figures 2.7 and 2.8.

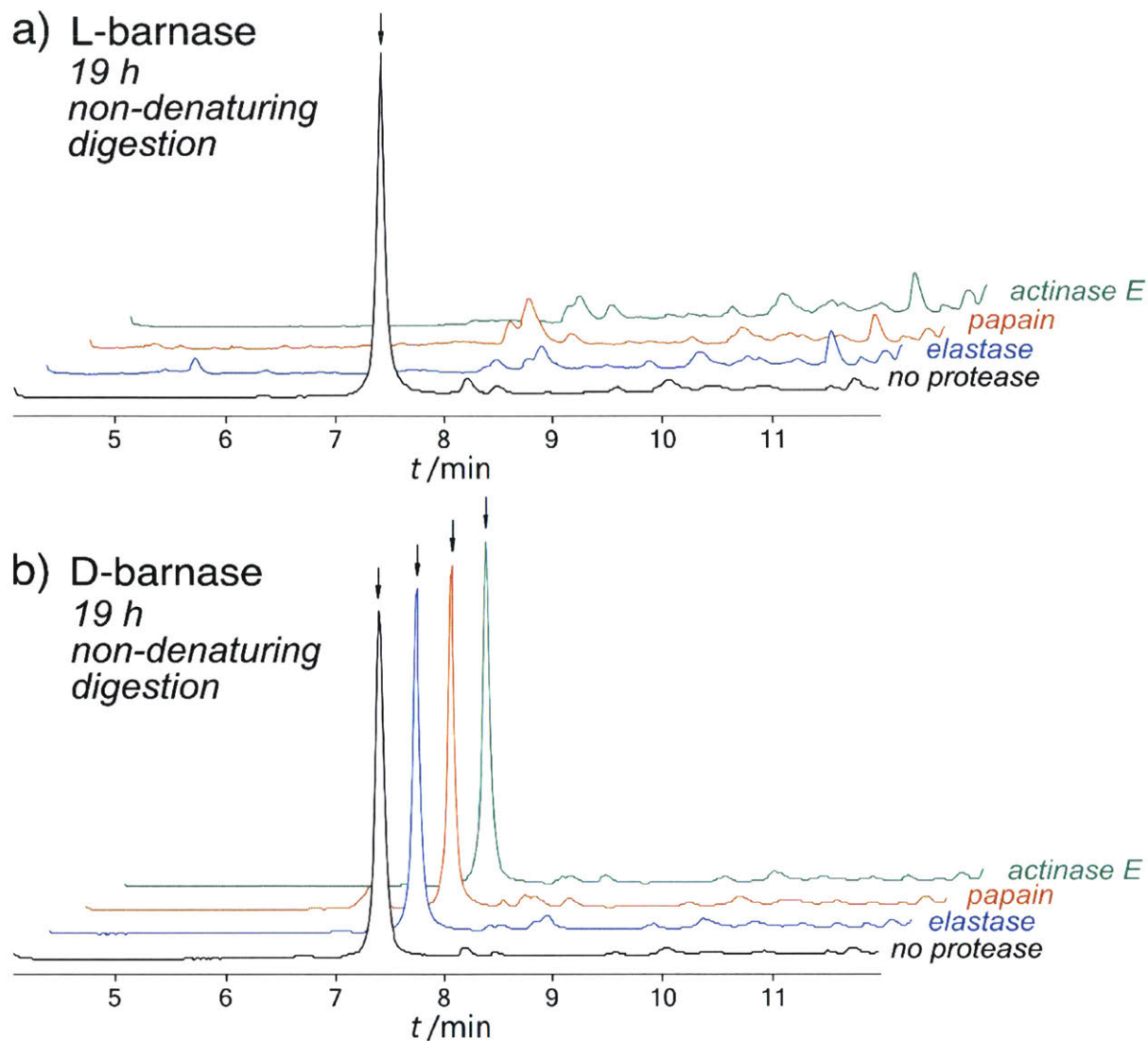
*Denaturing in-solution digestion.* The experiment was performed as described in section 2.2.3. HPLC-MS analysis was performed after 4 hours of incubation time to compare L- and D-barnase. Additionally, D-barnase was analyzed after 19 hours of incubation with proteinase K, and was also found to be completely stable. The reactions were quenched by adding excess 50% acetonitrile/50% water with 0.1% TFA added. The results of the analysis are summarized in Figure 2.9.

*Digestion of barnase with proteinase K.* The week-long non-denaturing in-solution digestion of D-barnase by proteinase K was performed under the following conditions. The digestion mixture containing 1.28 mg/mL D-barnase and 0.26 mg/mL proteinase K in 50 mM Tris, 100 mM NaCl, 5 mM CaCl<sub>2</sub> buffer (pH 7.5) in a total volume of 25  $\mu$ L was incubated at 37 °C for one week. A sample for HPLC-MS analysis (1  $\mu$ L) was taken every 48 hours, and additional proteinase K (1  $\mu$ L of 1.28 mg/mL stock in the same buffer) was added to the digestion mixture at the same time. As a negative control experiment, D-barnase was incubated under identical conditions free of proteinase K. The reactions were quenched by adding excess 50% acetonitrile/50% water with 0.1% TFA added. The HPLC-MS data of the digestion endpoint are summarized in Figure 2.10.



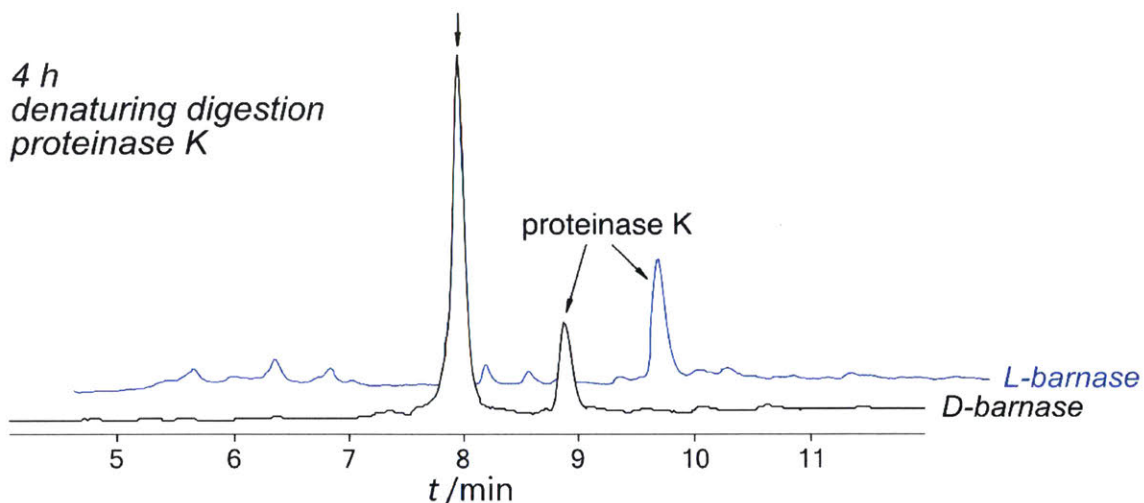
**Figure 2.7. Relative proteolytic stabilities of L- and D-barnase**

Comparison of stabilities of L- and D-barnase towards trypsin, chymotrypsin and proteinase K under non-denaturing digestion conditions using HPLC-MS (TIC chromatograms are displayed, 1-61 chromatographical method was used). Barnase peaks are labelled with arrows.

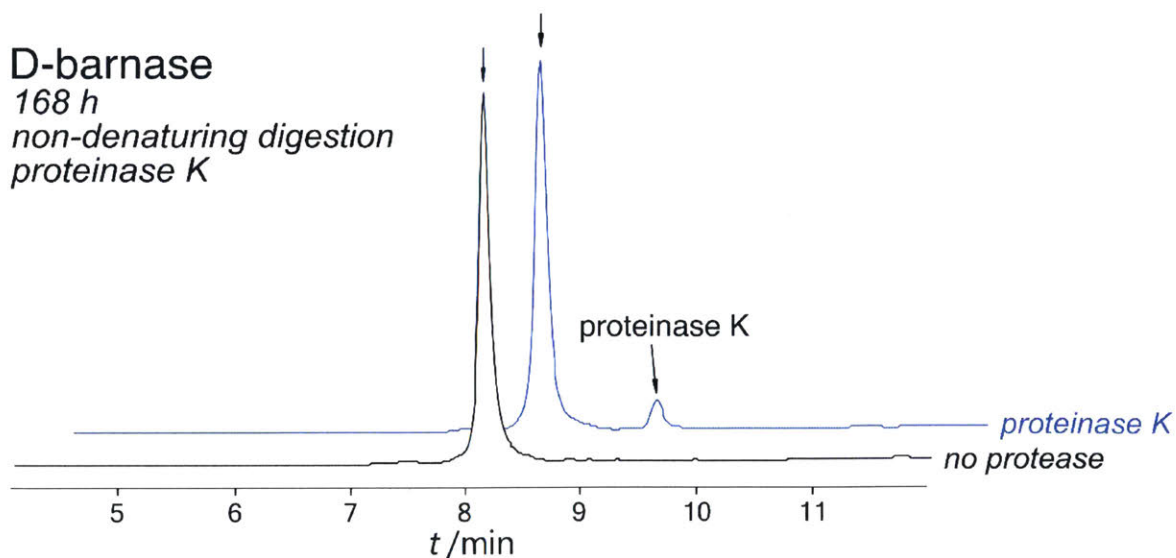


**Figure 2.8. Relative proteolytic stabilities of L- and D-barnase**

Comparison of stabilities of L- and D-barnase towards elastase, papain and actinase E under non-denaturing digestion conditions using HPLC-MS (TIC chromatograms are displayed, 5-65 chromatographical method was used). Barnase peaks are labelled with arrows.



**Figure 2.9. Proteolytic stability of D-barnase towards denaturing proteinase K digestion**  
 Comparison of stabilities of L- and D-barnase towards proteinase K under denaturing digestion conditions using HPLC-MS (TIC chromatograms are displayed). The barnase peak is labelled with an arrow. In the case of L-barnase no barnase peak was identified after 4 hours.



**Figure 2.10. Proteolytic stability of D-barnase towards extended proteinase K digestion**  
 Digestion of D-barnase by proteinase K under non-denaturing digestion conditions for 168 hours. HPLC-MS (TIC) chromatograms are displayed. Barnase peaks are labelled with arrows.

## 2.4. Acknowledgements

We are indebted to Prof. J. Stubbe, Prof. M. Shoulders, and M. Simon for critical discussions and helpful suggestions throughout the course of the work. We thank Prof. E. Nolan for providing some laboratory equipment used to carry out the study, as well as T. Nakashige and H. Chileveru for expert technical assistance and A. Rabideau for help with protein expression. This research was supported by Amgen summer graduate fellowship for A.V. and funded by MIT start-up funds, MIT Reed Fund, Desphande Center Innovation Award, and DARPA (Award #023504-001) for B.L.P. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of DARPA.

## 2.5. References

- (1) Kent, S.; Sohma, Y.; Liu, S.; Bang, D.; Pentelute, B.; Mandal, K. *J. Pept. Sci.* **2012**, *18*, 428–436.
- (2) Zawadzke, L.; Berg, J. *J. Am. Chem. Soc.* **1992**, *114*, 4002–4003.
- (3) Dintzis, H.; Symer, D.; Dintzis, R.; Zawadzke, L.; Berg, J. *Proteins Struct. Funct. Genet.* **1993**, *16*, 306–308.
- (4) Milton, R.; Milton, S.; Kent, S. *Science.* **1990**, *256*, 1445–1448.
- (5) Fitzgerald, M.; Chernushevich, I.; Standing, K.; Kent, S.; Whitman, C. *J. Am. Chem. Soc.* **1995**, *117*, 11075–11080.
- (6) Weinstock, M.; Jacobsen, M.; Kay, M. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 11679–11684.
- (7) Rushizky, G.; Greco, A.; Hartley, R.; Sober, H. *Biochemistry* **1963**, *2*, 787–793.
- (8) Buckle, A.; Fersht, A. *Biochemistry* **1994**, *33*, 1644–1653.
- (9) Mossakowska, D.; Nyberg, K.; Fersht, A. *Biochemistry* **1989**, *28*, 3843–3850.
- (10) Day, A.; Parsonage, D.; Ebel, S.; Brown, T.; Fersht, A. *Biochemistry* **1992**, *31*, 6390–6395.
- (11) Hartley, R.; Barker, E. *Nat. New Biol.* **1972**, *235*, 15–16.
- (12) Johnson, C.; Fersht, A. *Biochemistry* **1995**, *34*, 6795–6804.
- (13) Edelweiss, E.; Balandin, T.; Ivanova, J.; Lutsenko, G.; Leonova, O.; Popenko, V.; Sapozhnikov, A.; Deyev, S. *PLoS One* **2008**, *3*, e2434.
- (14) Mong, S.; Vinogradov, A.; Simon, M.; Pentelute, B. *Chembiochem* **2014**, *15*, 721–733.
- (15) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. a; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A.; Jensen, K.; Pentelute, B. *Chembiochem* **2014**, *15*, 713–720.
- (16) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chem.; Int. Ed.* **2011**, *50*, 7645–7649.
- (17) Wan, Q.; Danishefsky, S. *Angew. Chem.; Int. Ed.* **2007**, *46*, 9248–9252.
- (18) Kelemen, B.; Klink, T.; Behlke, M.; Eubanks, S.; Leland, P.; Raines, R. *Nucleic Acids Res.* **1999**, *27*, 3696–3701.
- (19) Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Taw, D.; Milo, R. *Biochemistry* **2011**, *50*, 4402–4410.
- (20) Li, Y.; Breaker, R. *J. Am. Chem. Soc.* **1999**, *121*, 5364–5372.
- (21) Hartley, R. *Trends Biochem. Sci.* **1989**, *14*, 450–454.
- (22) Cuchillo, C.; Nogués, M.; Raines, R. *Biochemistry* **2011**, *50*, 7835–7841.
- (23) Miller, S.; Simon, R.; Ng, S.; Zuckermann, R.; Kerr, J.; Moos, W. *Drug Dev. Res.* **1995**, *35*, 20–32.

- (24) Tugyi, R.; Uray, K.; Dora, I.; Fellingner, E.; Perkins, A.; Hudecz, F. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 413–418.
- (25) Narayanan, S.; Anwar, R. *Biochem. J.* **1969**, *114*, 11–17.
- (26) de Jersey, J. *Biochemistry* **1970**, *9*, 1761–1767.
- (27) Okorokov, A.; Hartley, R.; Panov, K. *Protein Expr. Purif.* **1994**, *5*, 547–552.





### Chapter 3. C-terminal Modification of Fully Unprotected Peptide Hydrazides via *in situ* Generation of Isocyanates

The work presented in this chapter was published in the following manuscript and is reproduced here with permission of American Chemical Society:

Vinogradov, A.; Simon, M.; Pentelute, B. C-terminal modification of fully unprotected peptide hydrazides via *in situ* generation of isocyanates. *Org. Lett.* **2016**, 18, 1222–1225. DOI: 10.1021/acs.orglett.5b03625

## 3.1. Introduction

Chemo- and regioselective conjugation of bioprobes to peptides and proteins is a key challenge in modern bioorganic chemistry. An ideal conjugation reaction should proceed rapidly in aqueous solvents without protecting groups and accept a wide range of probes and substrates. Several outstanding methods embrace these criteria, allowing excellent chemoselectivity in conjugation chemistry.<sup>1-10</sup> Many of these methods make use of the nucleophilic properties of peptide side chains, in particular cysteine.<sup>11-13</sup> Less common are methods to conjugate nucleophiles to electrophilic peptides.<sup>14-20</sup>

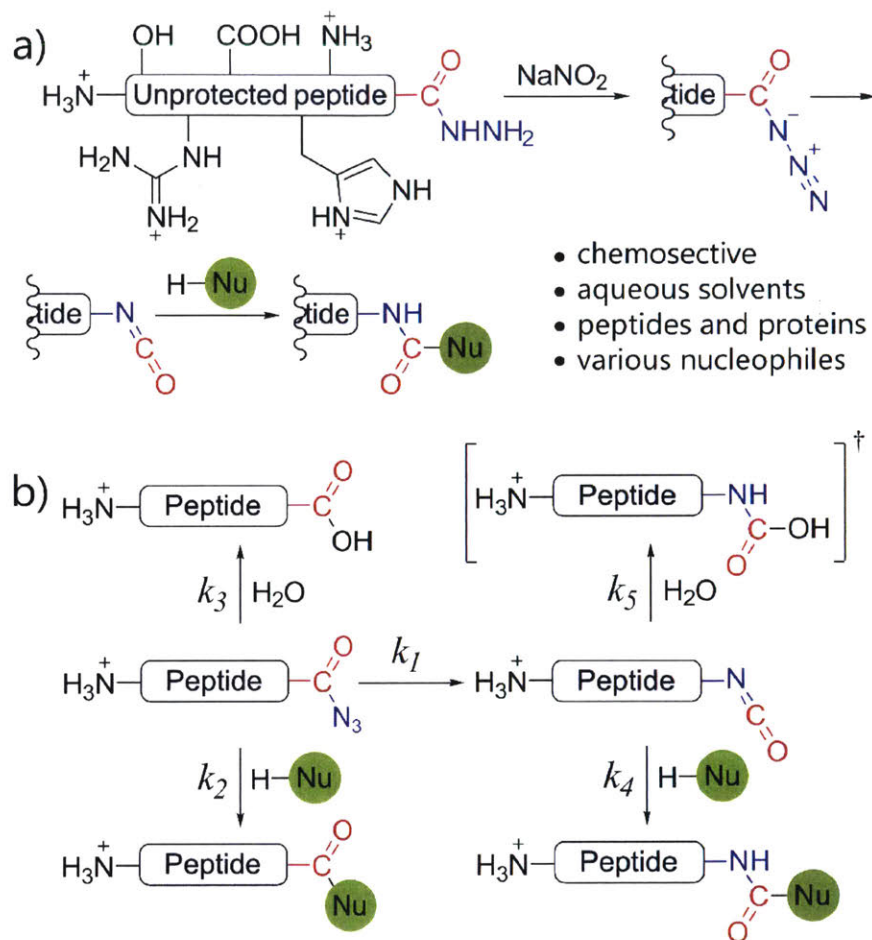
Isocyanates are excellent electrophiles, known to react with a broad array of nucleophilic species.<sup>21</sup> Isocyanates are invaluable precursors in synthesis of peptidomimetics containing urea and carbamate moieties.<sup>22,23</sup> Synthesis and transformations of peptide isocyanates have been reported for short protected fragments in organic solvents, with constituent amino acids often lacking functional side chains.<sup>24-26</sup> Thus far, the utilization of isocyanates as reactive handles for conjugation of nucleophiles to complex unprotected peptides in aqueous media is lacking.

Here we fill this gap for the facile generation of C-terminal electrophilic isocyanates. We found that fully unprotected C-terminal acyl azides, obtained by sodium nitrite oxidation of corresponding hydrazides, undergo a Curtius rearrangement to generate reactive isocyanates (Fig. 3.1a). These species cannot be isolated from aqueous solvents, but can rapidly and selectively react with an excess of various external nucleophiles to give conjugation products. To avoid hydrolysis of isocyanates, nucleophiles are added directly to peptide acyl azides, such that any freshly generated isocyanate reacts with the nucleophile rather than water.

## 3.2. Results and Discussion

### 3.2.1. Investigation of Conjugation Conditions

We first observed these transformations while preparing peptides for native chemical ligation using the recently reported protocol of Liu and co-workers.<sup>27,28</sup> We found that isocyanates are generated and then react with 4-mercaptophenylacetic acid (MPAA), to give carbamothioates as side products (Section 3.3.4 and Chapter 1). To study this side reaction in more detail, we prepared a ten residue model peptide, H<sub>2</sub>N-Ala-Gln-Val-Ile Asn-Thr-Phe-Asp-Gly-Val-CONHNH<sub>2</sub>. In a typical reaction, 2.80 mM peptide in aqueous phosphate buffer at pH 3.2 was oxidized with 20 mM NaNO<sub>2</sub> at -10 °C for 10 minutes and excess MPAA was added to the resulting acyl azide. The reaction was then warmed to room temperature (17 °C) and an aliquot of the crude reaction mixture was analyzed by HPLC-MS two hours later.



**Figure 3.1. *In situ* generation of peptide isocyanates**

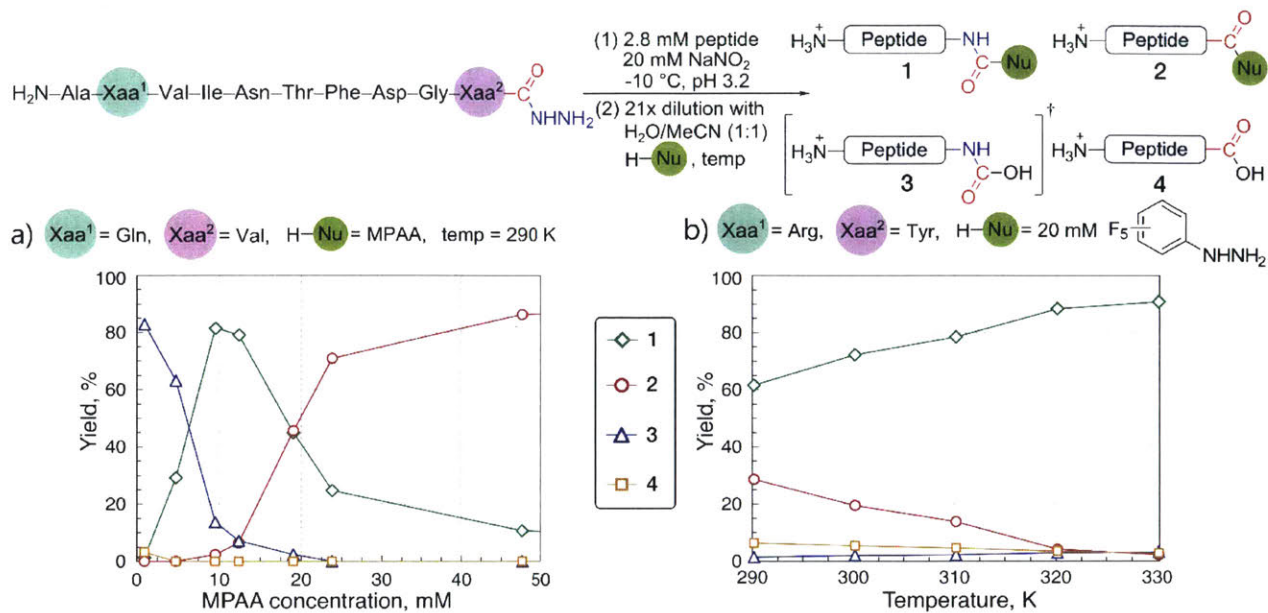
a) One pot oxidation of C-terminal peptide hydrazides followed by Curtius rearrangement and isocyanate conjugation. b) A scheme, illustrating formation of the major products observed throughout the work. †This product was not directly observed, rather it underwent facile decarboxylation during LCMS analysis.

After investigating different reaction conditions, we found that pH, solvent composition, and the order of the reagent addition had modest effect on the formation of the carbamothioate (Section 3.3.5). In contrast, we found that the concentrations of the reagents were crucial for promoting the Curtius rearrangement/isocyanate conjugation. To systematically investigate the effect of nucleophile concentration, we fixed the peptide acyl azide concentration at 133  $\mu\text{M}$  and added varying amounts of MPAA dissolved in water/acetonitrile (1: 1, v/v). We found that the carbamothioate **1** formed in 82% HPLC yield when 10 mM MPAA was used (Fig. 3.2a). We also observed that the further dilution of the MPAA nucleophile led to hydrolysis of isocyanate, while more concentrated nucleophile resulted in thioester formation.

We consistently detected four major products in the reaction mixture: a product of the isocyanate conjugation to nucleophile (Figure 3.2, compound **1**), a product of nucleophile conjugation to azide (**2**), isocyanate derived hydrolysis products (**3**), and carboxylate (azide hydrolysis product, **4**). Analysis of the kinetic scheme in Fig. 3.1b explains the results summarized in Fig. 3.2a: at high concentration of MPAA and/or peptide, the bimolecular reaction with rate constant  $k_2$  takes place leading to the corresponding product **2**. At very low concentrations of both reactants, the monomolecular Curtius rearrangement takes place faster than the direct conjugation leading to **2**, and the pseudo-monomolecular hydrolysis takes place faster than bimolecular conjugation which yields product **3**; this leads to the hydrolysis of the isocyanate observed at low MPAA concentrations. Fortunately, there is a window of concentrations where the Curtius rearrangement is faster than the azide conjugation, and the nucleophile conjugation to the isocyanate is faster than hydrolysis; this leads to a successful conjugation of nucleophiles to isocyanates to give **1**. It is challenging to write exact kinetic relationships based on this simplified idea due to the complex nature of interactions between isocyanates and their nucleophiles.<sup>29</sup> Nevertheless, we found that this analysis holds true for various nucleophiles investigated during the course of the study (see below) and, moreover, with isocyanate concentration fixed, the optimum nucleophile concentration does not change significantly, when different peptide isocyanates are used.

### 3.2.2. Investigation of Conjugation Scope

Next, we explored the nucleophile scope of the reaction with the same model peptide. Isocyanate conjugations were performed under the following standard conditions: 3.07 mM peptide hydrazide solution in oxidation buffer (200 mM  $\text{Na}_2\text{HPO}_4$  and 6 M  $\text{Gn}\cdot\text{HCl}$  in water, pH 3.2) was incubated in an ice-salt bath at  $-10\text{ }^\circ\text{C}$  for 5 min, and then 200 mM  $\text{NaNO}_2$  solution in water was carefully added to the peptide to a final  $\text{NaNO}_2$  concentration of 18 mM. The reaction was allowed to proceed for 10 minutes at  $-10\text{ }^\circ\text{C}$ . Then, a solution of nucleophile in water/acetonitrile (1: 1, v/v, 0.1% TFA added, pH  $\sim$  3.0-4.5) was added to the reaction mixture to a final peptide concentration of 133  $\mu\text{M}$  and the mixture was warmed to the desired temperature (usually 17 or 57  $^\circ\text{C}$ ). Depending on the temperature, the reaction mixture was analyzed using



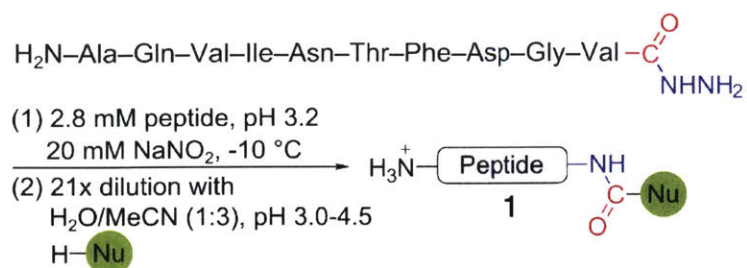
**Figure 3.2. Influence of the parameters on the outcome of the reaction**

a) Distribution of the major reaction products as a function of nucleophile (MPAA) concentration. All data points are in triplicates,  $\pm$  s.d. is smaller than the size of the marker. b) Distribution of the major reaction products as a function of temperature. All data points are single measurements.  $\dagger$ Hydrolysis was accompanied by formation of several products as summarized in Section 3.3.4.

HPLC-MS after 20 minutes (57 °C reaction) or 120 minutes (17 °C reaction). These reaction conditions are referred to as standard conditions for the rest of this chapter. We used strong nucleophiles that remained nucleophilic under acidic conditions and found (Table 3.1) that electron-deficient monosubstituted hydrazines (entries 1, 2) and hydrazides (entries 3, 4) are excellent nucleophiles and react selectively with isocyanates over a wide range of concentrations to give semicarbazides with high yield and selectivity. Using a large excess of these nucleophiles ensured that any remaining sodium nitrite is quenched and is inconsequential to the reaction progress. Additionally, we found that hydroxylamine derivatives (entries 5, 6) also conjugate to isocyanates albeit at higher concentrations and with yields lower than hydrazines. We also demonstrated that the dilution of the acyl azide with primary alcohols affords corresponding carbamates in high yield (entry 8). Various alkylthiols ( $\beta$ -mercaptoethane sulfonate,  $\beta$ -mercaptoethanol, and *tert*-butylthiol) failed to conjugate to peptide isocyanates under a number of conditions. Notably, in this study we did not observe products of intramolecular cyclizations, or formation of symmetrical ureas, a prominent side reaction in the acyl azide method of peptide synthesis.<sup>30,31</sup>

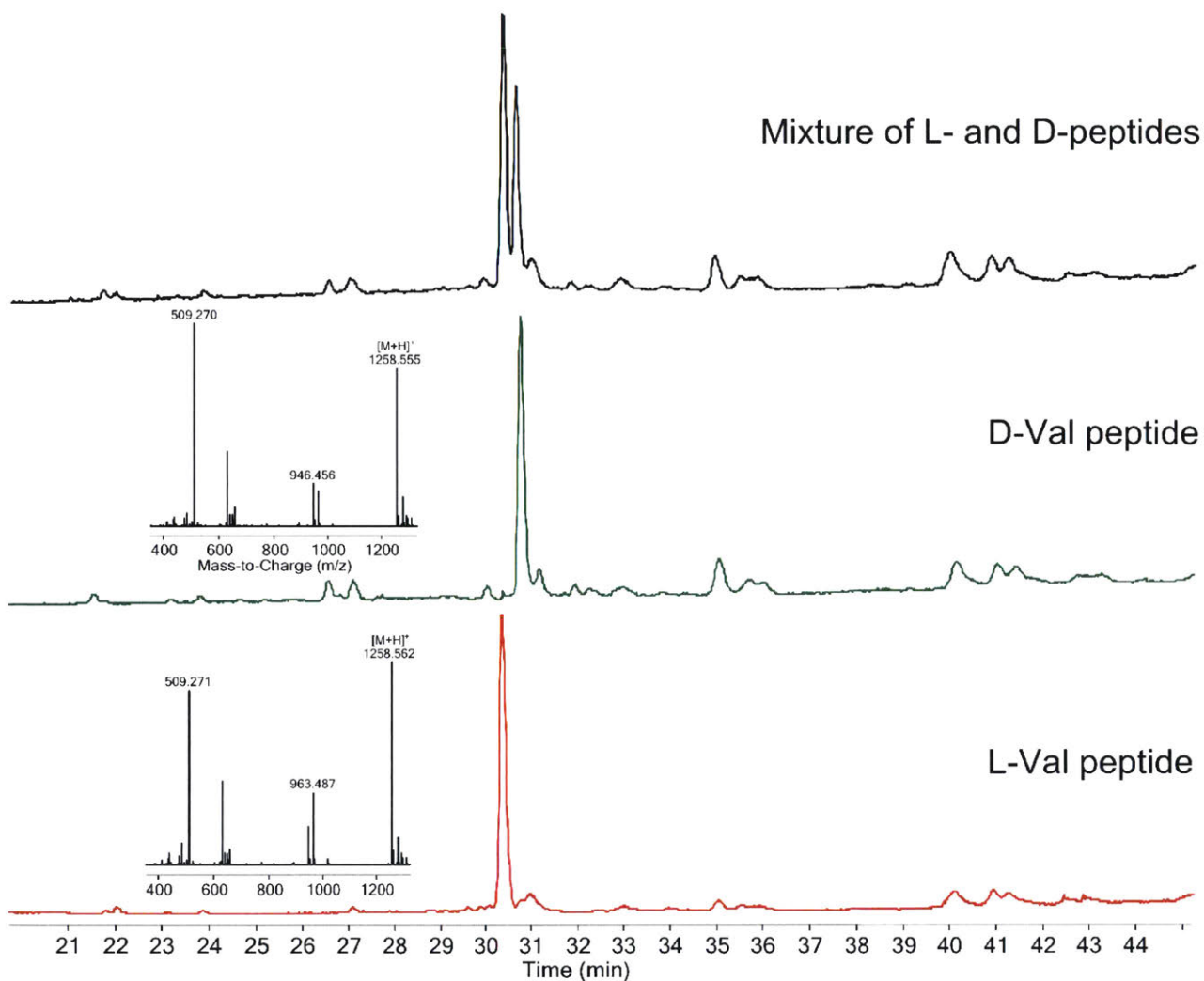
The side-chain of the C-terminal amino acid is reported to significantly affect the rate of the Curtius rearrangement. More sterically hindered amino acids, such as valine and isoleucine accelerate the rearrangement, while less sterically hindered glycine and proline significantly decelerate it.<sup>32</sup> Indeed, when we oxidized the model peptide H<sub>2</sub>N-Ala-Arg-Val-Ile-Asn-Thr-Phe-Asp-Gly-Tyr-CONHNH<sub>2</sub> under standard conditions and used 20 mM perfluorophenylhydrazine in water/acetonitrile (1: 1, v/v) as a nucleophile, we observed a modest 62% yield of semicarbazide, with N<sup>1</sup>-perfluorophenyl hydrazide being the side product (29% yield, Fig 3.2b). To accelerate the Curtius rearrangement and increase the selectivity of the transformation, we performed the reaction at higher temperatures, keeping the temperature of the oxidation step at -10 °C. We were delighted to find that the yield of semicarbazide rose almost linearly with increasing temperature, reaching 91% at 57 °C. Additionally, performing the reaction at 57 °C allowed us to reduce the overall reaction time to 30 minutes (a 10 minute oxidation step followed by a 20 minute rearrangement). Although the Curtius rearrangement is known to proceed with retention of configuration,<sup>33</sup> we confirmed that our conjugation does not lead to the racemization of the C-terminal residue in a separate experiment using two diastereomeric peptides: H<sub>2</sub>N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-**Val<sup>L</sup>**-CONHNH<sub>2</sub> and H<sub>2</sub>N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-**Val<sup>D</sup>**-CONHNH<sub>2</sub>. Both reactions were performed under standard conditions using perfluorophenylhydrazine as a nucleophile at 17 °C. To resolve diastereomeric semicarbazides (conjugation products) we performed HPLC-MS analysis with the following gradient: 5% B' in A' for 2 min, 5-65% B' in A' ramping linearly over 70 min, 65% B' in A' for 1 minute. Because diastereomers were resolved on RP-HPLC, and because each reaction led to the formation of a single, chromatographically resolved diastereomer product (Fig. 3.3), we concluded that the whole process is indeed racemization-free.

**Table 3.1. Nucleophile scope of the reaction**



entry	H—Nu	[H—Nu], mM	yield of <b>1</b> , % <sup>a</sup>
1		25	96 (82)
2		5.7 <sup>b</sup>	73 (48)
3		33	96 (73)
4		33	92 (69)
5		100	72 (53)
6	$\text{CH}_3\text{ONH}_2$	100	79 (42)
7		10	82 (60)
8	MeOH	95% (v/v)	89 (66)

<sup>a</sup>HPLC yields are listed. Isolated yields are provided in parentheses. <sup>b</sup>The nucleophile concentration was limited by its solubility.



**Figure 3.3. HPLC-MS analysis of two diastereomeric conjugation products**

HPLC-MS (TIC) chromatograms for diastereomeric  $\text{H}_2\text{N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-}^{\text{L}}\text{Val-CONHNH}_2$  and  $\text{H}_2\text{N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-}^{\text{D}}\text{Val-CONHNH}_2$  oxidation/conjugation reaction with MS insets for main products (semicarbazides).

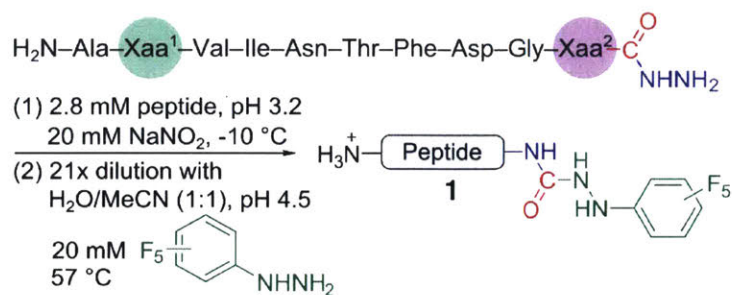


To further elucidate the effects of the C-terminal residue on the efficiency of the isocyanate conjugation we prepared a small library of peptides  $\text{H}_2\text{N-Ala-Xaa}_1\text{-Val-Ile-Asn-Thr-Phe-Asp-Gly-Xaa}_2\text{-CONHNH}_2$ , where all 20 proteogenic amino acids were scanned in the position  $\text{Xaa}_2$ .<sup>34,35</sup> Amino acids in the position  $\text{Xaa}_1$  were additionally varied to test the chemoselectivity of the transformations in the presence of potentially reactive peptide side chains. All reactions were carried out under standard conditions. With the exception of  $\text{Xaa}_2 = \text{Val, Ile, and Thr}$ , which were reacted at 17 °C, all peptides were incubated with the nucleophile at 57 °C for 20 minutes.

As summarized in Table 3.2, we found that the majority of C-terminal amino acids were compatible with the studied chemistry, and in all cases the isocyanate conjugation was compatible with residues in  $\text{Xaa}_1$  position; we did not observe products of the isocyanate cyclization to the  $\text{Xaa}_1$ . Peptides with C terminal hydrophobic residues gave corresponding semicarbazides in good yields, with  $\text{Xaa}_2 = \text{Gly}$  (entry 13) as the predictable exception. Interestingly, despite the fact that isocyanates are known to react with aromatic nucleophiles,<sup>21</sup> we found that C-terminal Tyr and Trp (entries 3 and 4) did not form cyclization products under the studied conditions and instead yielded the desired semicarbazides. Also compatible with the one-pot isocyanate conjugation were C-terminal Arg, Lys, Met, and Cys (entries 5, 6, 8, 11). For Cys the reaction was accompanied by the formation of a product +29 Da heavier than expected, which we attributed to the nitrothioite (RSNO) formation. Fortunately, nitrothioites can be reduced back to thiols.<sup>36</sup>

Several C-terminal amino acids were incompatible with the reported reaction. The side chains of Thr and Ser both cyclized with the isocyanate, which led to the formation of oxazolidin-2-ones (Section 3.3.6).<sup>37</sup> C-terminal His and Glu also led to poor yields of semicarbazides, giving significant amounts of carboxylate. This result may be explained by the cyclization of His and Glu side chains with the acyl azide, followed by the hydrolysis of the cyclic intermediate to release carboxylate.<sup>38,39</sup> Formation of the cyclic intermediate is supported by the fact that two chromatographically resolved peaks corresponding to peptide thioester were observed when  $\text{H}_2\text{N-Ala-Leu-Val-Ile-Asn-Thr-Phe-Asp-Gly-Glu-CONHNH}_2$  was subject to oxidation/thioesterification with 200 mM MPAA (Section 3.3.6). Consistent with previous reports,<sup>27</sup> we could not produce C-terminal hydrazides of Asn, Asp, and Gln due to the intramolecular cyclization that spontaneously occurs after the peptide is released from resin.

**Table 3.2. C-terminal amino acid scope of the reaction**

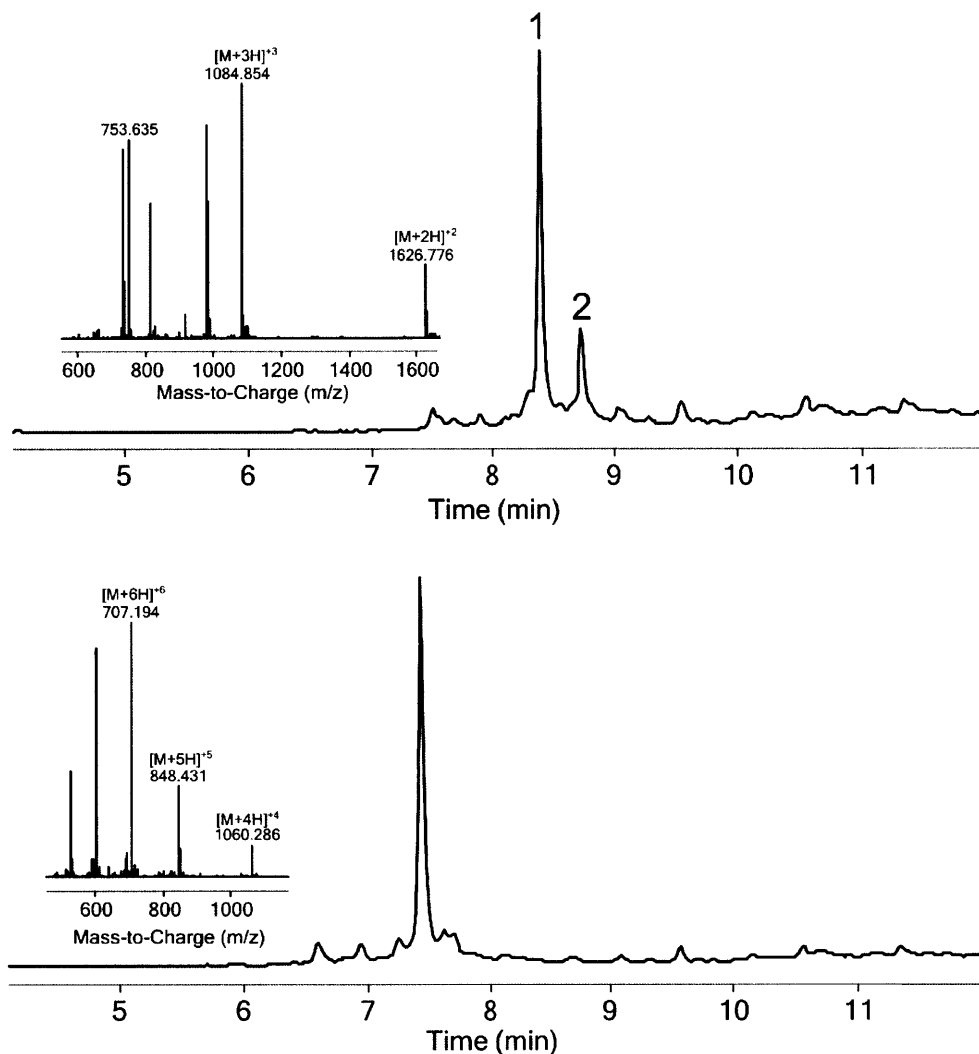


entry	Xaa <sup>1</sup>	Xaa <sup>2</sup>	HPLC yield of <b>1</b> , %
1	Gln	Val	96 <sup>a</sup>
2	Lys	Ile	95 <sup>a</sup>
3	Arg	Tyr	91
4	Lys	Trp	89
5	Trp	Arg	85
6	Tyr	Lys	84
7	Asp	Leu	84
8	Gly	Met	83
9	Glu	Ala	72
10	Lys	Pro	62
11	Pro	Cys	57
12	Met	Phe	51
13	Ser	Gly	44
14	Leu	Glu	24
15	Val	His	11
16	Asn	Ser	n. d.
17	His	Thr	n. d. <sup>a</sup>
18	Any	Asp, Asn, Gln	n. a. <sup>b</sup>

<sup>a</sup>Reaction was performed at 17 °C. <sup>b</sup>Fmoc-SPPS failed to produce corresponding hydrazides. n.d. = not determined, n. a. = not applicable.

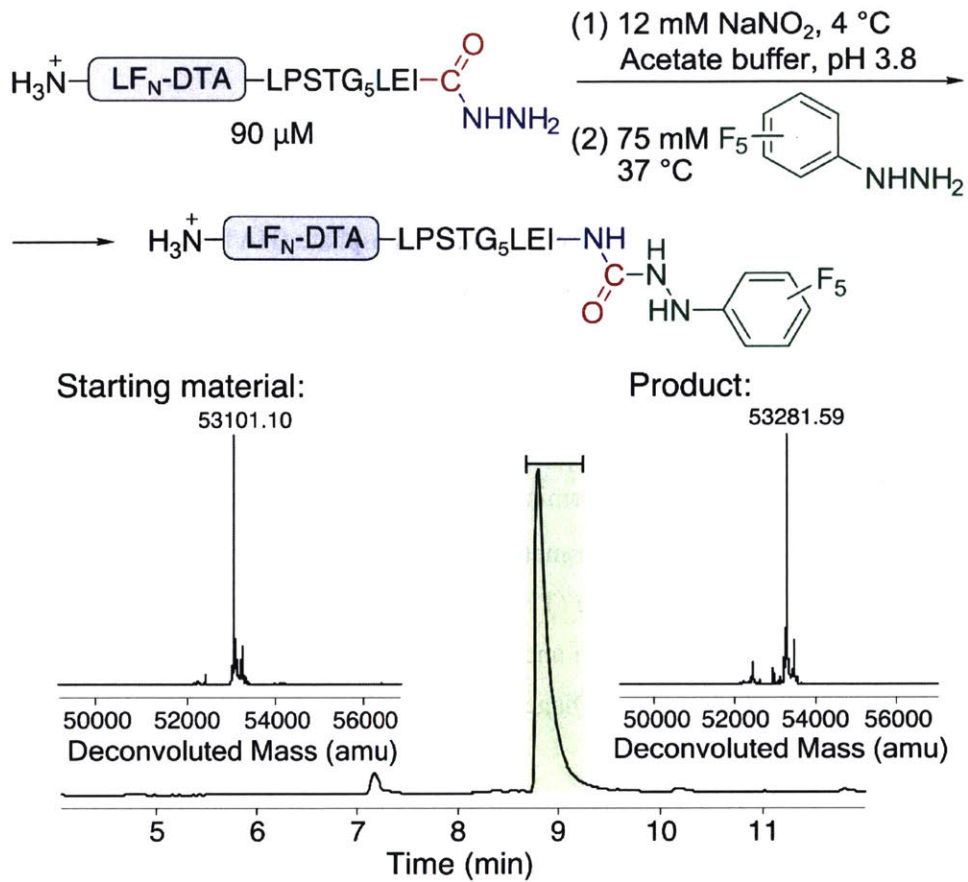
Finally, as a proof-of-concept, we demonstrated that the described chemistry can be used to conjugate nucleophiles to longer peptides and proteins. To this end, we successfully conjugated perfluorophenylhydrazine to 26-mer H<sub>2</sub>N-CDYLQTYHKLPDNYITKSEAQALGWV-CONHNH<sub>2</sub> and 37-mer H<sub>2</sub>N-(Acm)CSKGNLADVAPGKSIGGDIFSNREGKLPGKSGRTWRE(Cy)-CONHNH<sub>2</sub> peptides under standard conditions (Fig. 3.4). For protein conjugation, we installed a C-terminal hydrazide on a 53 kDa protein (LF<sub>N</sub>-DTA-Leu-Pro-Ser-Thr-Gly-Gly-His<sub>5</sub>) by means of Sortase A\* mediated ligation with Gly<sub>5</sub>-Leu-Glu-Ile-CONHNH<sub>2</sub>.<sup>40,41</sup> 12 mM NaNO<sub>2</sub> in water was added to 90 μM LF<sub>N</sub>-DTA-Leu-Pro-Ser-Thr-Gly<sub>5</sub>-Leu-Glu-Ile-CONHNH<sub>2</sub> in 50 mM acetate buffer at pH 3.8; after 10 minutes at 4 °C the protein acyl azide was diluted to 22.5 μM with 75 mM perfluorophenylhydrazine in the same buffer (pH 3.8), and the resulting solution was incubated for 1 hour at 37 °C. These conditions cleanly afforded perfluorophenyl-labeled C-terminal semicarbazide of LF<sub>N</sub>-DTA protein in 92% yield as estimated by the integration of the deconvolution spectrum (Fig. 3.5).

In summary, we have developed a method for the conjugation of nucleophiles to peptides and proteins, making use of C-terminal isocyanates generated in one pot from corresponding hydrazides. The reaction occurs in aqueous solvents; it is chemo- and regioselective, and it is orthogonal to unprotected amino acid side chains. As C-terminal hydrazides are readily available through standard Fmoc-SPPS, there is no need to preinstall electrophilic moieties during SPPS. The conjugation results in formation of peptidomimetic bonds, such as carbamates, semicarbazides, and carbamothioates, and is essentially irreversible. Optimal concentrations of reagents are crucial for the success of conjugation (Fig. 3.1b): various nucleophiles in 5-100 mM range yielded the desired products.



**Figure 3.4. Perfluorophenylhydrazine conjugation to longer peptides**

HPLC-MS (TIC) chromatograms for oxidation/conjugation reactions with the barnase fragments  $H_2N$ - $^{14}Cys$ - $^{39}Val$ -CONH $NH_2$  (26 residues, top) and  $H_2N$ - $^{40}Cys(Acm)$ - $^{76}Glu(Cy)$ -CONH $NH_2$  (37 residues, bottom). Top: peak 1 corresponds to the expected semicarbazide conjugation product with MS inset of the peak apex displayed on the left (calc. monoisotopic mass = 3250.5 Da). Peak 2 is an intermolecular disulfide of the conjugation product. Bottom: the main peak is the expected conjugation product with the MS inset of the peak apex on the left. (calc. monoisotopic mass = 4235.0 Da)



**Figure 3.5. C-terminal labeling of 53 kDa protein with perfluorophenylhydrazine**

HPLC-MS (TIC) chromatogram of the crude labeled protein is shown. Highlighted MS peak was deconvoluted (maximum entropy) to obtain the product deconvolution spectrum on the right. The starting material spectrum was obtained the same way (TIC not shown).

### 3.3. Experimental

#### 3.3.1. General

2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and N $\alpha$ -Fmoc protected D-amino acids were purchased from Chem-Impex International. N $\alpha$ -Fmoc protected L-amino acids were obtained through Advanced ChemTech, and Novabiochem.

2-Chlorotrityl chloride resin (200-400mesh, 1.2 mmol/g) was purchased from Chem-Impex International and was subsequently used to prepare 2-chlorotrityl hydrazine resin adhering to the described protocols.<sup>42</sup> The loading of the resin was determined to be 0.75 mmol/g.

*N,N*-Dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile were from EMD Millipore. Triisopropyl silane (TIPS), and 1,2-ethanedithiol (EDT) were from Alfa Aesar. Solvents for HPLC-MS were purchased from and Fluka. All other reagents were purchased from Sigma-Aldrich. Ni-NTA Agarose beads were from Qiagen.

#### 3.3.2. Peptide Synthesis

*Fmoc Solid Phase Peptide Synthesis.* Peptides were synthesized on our recently reported automated flow-based platform using a 2<sup>nd</sup> generation vessel as described in Chapter 1.<sup>34</sup> During synthesis all solvents and reagents were preheated to 60°C immediately before reaching the synthesis vessel. The following standard 94 second cycle was used to assemble all peptides, unless noted: amide bond formation (coupling) was performed for 14 s at 23 mL/min. Removal of coupling reagent (wash) was done in 30 s at 20 mL/min. N $\alpha$ -Fmoc removal (deprotection) was carried out for 20 s at 20 mL/min, and finally, the removal of deprotection reagent and products (second wash) was achieved in 30 s at 20 mL/min.

*Coupling.* Coupling was performed by delivering the following coupling solution at 20 mL/min for 14 s. The coupling solution consisted of N $\alpha$ -Fmoc and side chain protected amino acid dissolved in equimolar 0.4 M HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) in DMF. To activate the amino acid, DIEA (diisopropylethylamine) was injected at 4 ml/min. The mixed fluid then passed through a static mixer and a preheat loop to the synthesis vessel. In order to minimize racemization of Cys and His residues, only 2 mL/min of DIEA were injected when activating these amino acids. In all cases, this coupling solution contained at least four equivalents of activated amino acid with respect to the resin.

To attach C-terminal Val and Ile residues to the resin, 1 mmol of protected amino acid dissolved in 2.5 mL of 0.4 M HBTU in DMF was manually delivered to the synthesis vessel, then synthesis was paused for

10 min. An increased reaction time was necessary to quantitatively couple these sterically hindered residues to the resin.

Side chain protection was as follows: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Trt), Gln(Trt), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu).

*Wash.* Excess reagent and reaction by-products were washed out from the synthesis vessel with 10 mL of DMF delivered at 20 mL/min over 30 s.

*Deprotection.* N<sup>α</sup>-Fmoc protecting groups were removed with 6.6 mL of 20% (v/v) piperidine in DMF delivered at 20 mL/min over 20 s.

Peptide H<sub>2</sub>N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONHNH<sub>2</sub> was additionally synthesized on a larger scale using standard batch Fmoc SPPS procedures: 2.5 g of chlorotriyl hydrazide resin (0.25 mmol/g) was pre-swollen in DCM for 30 min, then drained. To attach C-terminal Val to the resin, 5 mmol of Fmoc protected valine was dissolved in 12.5 mL of 0.38 M HATU in DMF and 1.5 mL of DIEA, then added to the swollen resin. After 1.5 hours, the resin was drained and a second 5 mmol aliquot of Fmoc-Val-OH, prepared as above, was added. After a second 1.5 hour coupling, the resin was drained and washed 3 x 30 mL of DMF. Fmoc was removed by treatment with 20 mL of 20% piperidine in DMF for 20 min, and the solution was drained and the resin was washed with DMF 3 x 30 mL. Synthesis then proceeded according to the following procedure: 5 mmol of each protected amino acid was dissolved in 12.5 mL of 0.38 M HATU in DMF and 1.5 mL of DIEA, then added to the resin. After a 30 min coupling step, the resin was drained and washed with DMF three times (30 mL each). Fmoc was removed by treatment with 30 mL of 20% piperidine in DMF for 20 minutes, and the resin was again drained and washed as above. After the final Fmoc removal, the resin was washed 4 x 40 mL of DCM each, and dried *in vacuo*. All other peptides mentioned herein were synthesized and purified as previously described.<sup>35</sup>

*Cleavage of peptides from resin.* All peptides were cleaved from the resin and side chain deprotected with a standard cleavage cocktail of 2.5% (v/v) EDT, 2.5% (v/v) H<sub>2</sub>O, and 1% (v/v) TIPS in TFA for 7 minutes at 60 °C. In all cases, compressed nitrogen was used to evaporate the cleavage solution to dryness together with resin. The resulting mixtures were triturated and washed three times with cold diethyl ether, and, unless noted, dissolved in 95% water / 5% acetonitrile (0.1% TFA added). Resin was filtered and solutions were purified without intermediate lyophilization.

*Purification of Peptides.* All peptides were purified on a Waters 600 HPLC system with a Waters 484 or 486 UV detector using water with 0.1% TFA added (solvent A) and acetonitrile with 0.1% TFA added (solvent B) as solvents.

An Agilent Zorbax 300SB preparative C<sub>3</sub> column (300Å, 7µm, 21.2 x 250mm) was used at 20 mL/min linear flow rate for the purification of peptide hydrazides. Unless noted, the following gradient was used to purify peptide hydrazides: 5% B in A for 5 minutes, then 5%-35% B ramping linearly over 90 minutes.

An Agilent Zorbax 300SB semi-preparative C<sub>3</sub> column (300 Å, 5 µm, 9.4 x 250 mm) was used at 5 mL/min linear flow rate for the purification of isocyanate conjugation reactions. Unless noted, the following gradient was used to purify isocyanate conjugation reactions: 5% B in A for 10 minutes; then 5%-15% B ramping linearly over 10 minutes; then 15%-45% B ramping linearly over 90 minutes. Fractions were collected and screened for the desired material using HPLC-ESI-TOF as described below.

### 3.3.3. HPLC-MS Analysis

Unless noted, all peptides, proteins and reaction mixtures were analyzed on an Agilent 6520 Accurate Mass Q-TOF LC-MS using an Agilent Zorbax 300SB C<sub>3</sub> column (300 Å, 5 µm, 2.1 x 150 mm) run with the following method. At 40 °C and a flow rate of 0.8 mL/min, the following gradient was used: 1% acetonitrile with 0.1% formic acid (FA, solvent B') in water with 0.1% FA (solvent A') for 2 min, 1-61% B' in A' ramping linearly over 11 min, 61% B' in A' for 1 minute. Typically, 100 – 1000 µg/mL solutions of peptides and proteins were subject to analysis. More concentrated solutions were diluted to appropriate concentrations with 50% A/50% B. Unless noted, all chromatograms shown in this work are plots of total ion current (TIC) versus time. HPLC yields of reactions were calculated via manual integration of the area under peaks corresponding to reaction products on TIC vs. time chromatograms. Generally, TIC and UV<sub>214</sub> integration gave converging results, consistent with previous reports;<sup>43</sup> TIC integration was a preferred method of characterization for practical reasons. The data were analyzed using Agilent MassHunter Qualitative analysis software. All MS deconvolution spectra were obtained using the maximum entropy algorithm.

### 3.3.4. Discovery of Isocyanate Generation

Some data shown in this section are a more detailed reproduction of the data mentioned in Chapter 1 and in Mong *et al.*<sup>35</sup> describing the total synthesis of barnase. Originally, during oxidation/thioesterification of barnase fragments, we consistently observed formation of MPAA thioester, contaminated with a side-product 15.0 Da more massive than thioester. It was observed for two different fragments.

In particular, during these studies we adhered to the standard oxidation/thioesterification protocol described by Fang *et al.*<sup>27</sup> Thus, 2.00 mg (1.47 µmol) of H<sub>2</sub>N-Gly<sub>5</sub>-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONHNH<sub>2</sub> were dissolved in 0.48 mL of oxidation buffer (200 mM Na<sub>2</sub>HPO<sub>4</sub> and 6 M Gn·HCl in water, pH 3.2). The solution was incubated in an ice-salt bath at -18°C for 10 min and then 48 µL of 200 mM NaNO<sub>2</sub> solution in water was added to it dropwise while stirring. After 20 min an analytical sample

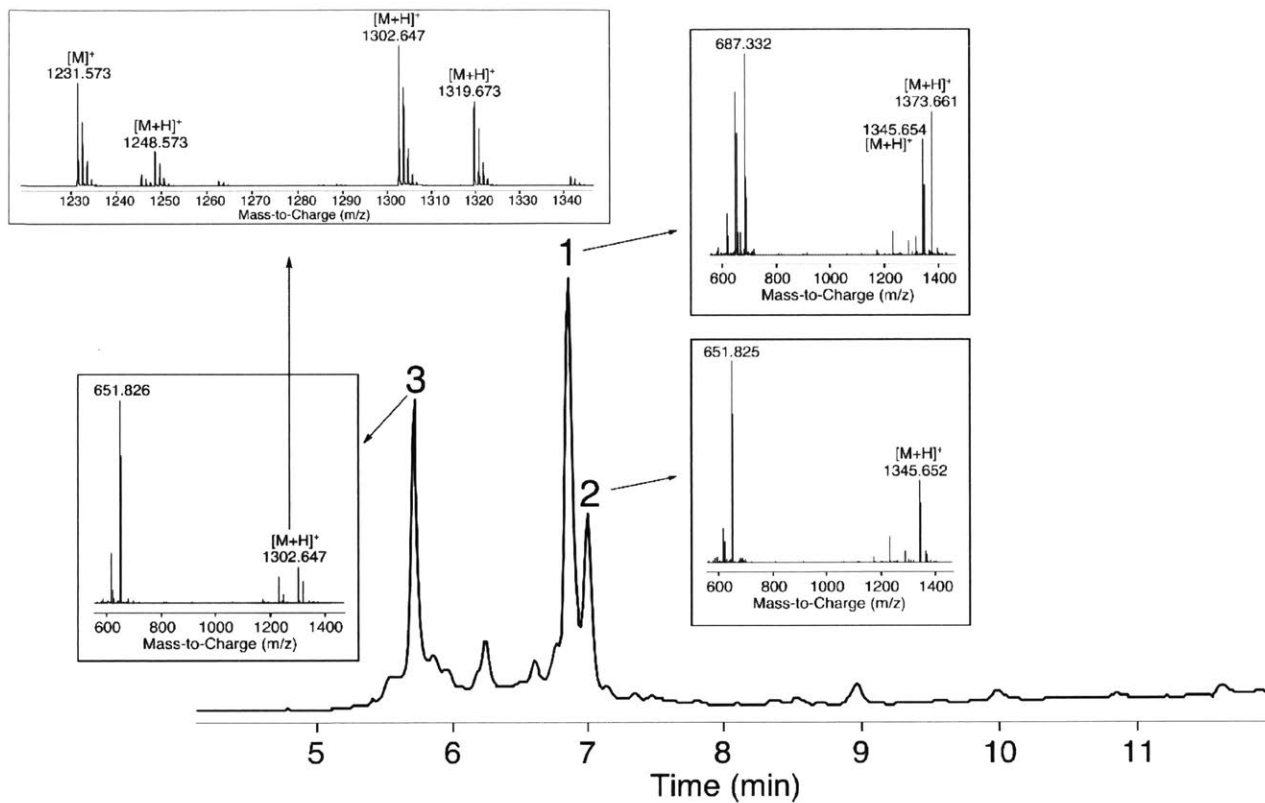


for LC-MS was taken (Fig. 3.6) and then 0.48 mL of 0.2 M MPAA in ligation buffer (200 mM Na<sub>2</sub>HPO<sub>4</sub> and 6 M Gn·HCl in water, pH 6.8) were added to the reaction. Upon complete addition of the MPAA solution, another analytical sample for the LC-MS was taken (Fig. 3.7).

The data presented in Fig. 3.6 suggest that the hydrazide is successfully oxidized to form the acyl azide (peak 1). Two sets of ions (one is -28 Da relative to another) for peak 1 may be explained by the fragmentation of azide in the mass spectrum. Indeed, this behavior is quite common for azides.<sup>44</sup> In contrast, peak 2 is chromatographically resolved from the azide peak and there is no parent azide ion, which allowed us to conclude that peak 2 is a real compound 28 Da lighter than the azide. The mass difference of 28 Da corresponds to a loss of a nitrogen molecule, which is consistent with the hypothesis that the acyl azide undergoes a Curtius rearrangement to release nitrogen and form an isocyanate. Analysis of peak 3 further supports this conclusion. The four major ions observed in the MS of the apex of peak 3 are consistent with isocyanate hydrolysis, which may be explained by the degradation of the C-terminal amino acid, which is accompanied by many intermediates, most of which can be observed in the MS (Fig. 3.8).

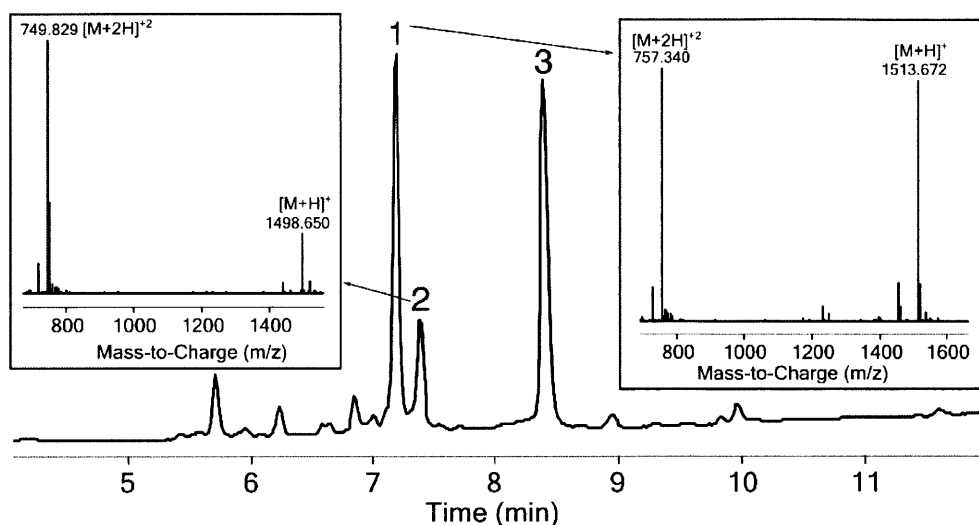
We observed similar ions series for most peptides in this study. Because in most cases we observed several products stemming from isocyanate hydrolysis, we refer to them collectively as ‘isocyanate hydrolysis products’ in this chapter.

Figure 3.7 shows the crude reaction mixture after addition of MPAA. Formation of peak 2 (+15 Da relative to the expected thioester) is consistent with the hypothesis that the acyl azide rearranges to isocyanate, because isocyanate may react with MPAA to give carbamothioate, which is 15 Da heavier than thioester.



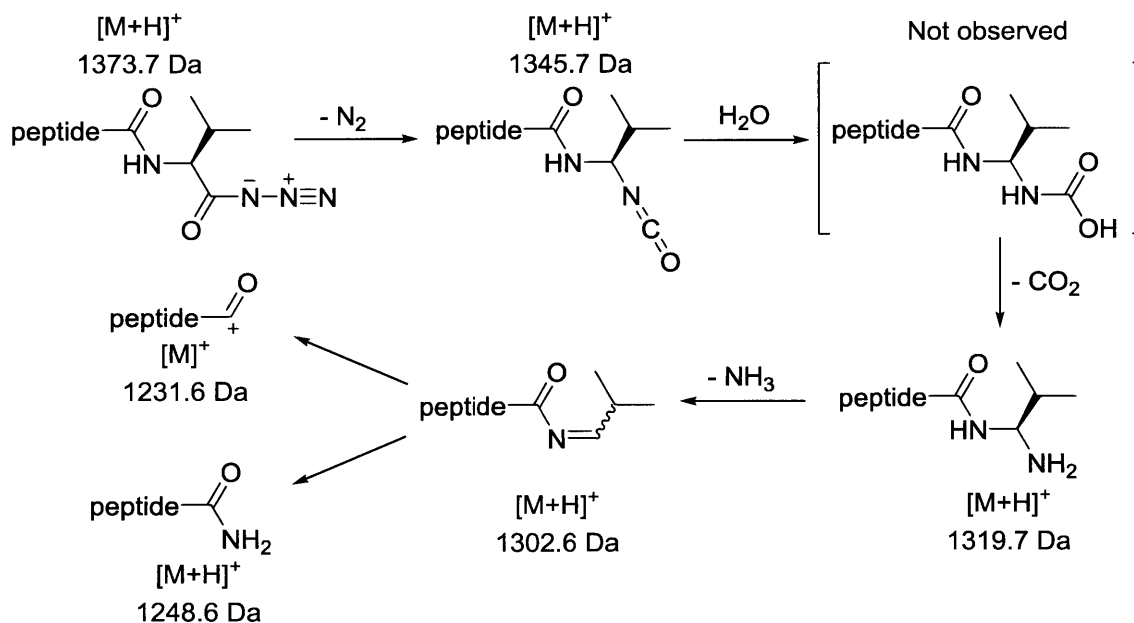
**Figure 3.6. HPLC-MS analysis of the isocyanate formation**

HPLC-MS (TIC) chromatogram for Gly5-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONHNH<sub>2</sub> oxidation reaction with MS insets for main products. Peak 1 corresponds to the expected acyl azide product (calc. monoisotopic mass = 1372.7 Da). The chromatogram shows the formation of isocyanate (peak 2 with MS inset of the peak apex on the bottom right) during hydrazide oxidation in the studied system. Peak 3 with MS insets on the left corresponds to isocyanates hydrolysis products.



**Figure 3.7. HPLC-MS analysis of a crude conjugation reaction with MPAA**

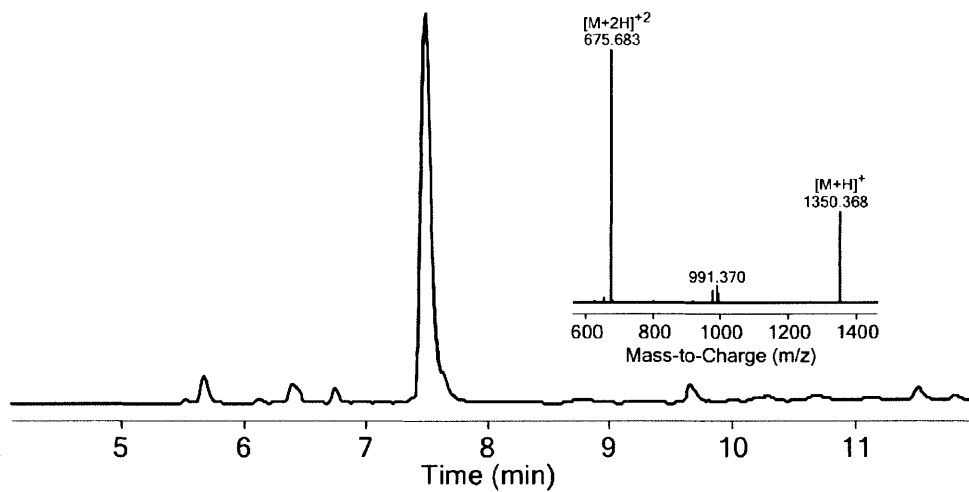
HPLC-MS (TIC) chromatogram for  $\text{H}_2\text{N-Gly}_5\text{-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONHNH}_2$  oxidation/transesterification reaction with MS insets for main products. Peak 2 corresponds to the expected thioester product (calc. monoisotopic mass = 1498.6 Da). Peak 1 is the unexpected carbamothioate formed from the corresponding isocyanate. Peak 3 is MPAA disulfide.



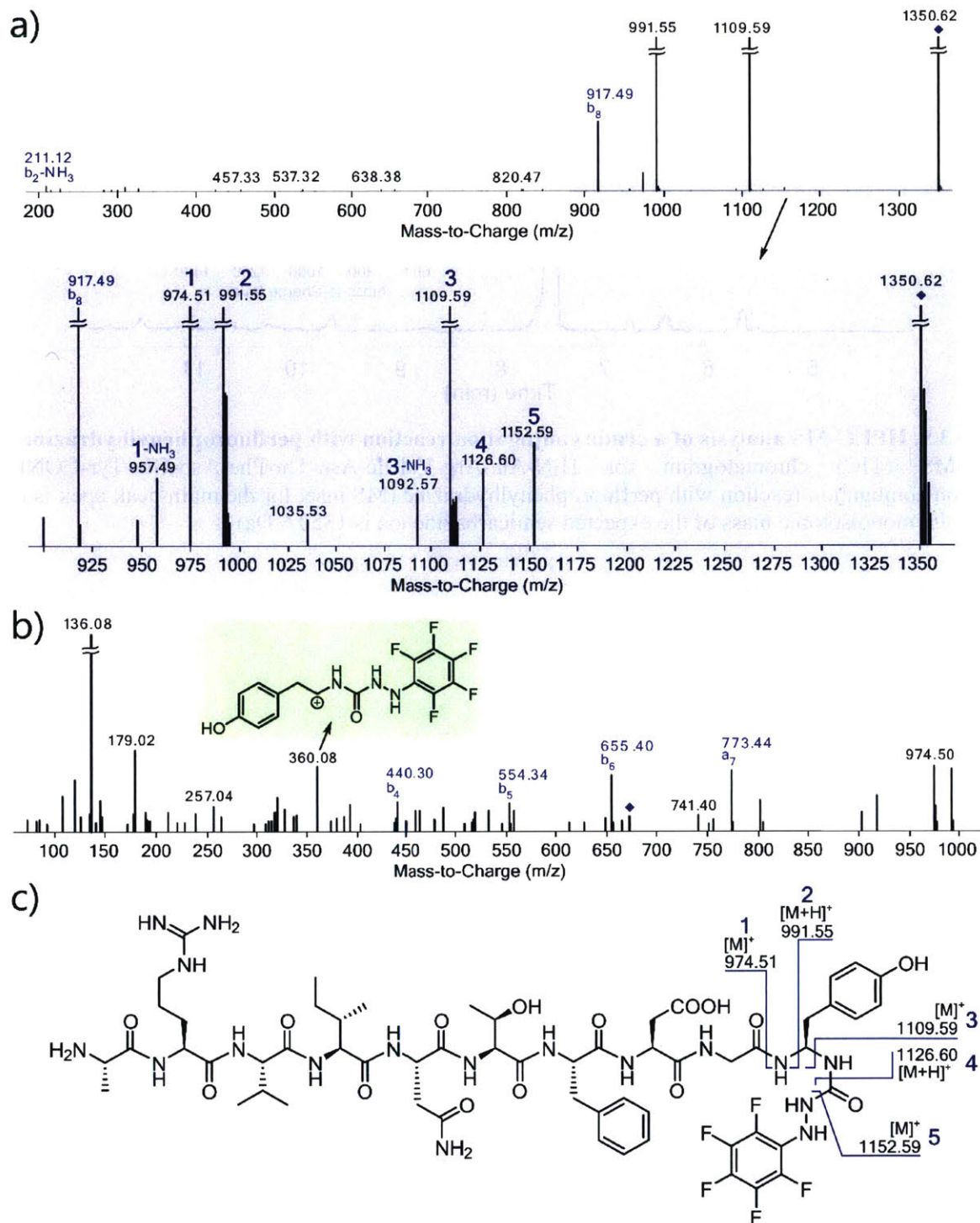
**Figure 3.8. Transformations of the C-terminal isocyanate in water**

To exclude the possibility of side chains reacting with sodium nitrite or MPAA, we prepared the peptide H<sub>2</sub>N-Gly<sub>5</sub>-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Val-COOH and subjected it to the reaction conditions described above. The peptide was unreactive under these conditions, suggesting that the +15 Da product indeed stems from the C-terminal hydrazide oxidation.

Finally, we confirmed the structure of the conjugation product using MS/MS analysis. To this end, we performed oxidation/conjugation of a model peptide H<sub>2</sub>N-Ala-Arg-Val-Ile-Asn-Thr-Phe-Asp-Gly-Tyr-CONHNH<sub>2</sub> under standard conditions with 15 mM perfluorophenylhydrazine as a nucleophile at 57 °C. HPLC-MS chromatogram for the crude reaction mixture shown in Figure 3.9 confirmed that the nucleophile was conjugated to the peptide. We then performed the secondary MS analysis of the main product [M+H]<sup>+</sup> and [M+2H]<sup>+2</sup> ions. As Figure 3.10 shows, the fragmentation patterns of both ions match the expected structure well. These data unambiguously indicate that perfluorophenylhydrazine conjugated to the C-terminus of the peptide. More importantly, we observed the extensive fragmentation of the C-terminal moiety, which confirmed the presence of the semicarbazide functional group. Thus, we concluded that C-terminal acyl azides rearrange to isocyanates, which react with external nucleophiles to give conjugation products.



**Figure 3.9. HPLC-MS analysis of a crude conjugation reaction with perfluorophenylhydrazine**  
 HPLC-MS (TIC) chromatogram for H<sub>2</sub>N-Ala-Arg-Val-Ile-Asn-Thr-Phe-Asp-Gly-Tyr-CONHNH<sub>2</sub>  
 oxidation/conjugation reaction with perfluorophenylhydrazine (MS inset for the main peak apex is on the  
 right; calc. monoisotopic mass of the expected semicarbazide ion is 1350.6 Da)



**Figure 3.10. MS/MS analysis of the synthesized semicarbazide**

a) Top: MS/MS chromatogram for  $[M+H]^+$  ion of the conjugation product. The parent ion is labeled as diamond. Bottom: zoom-in of the 900-1350 m/z region of the spectrum on the top. Peak labels correspond to fragments shown in panel c). b) MS/MS chromatogram for  $[M+2H]^{2+}$  ion of the conjugation product. The parent ion is labeled as diamond. Some  $a_n$  and  $b_n$  fragments are labeled. The C-terminal fragment containing semicarbazide was also identified and labeled. c) Structural formula of the conjugation product indicating the most important fragments.

### 3.3.5. Optimization of Conjugation Conditions

To optimize reaction conditions we chose H<sub>2</sub>N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONHNH<sub>2</sub> as a model peptide and MPAA as a nucleophile. 4.14 mg (3.85 μmol) of peptide hydrazide were dissolved in 1.25 mL of oxidation buffer (200 mM Na<sub>2</sub>HPO<sub>4</sub> and 6 M Gn·HCl in water, pH 3.2). The solution was incubated in an ice-salt bath at -10 °C for 5 min and then 125 μL of 200 mM NaNO<sub>2</sub> solution in water was added to it dropwise. After 10 min, 100 μl oxidized peptide solution was added to each of the following solutions:

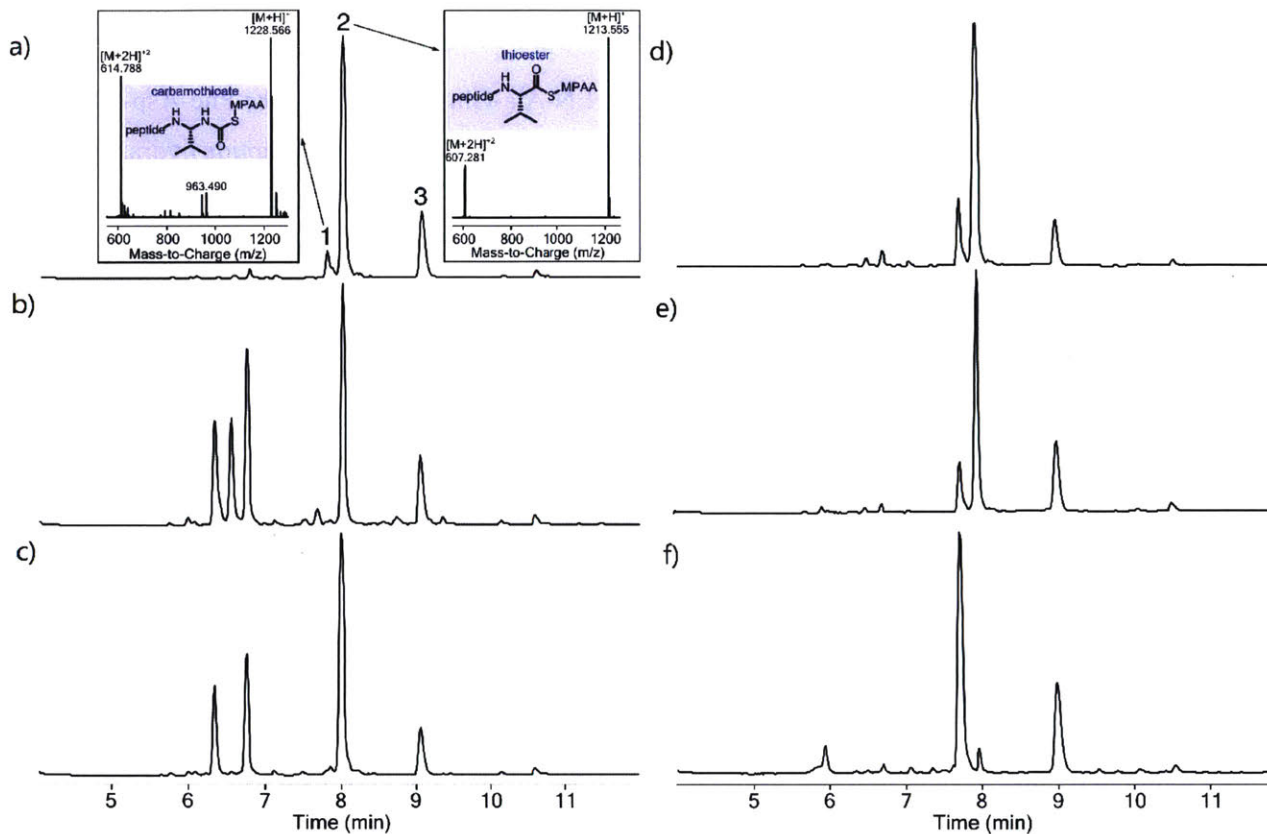
- a) 100 μl of 200 mM MPAA in 200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn·HCl buffer, pH 4.5 (thioester control experiment).
- b) 100 μl of 200 mM MPAA in 200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn·HCl buffer, pH 9.4.
- c) 100 μl of 200 mM MPAA in 200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn·HCl buffer, pH 4.5. pH of the resulting solution was immediately adjusted to 7.4.
- d) 200 μl of acetonitrile. After 5 minutes at rt, 100 μl of 200 mM MPAA in 200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn·HCl buffer, pH 4.5 was added.
- e) 300 μl of 67 mM MPAA in water/acetonitrile (1: 1, v/v) with 0.1% TFA added.
- f) 100 μl of 200 mM MPAA in 200 mM Na<sub>2</sub>HPO<sub>4</sub>, 6 M Gn·HCl buffer, pH 4.5. The resulting solution was then immediately added to 1.8 mL water/acetonitrile (1: 1, v/v) with 0.1% TFA (final pH ~ 3.0).

In each case the reaction was allowed to proceed for 120 minutes before an HPLC-MS sample was taken. Data shown in Figure 3.11 suggest that the exact order of reagent addition, pH of the reaction, or solvent composition had little influence on isocyanate generation. In fact, for reactions a) – e) we did not observe the formation of carbamothioate or other isocyanate derived products to any significant extent. In contrast, when we diluted both reagents (peptide to a final concentration of 133 μM, and MPAA to ~ 10 mM) we observed formation of carbamothioate as the primary reaction product.

To elucidate the solvent effects we then performed 3 more reactions. 0.65 mg peptide was oxidized as above, and then 35 μl of oxidized peptide solution was added to each of the following solutions:

- a) 1.0 mL 8 mM MPAA solution in water/acetonitrile (1: 1, v/v) with 0.1% TFA;
- b) 1.0 mL 8 mM MPAA solution in water/acetonitrile (3: 1, v/v) with 0.1% TFA;
- c) 1.0 mL 8 mM MPAA solution in water with 0.1% TFA.

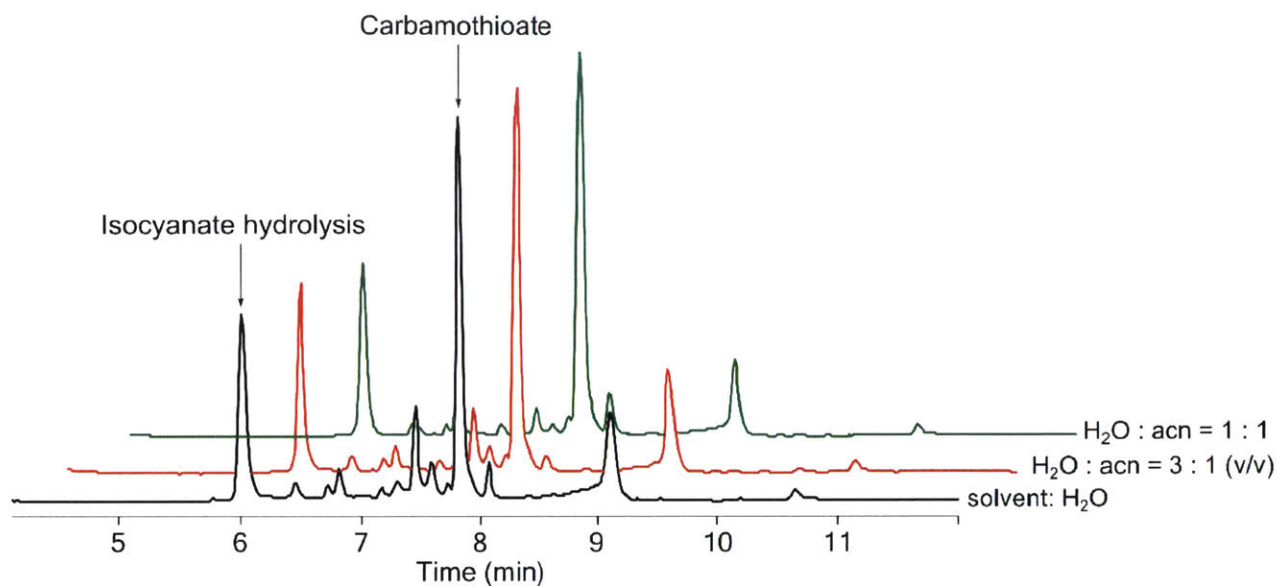
In each case, the reaction was allowed to proceed for 120 minutes before an HPLC-MS sample was taken (Fig. 3.12). We concluded that the concentration of acetonitrile in the solvent plays a minor role in the overall process and opted for water/acetonitrile (1: 1, v/v) as a universal solvent for further experiments.



**Figure 3.11. Optimization of reaction conditions**

HPLC-MS (TIC) chromatograms for  $\text{H}_2\text{N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONH}_2$  oxidation/conjugation reaction with MPA. Peak 1 is the carbamothioate, Peak 2 is the thioester, and Peak 3 is MPA disulfide. Chromatogram indices correspond to the reaction conditions specified above.





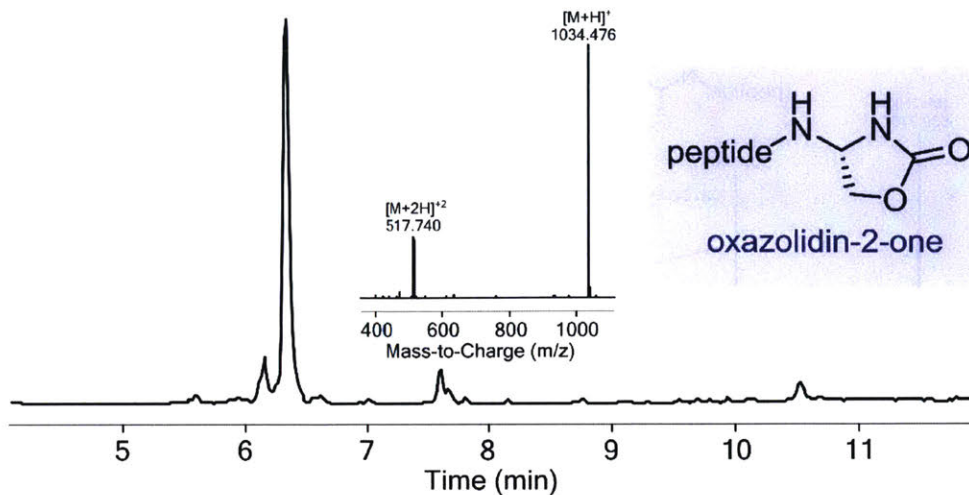
**Figure 3.12. The effect of solvent on the outcome of the conjugation**

Overlaid HPLC-MS (TIC) chromatograms for H<sub>2</sub>N-Ala-Gln-Val-Ile-Asn-Thr-Phe-Asp-Gly-Val-CONHNH<sub>2</sub> oxidation/conjugation reaction in different solvents. Although the cleanest product was obtained in water/acetonitrile (1: 1, v/v), the influence of acetonitrile concentration is minor.

### 3.3.6. Side Reactions Associated with Some C-terminal Amino Acids

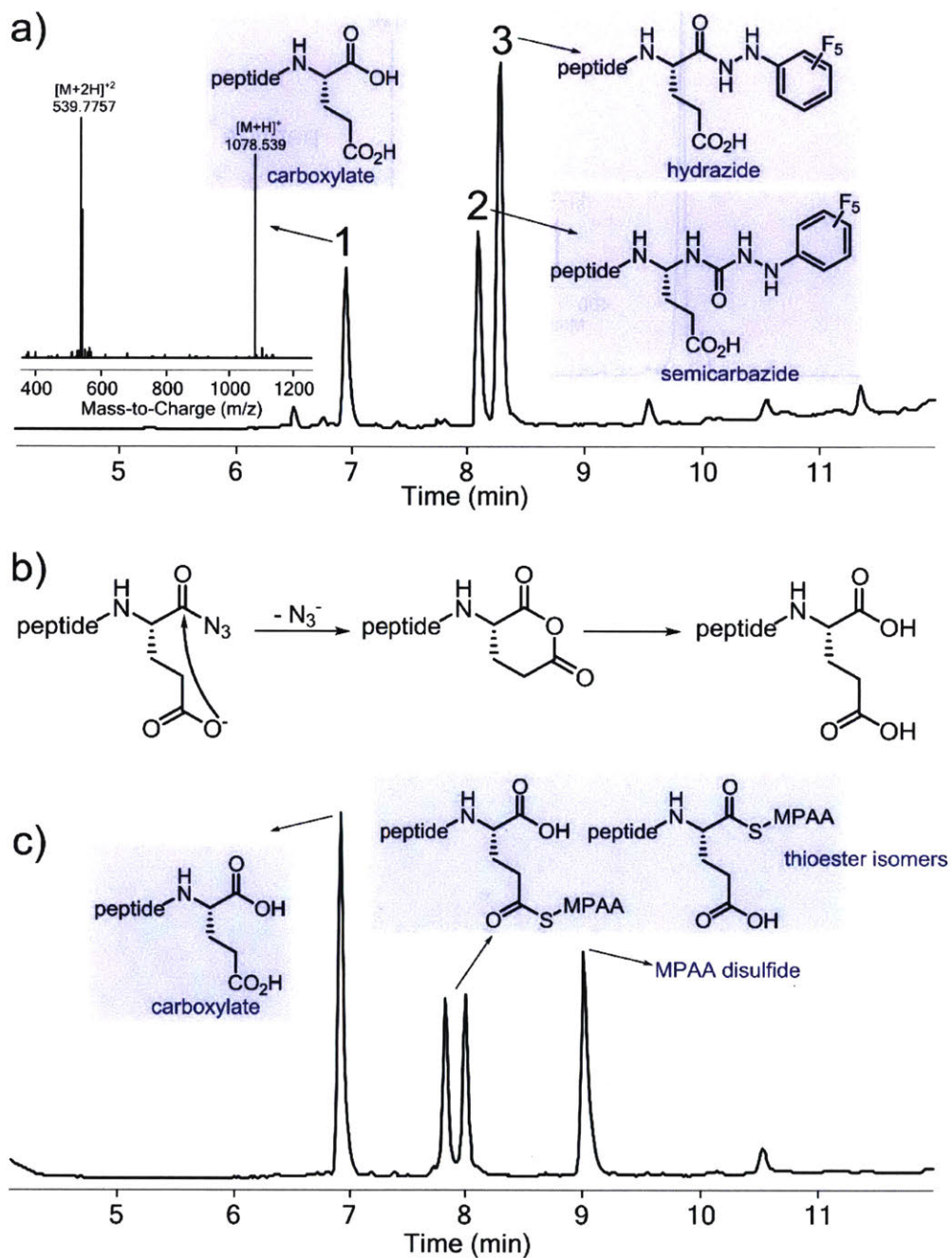
Reactions described in this section were performed under standard conditions using 20 mM perfluorophenylhydrazine as a nucleophile at 57 °C, except in case of H<sub>2</sub>N-Ala-His-Val-Ile-Asn-Thr-Phe-Asp-Gly-Thr-CONHNH<sub>2</sub> the reaction was at 17 °C. Isocyanates formed from peptides with C-terminal Thr and Ser amino acids cyclized to the C-terminal side chain yielding oxazolidin-2-ones. The HPLC-MS trace of one of these reactions is demonstrated in Figure 3.13.

Peptides with C-terminal His and Glu also conjugated poorly to perfluorophenylhydrazine. H<sub>2</sub>N-Ala-Val-Val-Ile-Asn-Thr-Phe-Asp-Gly-**His**-CONHNH<sub>2</sub> and H<sub>2</sub>N-Ala-Leu-Val-Ile-Asn-Thr-Phe-Asp-Gly-**Glu**-CONHNH<sub>2</sub> both showed significant products 14 Da lighter than the starting hydrazide, which were presumably the C-terminal carboxylate formed from hydrolysis of the azide (Fig. 3.14a). We hypothesized that in both cases the azide cyclized to the C-terminal residue's side-chain, and then this cyclic intermediate was hydrolyzed to give the carboxylate (Fig. 3.14b). To test this possibility, we performed oxidation/thioesterification of H<sub>2</sub>N-Ala-Leu-Val-Ile-Asn-Thr-Phe-Asp-Gly-Glu-CONHNH<sub>2</sub> under the standard conditions described in Fang *et al.*<sup>27</sup> As shown in Fig. 3.14c, the reaction yielded the carboxylate and two thioester isomers. When H<sub>2</sub>N-Ala-Val-Val-Ile-Asn-Thr-Phe-Asp-Gly-**His**-CONHNH<sub>2</sub> was subject to the same oxidation/thioesterification conditions, a significant amount of the carboxylate was observed, which is not typical for peptides bearing other C-terminal amino acids. These data are consistent with our initial hypothesis that azides of His and Glu cyclize with their side-chains.



**Figure 3.13. Formation of oxazolidin-2-ones for Ser/Thr-containing peptides**

HPLC-MS (TIC) chromatogram for H<sub>2</sub>N-Ala-Asn-Val-Ile-Asn-Thr-Phe-Asp-Gly-Ser-CONHNH<sub>2</sub> oxidation/conjugation reaction with MS inset for the main product.

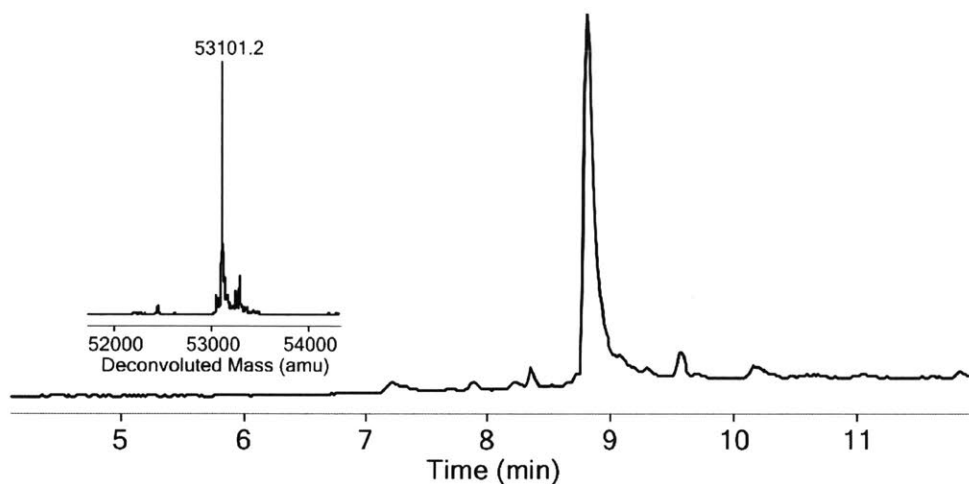


**Figure 3.14. Oxidation of peptides bearing C-terminal Glu or His hydrazides**

a) HPLC-MS (TIC) chromatogram for  $\text{H}_2\text{N}-\text{Ala}-\text{Leu}-\text{Val}-\text{Ile}-\text{Asn}-\text{Thr}-\text{Phe}-\text{Asp}-\text{Gly}-\text{Glu}-\text{CONHNH}_2$  oxidation/conjugation reaction with MS inset for carboxylate. b) A proposed scheme for carboxylate formation. c) HPLC-MS (TIC) chromatogram for  $\text{H}_2\text{N}-\text{Ala}-\text{Leu}-\text{Val}-\text{Ile}-\text{Asn}-\text{Thr}-\text{Phe}-\text{Asp}-\text{Gly}-\text{Glu}-\text{CONHNH}_2$  oxidation/thioesterification reaction (200 mM MPAAs).

### 3.3.7. Sortase A Mediated Ligation of Peptide Hydrazide to LF<sub>N</sub>-DTA

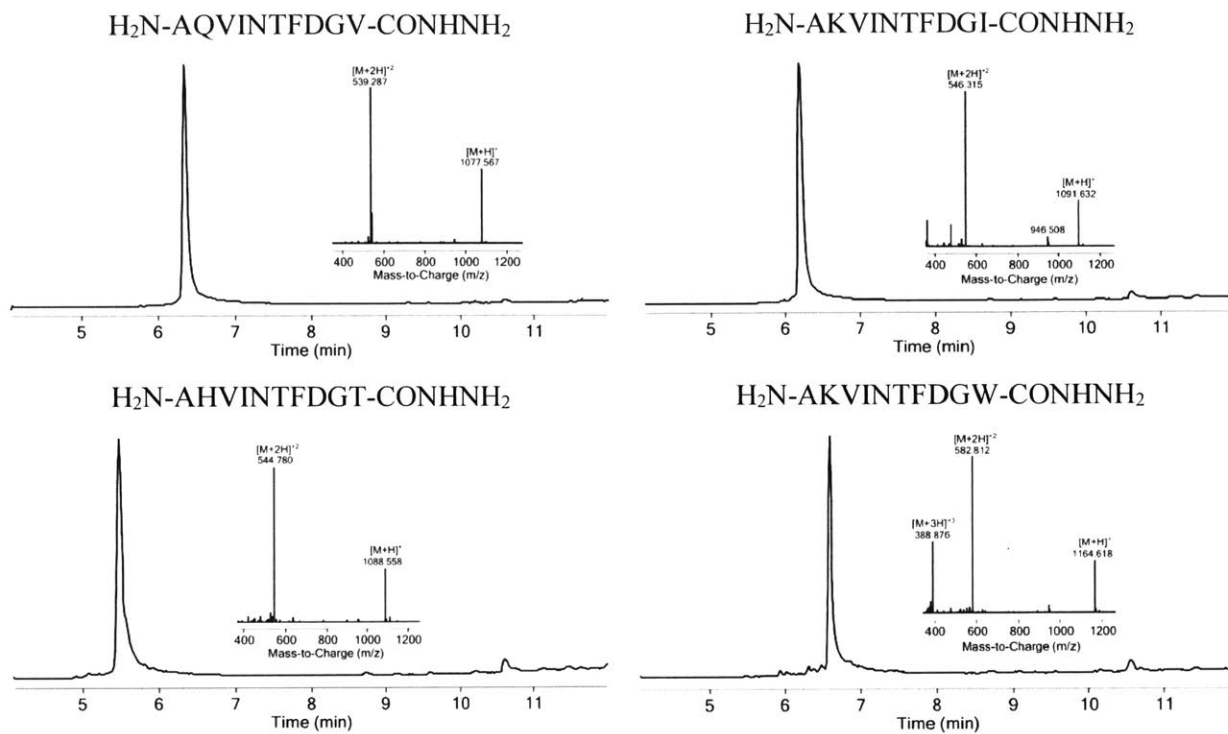
LF<sub>N</sub>-DTA protein (N-terminal domain of Anthrax toxin lethal factor produced as a C-terminal fusion to the A subunit of Diphtheria toxin) was expressed and purified as previously described.<sup>45</sup> To produce a C-terminal protein hydrazide we employed a Sortase A\* mediated ligation of H<sub>2</sub>N-Gly<sub>5</sub>-Leu-Glu-Ile-CONHNH<sub>2</sub> to LF<sub>N</sub>-DTA-Leu-Pro-Ser-Thr-Gly<sub>2</sub>-His<sub>5</sub>. Ligation was carried out under the following conditions: a mixture of 300 µl Tris buffer (125 mM Tris, 100 mM NaCl, 1 mM CaCl<sub>2</sub>, pH 7.5) containing 38 µM LF<sub>N</sub>-DTA, 350 µM peptide hydrazide, 2.5 µM Sortase A\*, and 80 µl Ni-NTA beads was incubated at 17 °C for 30 minutes while nutating gently. Afterwards, the reaction mixture was centrifuged at 13000 rpm for 5 minutes. The beads were washed two times with Tris buffer (150 µl), and then with 40 mM imidazole in Tris buffer. The washes were combined with the reaction supernatant and concentrated over an Amicon Ultra-4 Ultracel-10K centrifugation filter. To remove excess peptide hydrazide and imidazole, the protein solution was subject to buffer exchange into Tris buffer using a HiPrep 26/10 Desalting column. The reaction was analyzed with HPLC-MS to confirm complete formation of the desired C-terminal protein hydrazide. The MS spectrum of the resulting protein was deconvoluted using the maximum entropy algorithm to obtain deconvoluted the protein spectrum in Figure 3.15.



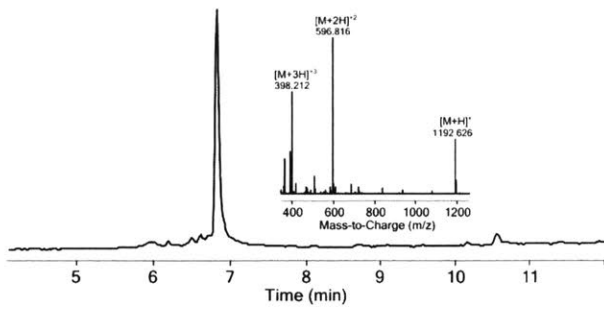
**Figure 3.15. HPLC-MS of C-terminal hydrazide of LF<sub>N</sub>-DTA**

HPLC-MS (TIC) chromatogram for LF<sub>N</sub>-DTA-Leu-Pro-Ser-Thr-Gly<sub>5</sub>-Leu-Glu-Ile-CONHNH<sub>2</sub> obtained via Sortase A\* mediated ligation of H<sub>2</sub>N-Gly<sub>5</sub>-Leu-Glu-Ile-CONHNH<sub>2</sub> to LF<sub>N</sub>-DTA-Leu-Pro-Ser-Thr-Gly<sub>2</sub>-His<sub>5</sub>. The deconvoluted MS on the left was obtained using the maximum entropy algorithm.

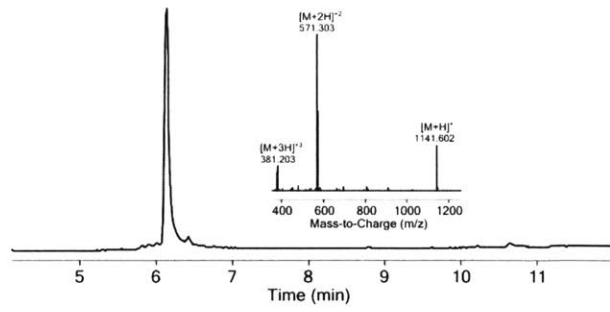
### 3.3.8. Peptide Synthesis: Analytical Data



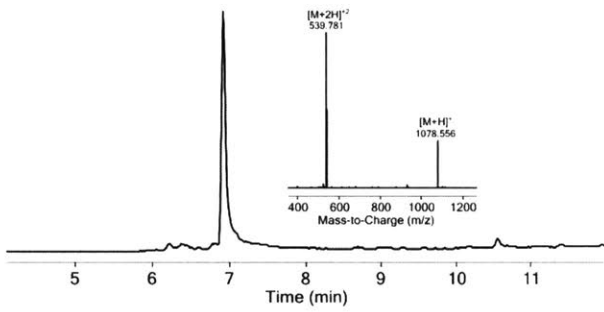
H<sub>2</sub>N-AVVINTFDGR-CONHNH<sub>2</sub>



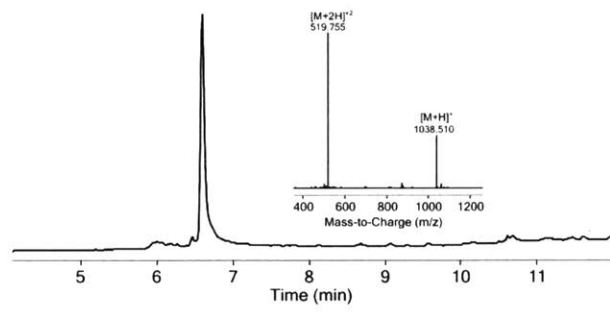
H<sub>2</sub>N-AYVINTFDGK-CONHNH<sub>2</sub>



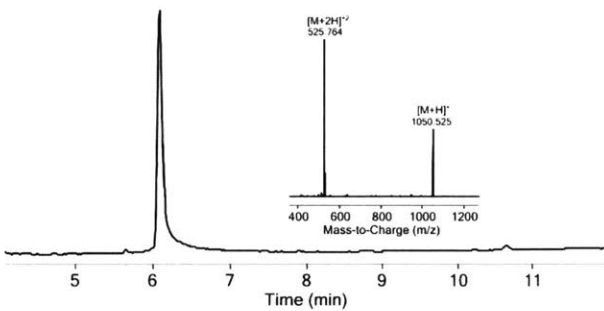
H<sub>2</sub>N-ADVINTFDGL-CONHNH<sub>2</sub>



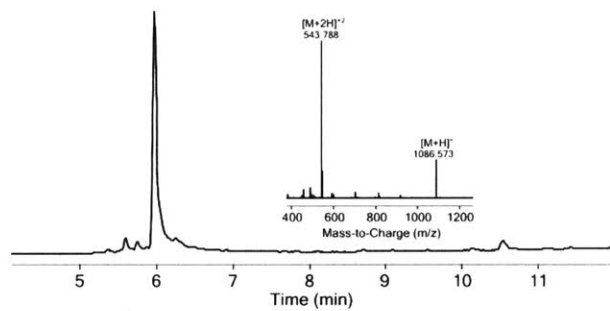
H<sub>2</sub>N-AGVINTFDGM-CONHNH<sub>2</sub>



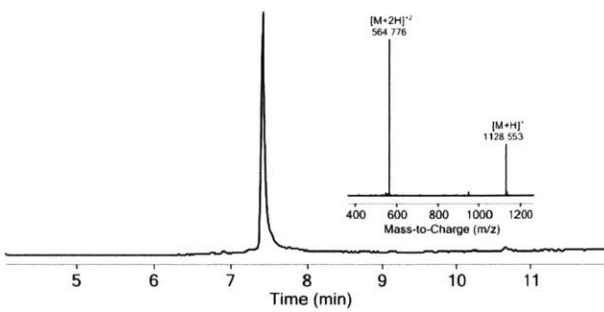
H<sub>2</sub>N-AEVINTFDGA-CONHNH<sub>2</sub>



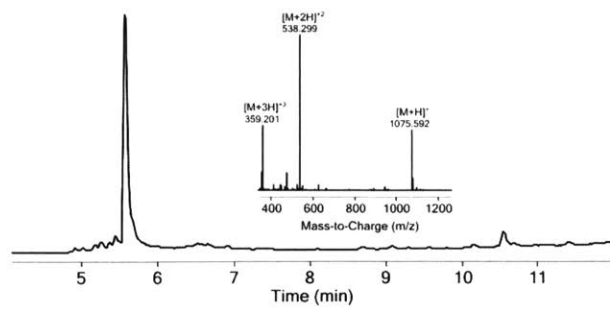
H<sub>2</sub>N-AVVINTFDGH-CONHNH<sub>2</sub>



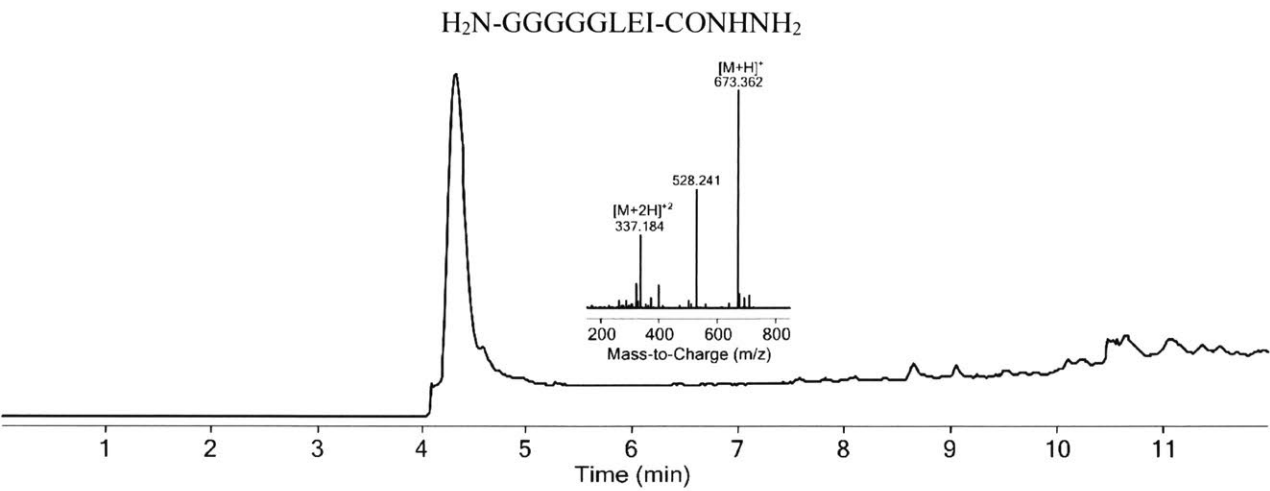
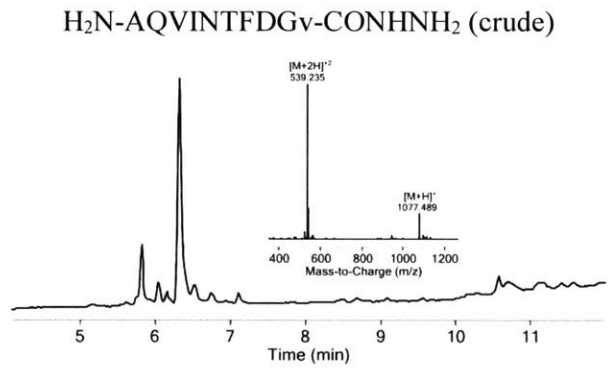
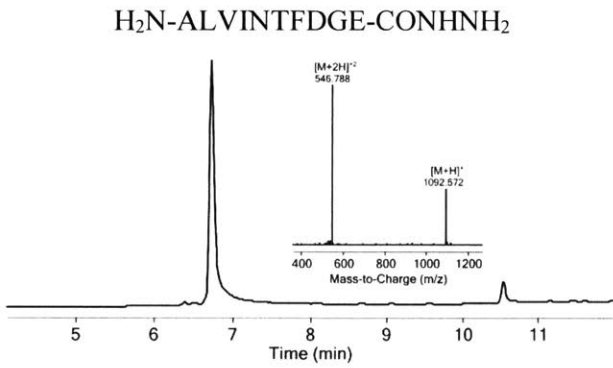
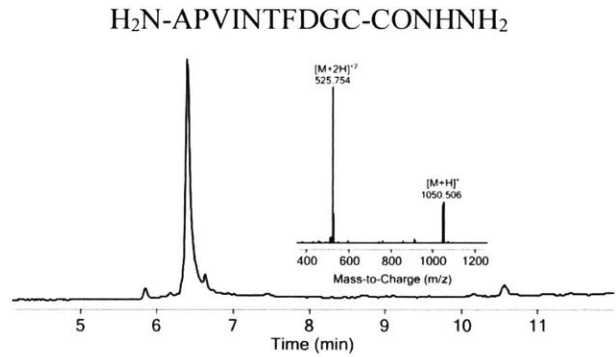
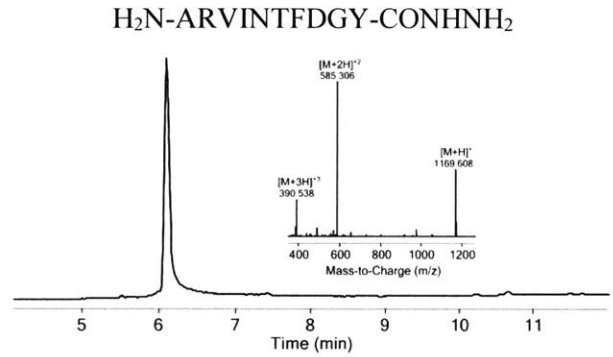
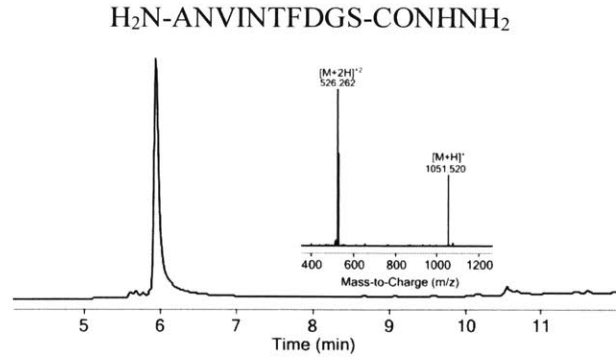
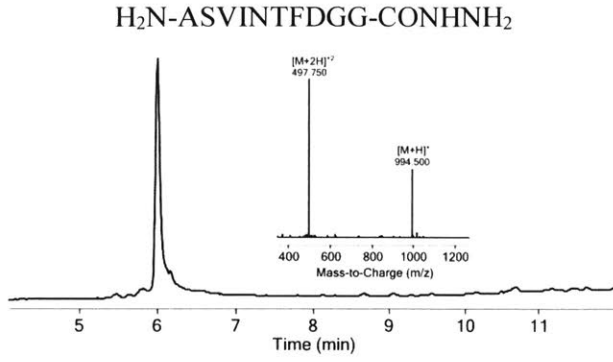
H<sub>2</sub>N-AMVINTFDGF-CONHNH<sub>2</sub>



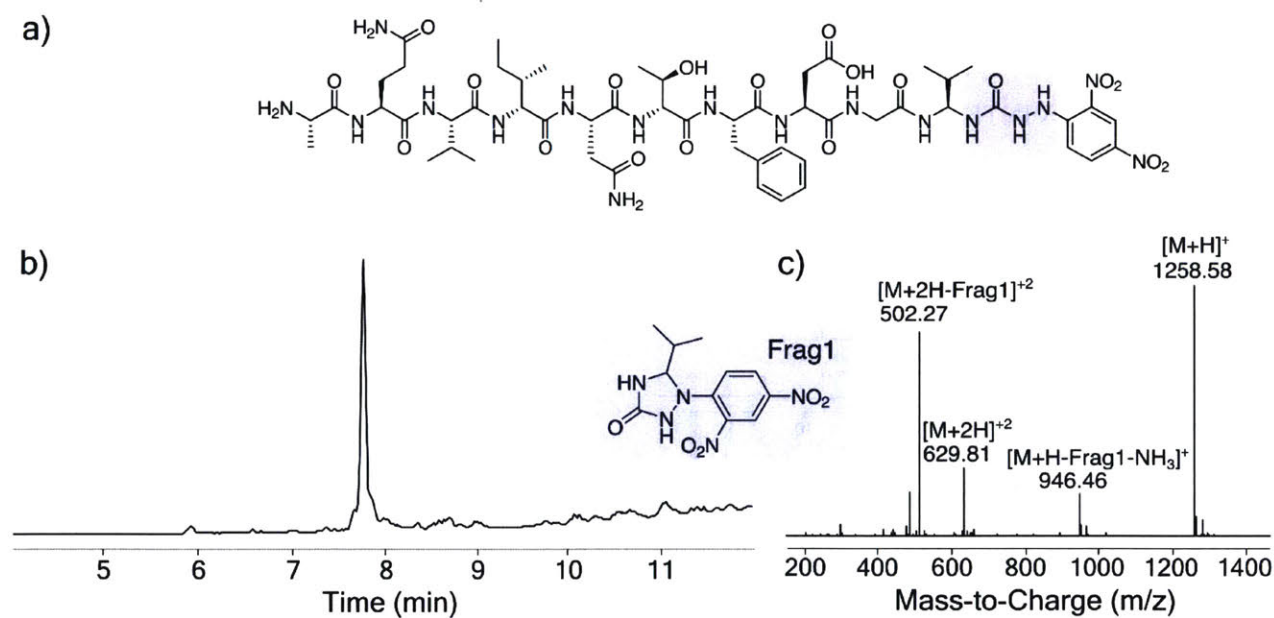
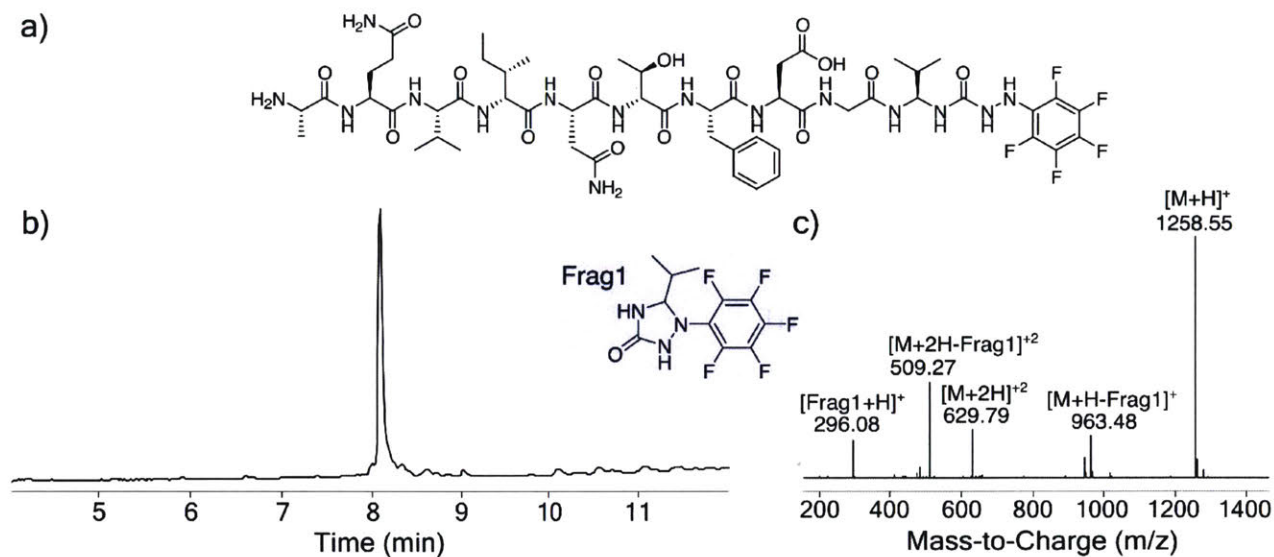
H<sub>2</sub>N-AKVINTFDGP-CONHNH<sub>2</sub>

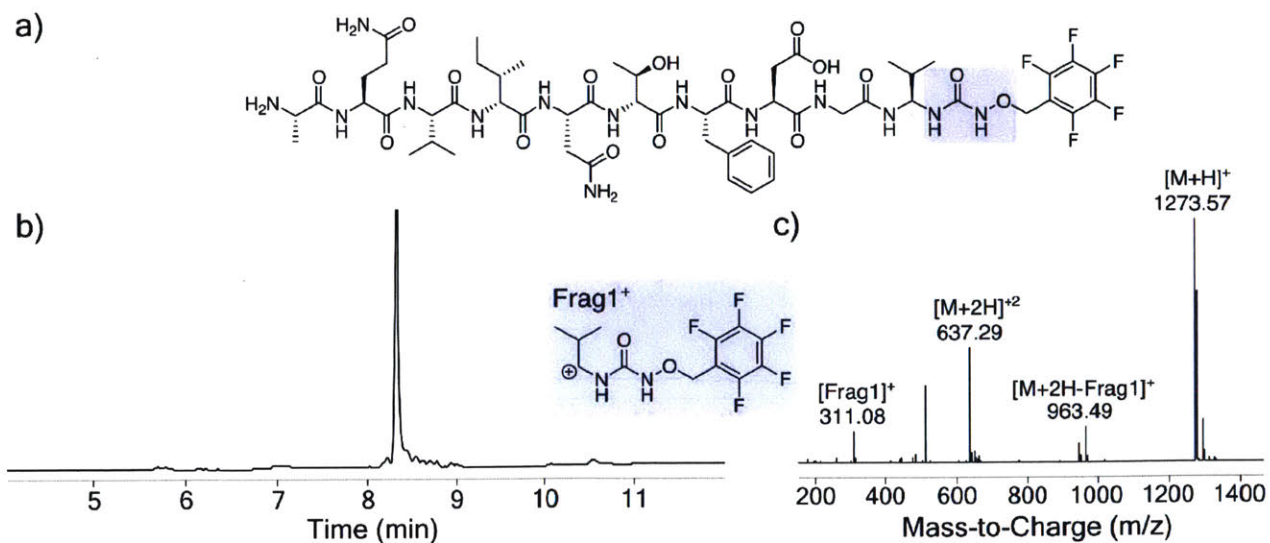
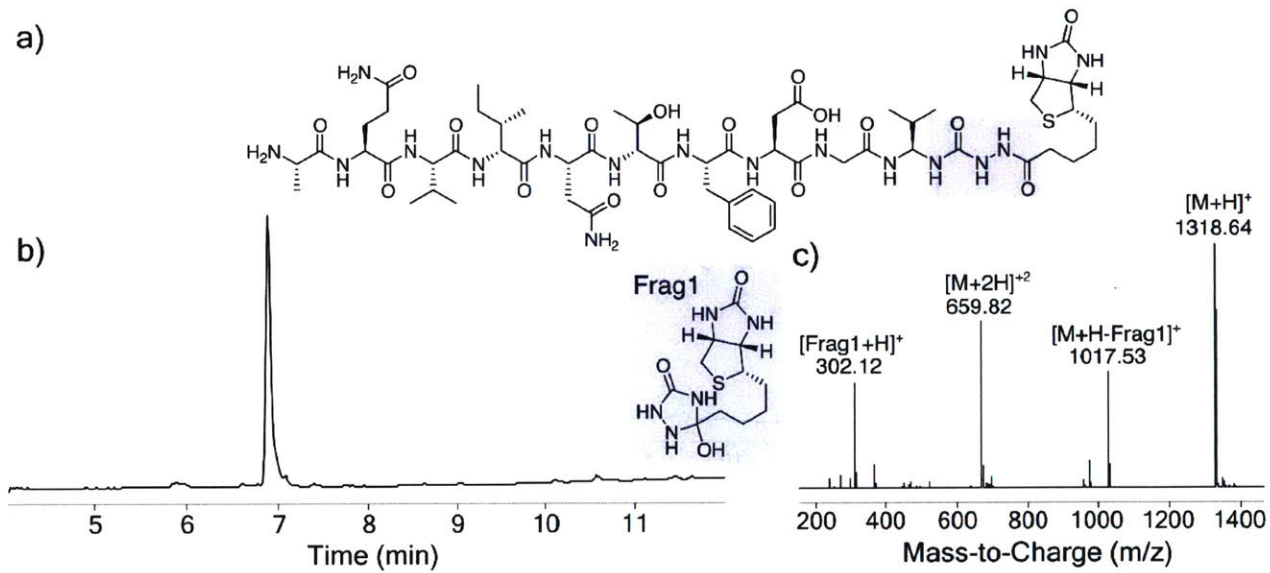
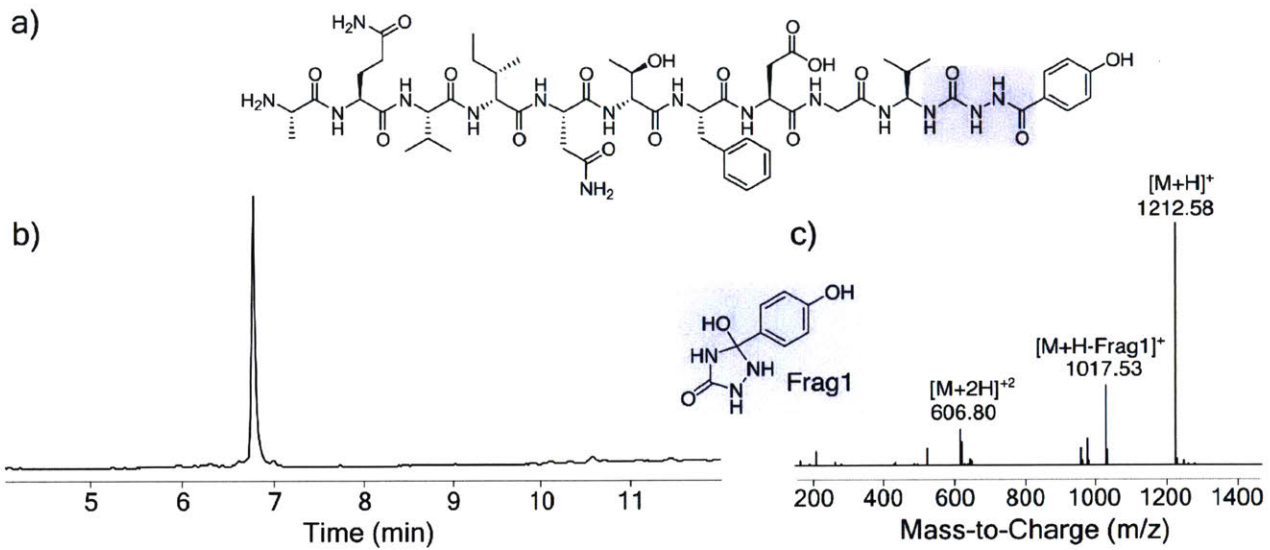


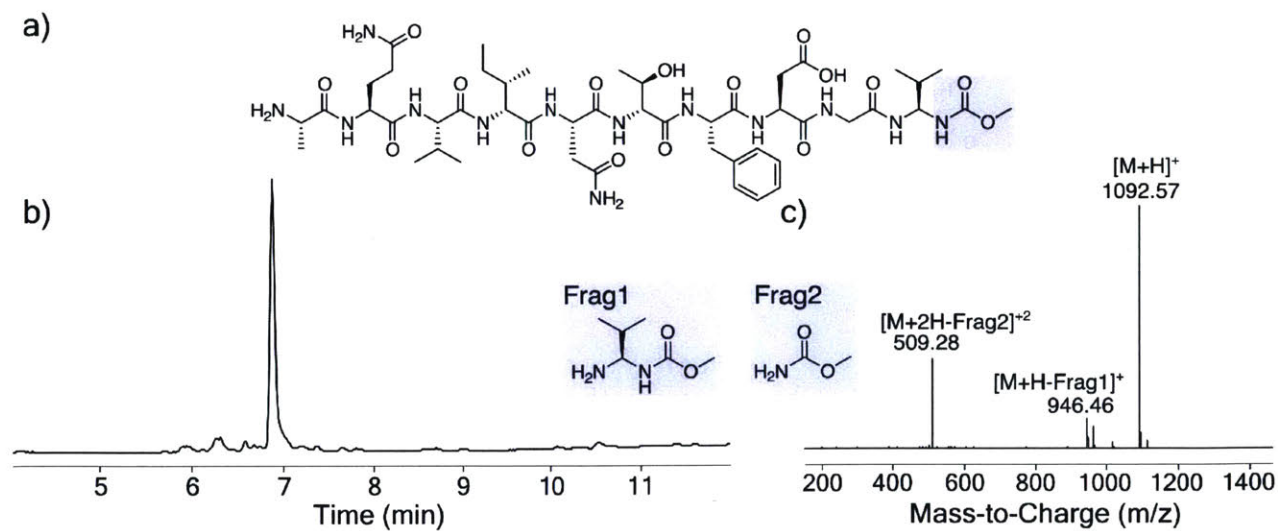
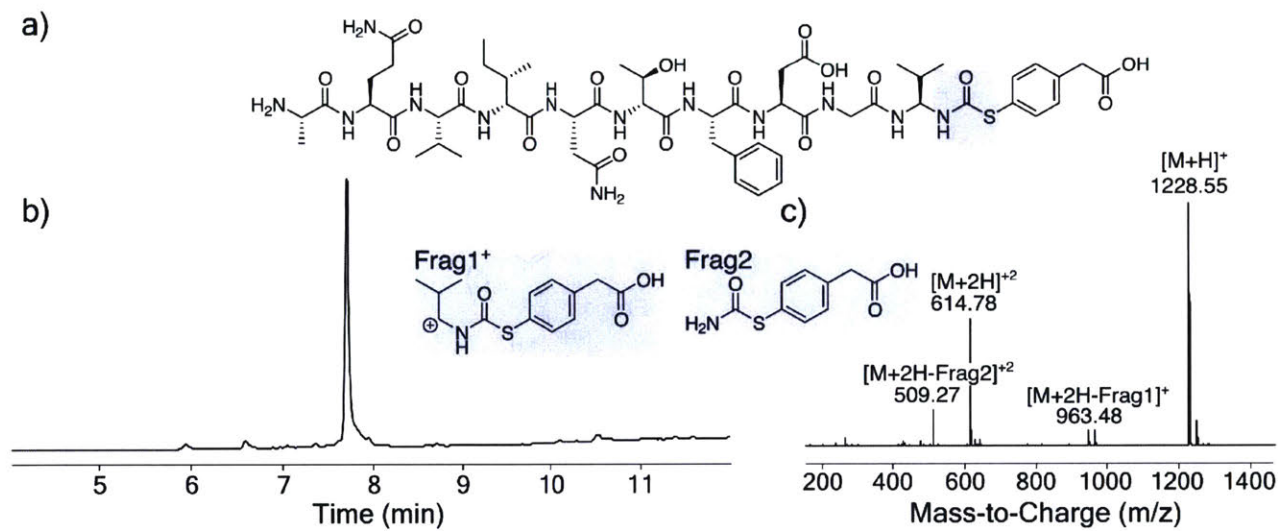
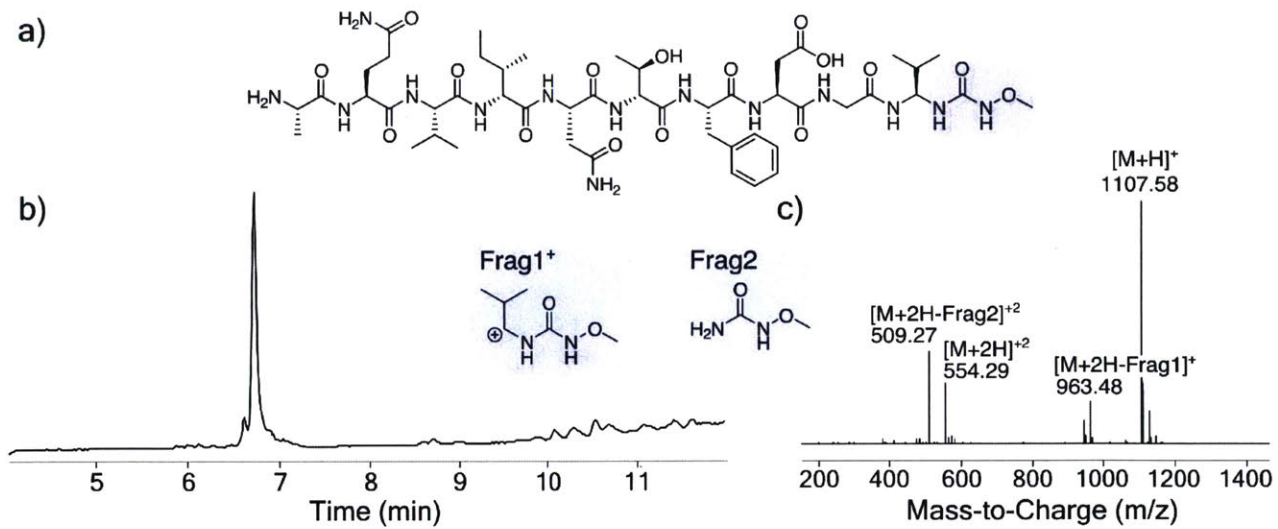


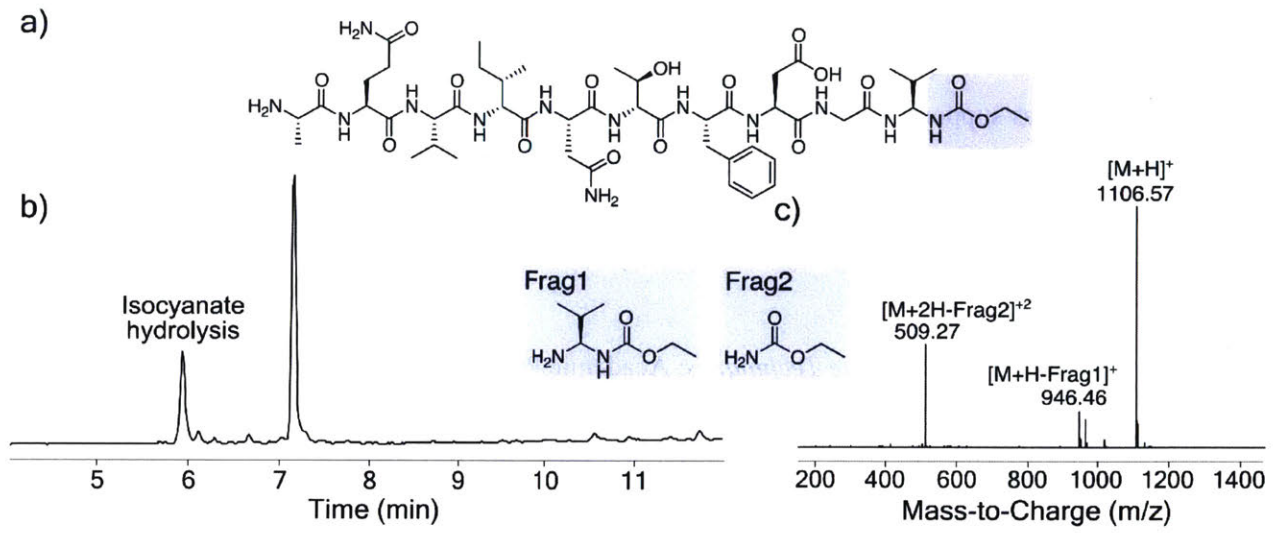


### 3.3.9. Nucleophile Scope of Conjugation: Analytical Data









### 3.5. Acknowledgements

This research was supported by the MIT Deshpande Center (B. L. P.), and Amgen Summer Graduate Fellowship for A. A. V. We also thank Dr. Amy Rabideau (MIT Chemistry Department), and Ms. Hansol Kang (MIT Chemistry Department) for providing protein samples.

### 3.6. References

- (1) Hermanson, G. *Bioconjugate Techniques*; Academic Press: London, 2013.
- (2) Kalia, J.; Raines, R. *Curr. Org. Chem.* **2010**, *14*, 138–147.
- (3) Baslé, E.; Joubert, N.; Pucheault, M. *Chem. Biol.* **2010**, *17*, 213–227.
- (4) Sletten, E.; Bertozzi, C. *Angew. Chemie Int. Ed.* **2009**, *48*, 6974–6998.
- (5) Debets, M.; van Hest, J.; Rutjes, F. *Org. Biomol. Chem.* **2013**, *11*, 6439–6455.
- (6) Noda, H.; Erős, G.; Bode, J. *J. Am. Chem. Soc.* **2014**, *136*, 5611–5614.
- (7) Behrens, C.; Hooker, J.; Obermeyer, A.; Romanini, D.; Katz, E.; Francis, M. *J. Am. Chem. Soc.* **2011**, *133*, 16398–16401.
- (8) Obermeyer, A.; Jarman, J.; Netirojjanakul, C.; El Muslemany, K.; Francis, M. *Angew. Chem. Int. Ed. Engl.* **2013**, *53*, 1057–1061.
- (9) Lemieux, G.; De Graffenried, C.; Bertozzi, C. *J. Am. Chem. Soc.* **2003**, *125*, 4708–4709.
- (10) Agard, N.; Prescher, J.; Bertozzi, C. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
- (11) Chalker, J.; Bernardes, G.; Lin, Y.; Davis, B. *Chem. - Asian J.* **2009**, *4*, 630–640.
- (12) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science.* **1994**, *266*, 776–778.
- (13) Dondoni, A. *Angew. Chem., Int. Ed.* **2008**, *47*, 8995–8997.
- (14) Carrico, I.; Carlson, B.; Bertozzi, C. *Nat. Chem. Biol.* **2007**, *3*, 321–322.
- (15) Dirksen, A.; Dawson, P. *Bioconjug. Chem.* **2008**, *19*, 2543–2548.
- (16) Cornish, V.; Hahn, K.; Schultz, P. *J. Am. Chem. Soc.* **1996**, *118*, 8150–8151.
- (17) Bernardes, G.; Chalker, J.; Errey, J.; Davis, B. *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053.
- (18) Besret, S.; Ollivier, N.; Blanpain, A.; Melnyk, O. *J. Pept. Sci.* **2008**, *14*, 1244–1250.
- (19) Melnyk, O.; Ollivier, N.; Besret, S.; Melnyk, P. *Bioconjug. Chem.* **2014**, *25*, 629–639.
- (20) Besret, S.; Vicogne, J.; Dahmani, F.; Fafeur, V.; Desmet, R.; Drobecq, H.; Romieu, A.; Melnyk, P.; Melnyk, O. *Bioconjug. Chem.* **2014**, *25*, 1000–1010.
- (21) Ozaki, S. *Chem. Rev.* **1972**, *72*, 457–496.
- (22) Burgess, K.; Ibarzo, J.; Linthicum, D.; Russell, D.; Shin, H.; Shitangkoon, A.; Totani, R.; Zhang, A. *J. Am. Chem. Soc.* **1997**, *119*, 1556–1564.
- (23) Boeijen, A.; Liskamp, R. *Eur. J. Org. Chem.* **1999**, 2127–2135.
- (24) Nowick, J.; Holmes, D.; Noronha, G.; Smith, E.; Nguyen, T.; Huang, S.-L. *J. Org. Chem.* **1996**, *61*, 3929–3934.
- (25) Patil, B.; Vasanthakumar, G.-R.; Suresh Babu, V. *J. Org. Chem.* **2003**, *68*, 7274–7280.
- (26) Sureshbabu, V.; Patil, B.; Venkataramanarao, R. *J. Org. Chem.* **2006**, *71*, 7697–7705.
- (27) Fang, G.; Li, Y.; Shen, F.; Huang, Y.; Li, J.; Lin, Y.; Cui, H.; Liu, L. *Angew. Chem., Int. Ed.* **2011**, *50*, 7645–7649.
- (28) Fang, G.; Wang, J.; Liu, L. *Angew. Chem., Int. Ed.* **2012**, *51*, 10347–10350.
- (29) Raspoet, G.; Nguyen, M.; McGarraghy, M.; Hegarty, A. *J. Org. Chem.* **1998**, *63*, 6878–6885.
- (30) Inouye, K.; Watanabe, K.; Shin, M. *J. Chem. Soc. Perkin Trans. 1* **1977**, 1905–1911.
- (31) Schnabel, E. *Liebigs Ann. Chem.* **1962**, *659*, 168–184.
- (32) Okada, Y.; Tsuda, Y.; Yagyu, M. *Chem. Pharm. Bull.* **1980**, *28*, 2254–2258.
- (33) Bräse, S.; Gil, C.; Knepper, K.; Zimmermann, V. *Angew. Chem., Int. Ed.* **2005**, *44*, 5188–5240.

- (34) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. a; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A.; Jensen, K.; Pentelute, B. *Chembiochem* **2014**, *15*, 713–720.
- (35) Mong, S.; Vinogradov, A.; Simon, M.; Pentelute, B. *Chembiochem* **2014**, *15*, 721–733.
- (36) Zheng, J.; Tang, S.; Huang, Y.; Liu, L. *Acc. Chem. Res.* **2013**, *46*, 2475–2484.
- (37) Fruton, J. *J. Biol. Chem.* **1942**, *146*, 463–470.
- (38) Weinstock, M.; Jacobsen, M.; Kay, M. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 11679–11684.
- (39) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, 3267–3272.
- (40) Navarre, W.; Schneewind, O. *Mol. Microbiol.* **1994**, *14*, 115–121.
- (41) Chen, I.; Dorr, B.; Liu, D. *Proc. Natl. Acad. Sci.* **2011**, *108*, 11399–11404.
- (42) Stavropoulos, G.; Gatos, D.; Magafa, V.; Barlos, K. *Lett. Pept. Sci.* **1995**, *2*, 315–318.
- (43) Cohen, D.; Zhang, C.; Pentelute, B.; Buchwald, S. *J. Am. Chem. Soc.* **2015**, 9784–9787.
- (44) Li, Y.; Hoskins, J.; Sreerama, S.; Grayson, S. *Macromolecules* **2010**, *43*, 6225–6228.
- (45) Ling, J.; Policarpo, R.; Rabideau, A.; Liao, X.; Pentelute, B. *J. Am. Chem. Soc.* **2012**, *134*, 10749–10752.





## Chapter 4. Macrocyclization of Unprotected Peptide Isocyanates

The work presented in this chapter was published in the following manuscript and is reproduced here with permission of American Chemical Society:

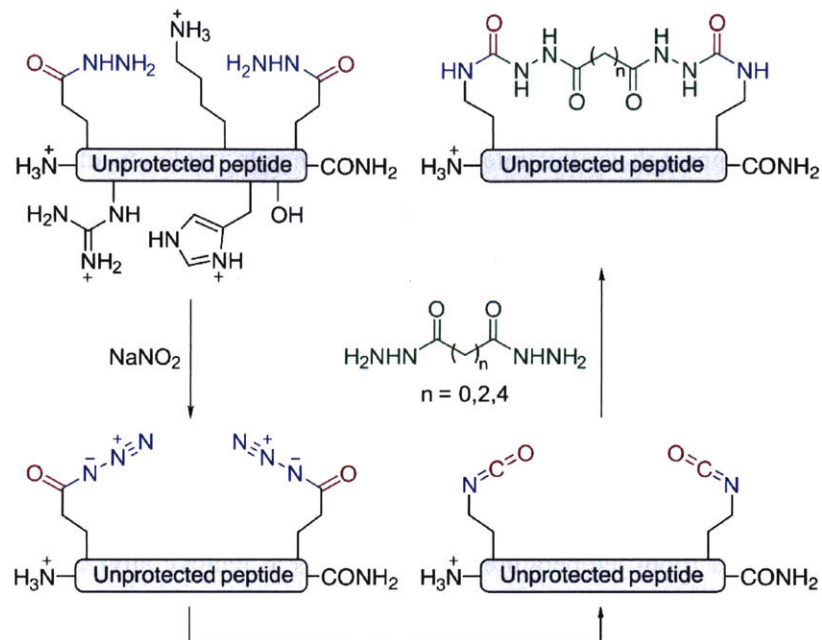
Vinogradov, A.; Choo, Z.-N.; Totaro, K.; Pentelute, B. Macrocyclization of Unprotected Peptide Isocyanates. *Org. Lett.* **2016**, *18*, 1226–1229. DOI: 10.1021/acs.orglett.5b03626

## 4.1. Introduction

In this chapter, we further explore the utility of the chemistry discovered and discussed in detail in Chapter 3. Here, we extended the methods for the generation and utilization of C-terminal isocyanates of unprotected peptides and proteins in aqueous solvents, and studied the possibility of preparing similar constructs bearing two side-chain isocyanate moieties. We developed associated synthesis protocols and applied this methodology to explore the utility of a new two-component peptide macrocyclization strategy based on the discovered chemistry.

Peptide macrocyclization is a powerful tool for enhancing and modulating bioactivity, cell permeability, proteolytic stability and other important properties of peptides. The pharmaceutical promise of macrocyclized or “stapled” peptides has drawn serious attention from the scientific community, resulting in a number of peptide macrocyclization strategies. The best known approach involves ring-closing metathesis of  $\alpha,\alpha$ -disubstituted olefin-bearing amino acid residues and results in all-hydrocarbon bridge tethering two residues on the same helical face.<sup>1,2</sup> Successful applications of this approach include the development of p53/MDM2 interaction inhibitors,<sup>3-5</sup> BCL-2 family proteins activators and repressors,<sup>6-9</sup>  $\beta$ -catenin binders,<sup>10,11</sup> and others.<sup>12,13</sup> Among other techniques now available for macrocyclization are cysteine modification reactions, including arylation,<sup>14,15</sup> alkylation<sup>16-18</sup> and disulfide bond formation;<sup>19</sup> side-chain to side-chain reactions, such as lactam<sup>20,21</sup> and oxime formation;<sup>22</sup> and cycloaddition reactions.<sup>23</sup> Combined together, these impressive advances in the field of peptide macrocyclization enable access to a wide range of structurally and chemically different substrates.

In this study we investigated the possibility of using isocyanates as reactive handles to achieve facile two-component cyclization with bifunctional nucleophiles, such as hydrazides of dicarboxylic acids (Fig. 4.1). This approach enables the formation of cyclic systems of variable size and topology in a modular fashion. Additionally, the cyclization proceeds selectively over unprotected amino acid side-chains and results in unique, hydrophilic bridges of variable rigidity, which modulate the biochemical properties of peptides.



**Figure 4.1. The concept of the study**

Two-component macrocyclization of unprotected peptide isocyanates with dicarboxylic acid hydrazides.

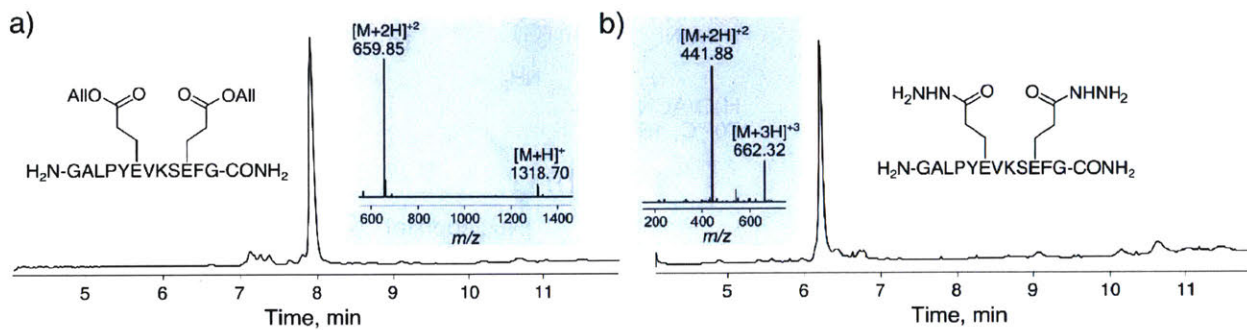
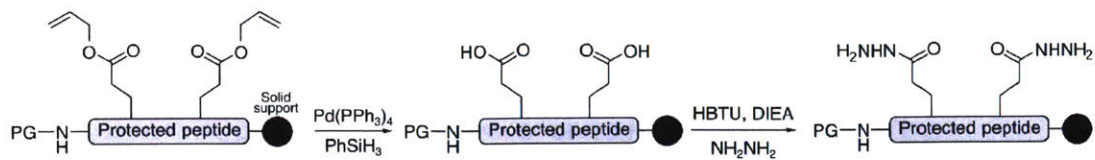
## 4.2. Results and Discussion

### 4.2.1. Establishment of the Macrocyclization Conditions

The idea behind our approach is illustrated in Fig. 4.1. Facile oxidation of an unprotected peptide containing two glutamic acid  $\gamma$ -hydrazides with sodium nitrite in water at pH 3-4 yields the peptide containing two side-chain acyl azides.<sup>24</sup> Without intermediate isolation, the peptide is then introduced to a bi-functional nucleophile to allow for a Curtius rearrangement followed by a cyclization between generated isocyanates and the nucleophile. If dihydrazides are used as the nucleophile, a double N'-substituted semicarbazide bridge tethering two diaminobutyric acid side-chains is constructed. Starting peptide hydrazides can be prepared<sup>25,26</sup> on-resin during peptide synthesis from  $\gamma$ -allyl-protected glutamic acid in two straightforward steps. First, the fully protected peptide is treated with 0.2 equivalents of Pd(PPh<sub>3</sub>)<sub>4</sub> and 20 equivalents of SiPhH<sub>3</sub> at room temperature for 20 minutes to remove allyl ester protection from the glutamic acid side-chain. Then, after washing the resin with DCM and DMF, the free acid is activated with 0.38M HBTU in DMF and 2 equivalents of DIEA for two minutes, and introduced to a 0.5 M solution of hydrazine in DMF for 10 minutes, which cleanly yields glutamic acid  $\gamma$ -hydrazides (Fig. 4.2).

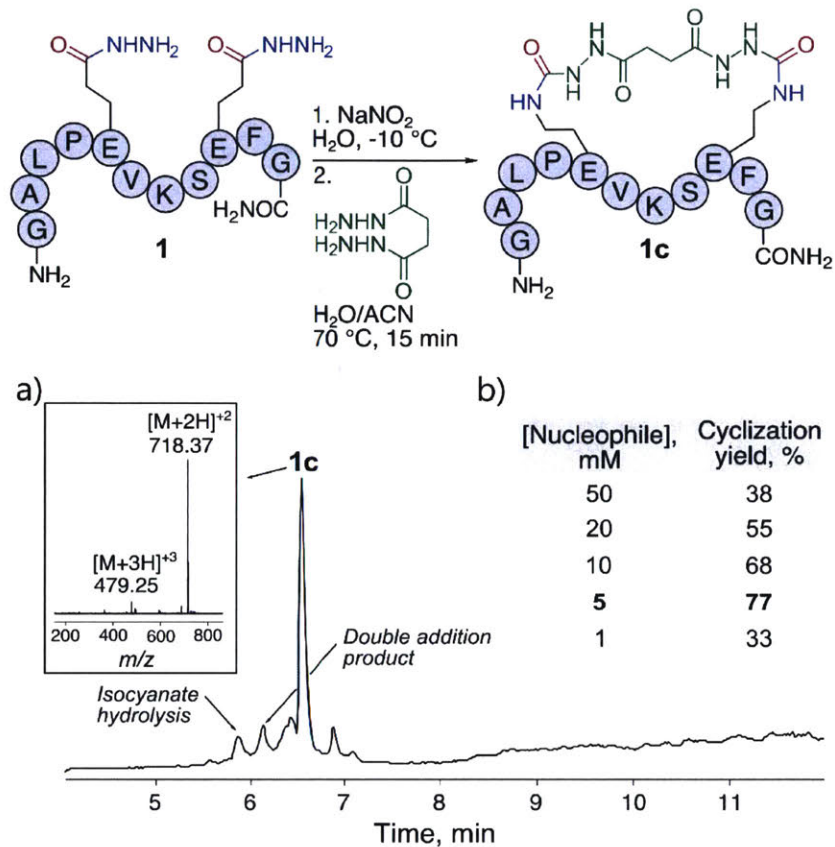
To investigate the feasibility of our approach, we prepared a model peptide, NH<sub>2</sub>-GALPYXVKSXFG-2, where X is glutamic acid  $\gamma$ -hydrazide (**1**) and studied its cyclization with succinic dihydrazide (Fig. 4.3a). In our previous study in Chapter 3, we established that nucleophile concentration is key to successful isocyanate conjugation in aqueous solvents. Thus, we began by screening various concentrations of succinic dihydrazide. In a typical reaction, 3.1 mM peptide in 200 mM sodium phosphate aqueous buffer at pH 3.2 was oxidized with 20 mM sodium nitrite at -10 °C for 10 minutes. An excess of succinic dihydrazide in water/acetonitrile (1:3, v/v) was then added to the peptide, and the mixture thermostated at 70 °C for 10 minutes. The reaction outcome was analyzed by HPLC-MS. As summarized in Fig. 4.3b, we investigated five different nucleophile concentrations and found that the reaction proceeded most efficiently at 5 mM succinic dihydrazide, yielding 77% macrocyclized product. Lower nucleophile concentrations resulted in significant hydrolysis of at least one isocyanate functionality, which abolished the cyclization, while more concentrated nucleophile promoted double addition over the cyclization.

Notably, the Curtius rearrangement and the subsequent conjugation to isocyanates and not directly to acyl azides was instrumental for the success of macrocyclization: our attempts to cyclize acyl azides with succinic dihydrazide without the rearrangement failed repeatedly and led to a complex mixture of products. Accordingly, we oxidized dihydrazide **1** with sodium nitrite essentially as described above to yield a peptide bearing two acyl azide moieties. Since acyl azides are much less reactive than isocyanates, we co-incubated the reaction partners at higher concentrations (1.0-1.5 mM peptide and 5-50 mM succinic dihydrazide at room temperature or 60 °C) trying to mimic established protocols for acyl azides conjugation.<sup>24</sup> In all cases



**Figure 4.2. On resin synthesis of glutamic acid  $\gamma$ -hydrazides**

a) HPLC-MS (TIC) chromatogram with MS inset for crude peptide **1** prior to allyl ester removal. b) HPLC-MS (TIC) chromatogram with MS inset for crude peptide **1** after allyl ester removal and hydrazine coupling.



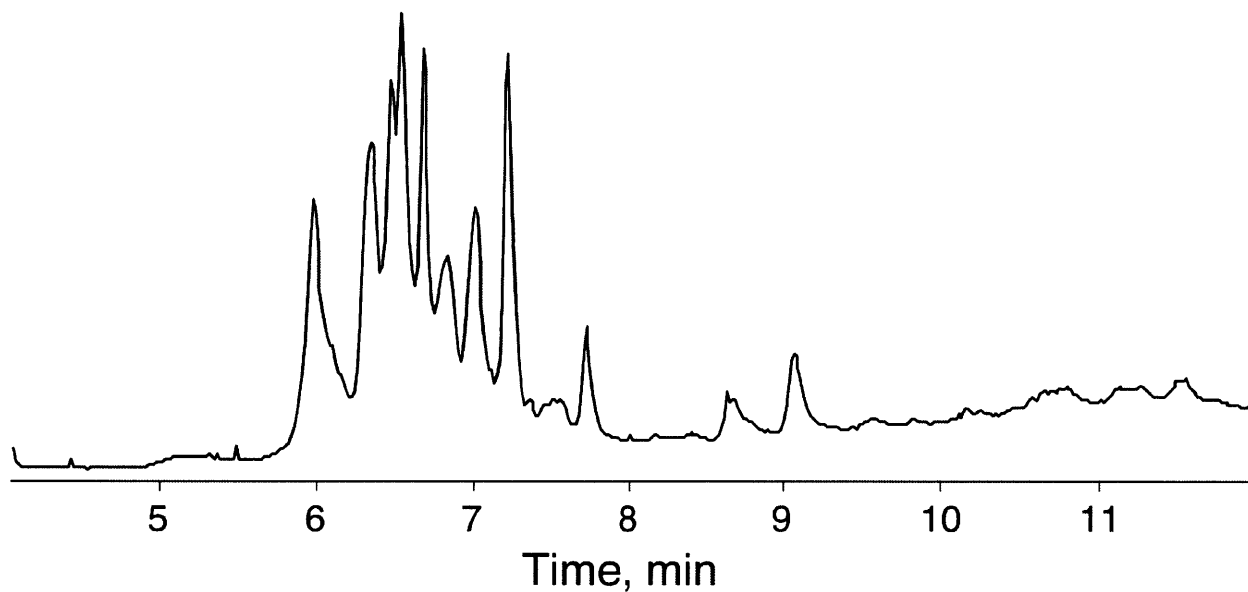
**Figure 4.3. Development of isocyanate macrocyclization reaction**

a) HPLC-MS (TIC) chromatogram of the crude macrocyclization reaction between peptide **1** and succinic dihydrazide with MS inset for the expected cyclized product. The peptide was oxidized with 20 mM  $\text{NaNO}_2$  at  $-10^\circ\text{C}$  for 10 minutes, diluted into a 5 mM solution of the nucleophile in water/acetonitrile (1:3, v/v) mixture and incubated at  $70^\circ\text{C}$  for 10 minutes. b) Yield of **1c** as a function of nucleophile concentration.

we observed a complex, virtually inseparable mixture of products, and did not detect the expected cyclic product by HPLC/MS. A representative LCMS chromatogram, which was obtained for the reaction between 1.5 mM peptide **1** and 50 mM succinic dihydrazide at 60 °C is illustrated in the Fig. 4.4.

Encouraged by the results of our model studies, we decided to further explore the substrate scope, the nucleophile influence, and the effects of the cycle size on the outcome of the reaction (Fig. 4.5). To this end, we prepared two more peptides: the first, NH<sub>2</sub>-GAXPYLVKSXFG-CONH<sub>2</sub> (**2**), had two hydrazides separated by six amino acids (*i, i+7* cycle), and the second, NH<sub>2</sub>-GALPYXVKSEFV-CONHNH<sub>2</sub> (**3**), was a peptide containing a single glutamic acid  $\gamma$ -hydrazide and a valine C-terminal hydrazide separated by five amino acids to study the possibility of a side-chain to C-terminus cyclization. As linkers for the cyclization, we chose the commercially available dihydrazides of variable reactivity, size and rigidity. Carbohydrazide (**a**) and oxalyldihydrazide (**b**) represented short, rigid, and generally less reactive linkers, while succinic dihydrazide (**c**) and adipic dihydrazide (**d**) were longer staples with several sp<sup>3</sup> carbon centers to allow for greater flexibility of resulting constructs.

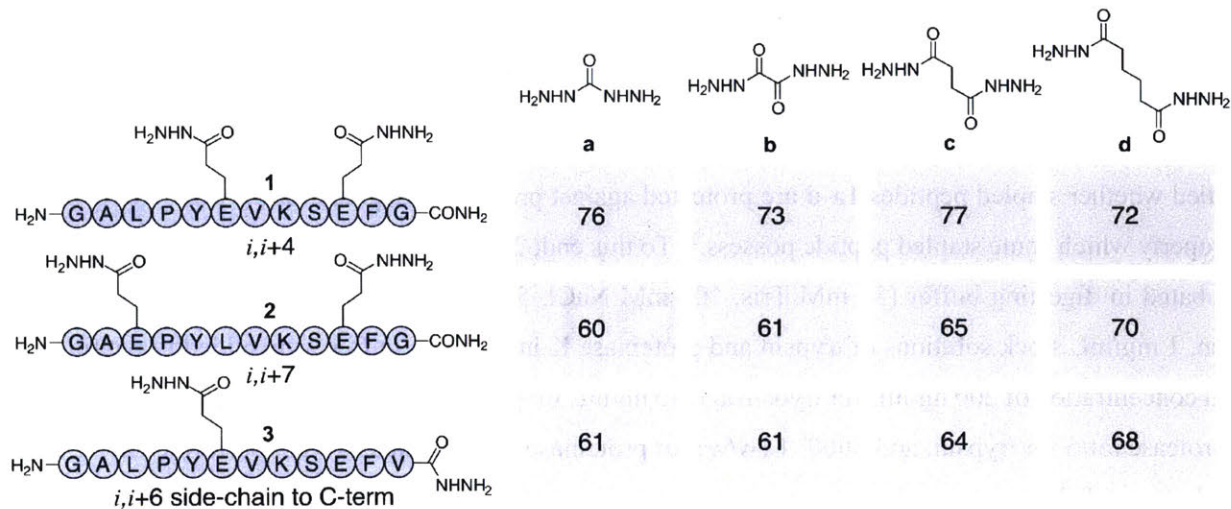
As summarized in Fig. 4.5, we performed a total of twelve reactions to study the cyclization. The peptides were incubated with 5 mM nucleophiles essentially following the protocol outlined above, and the reaction mixtures were analyzed by HPLC/MS. Additionally, peptides **1** and **1a-d** were isolated by RP-HPLC and characterized by <sup>1</sup>H and <sup>13</sup>C NMR to confirm the product identity and purity (spectra are shown in Section 4.3.9). In all cases, we observed smooth cyclization with similar reaction profiles, although reaction yields varied for different constructs. Peptide **1**, which had isocyanates spaced in the (*i, i+4*) manner, generally cyclized more efficiently than peptides **2** and **3**, indicating, not surprisingly, that the cyclization is more robust for smaller ring sizes. Interestingly, we also found that the side-chain to C-terminus cyclization of **3** proceeded nearly as efficiently as the side-chain to side-chain cyclization of **2** for all studied linkers which suggests that the cycle size is the primary factor influencing the reaction yield. From the linker standpoint, we found no major discrepancy in reactivity between four studied dihydrazides, although the cyclization with linker **c** generally led to higher product yields than the other linkers, possibly due to a more optimal size and substantial flexibility of the resulting ring.



**Figure 4.4. Attempts to macrocyclize acyl azides**

HPLC-MS (TIC) chromatogram of the crude reaction mixture between double acyl azide of peptide **1** and succinic dihydrazide. None of the major products correspond to the expected cyclized peptide.





**Figure 4.5. Substrate and linker scope of the isocyanate macrocyclization**

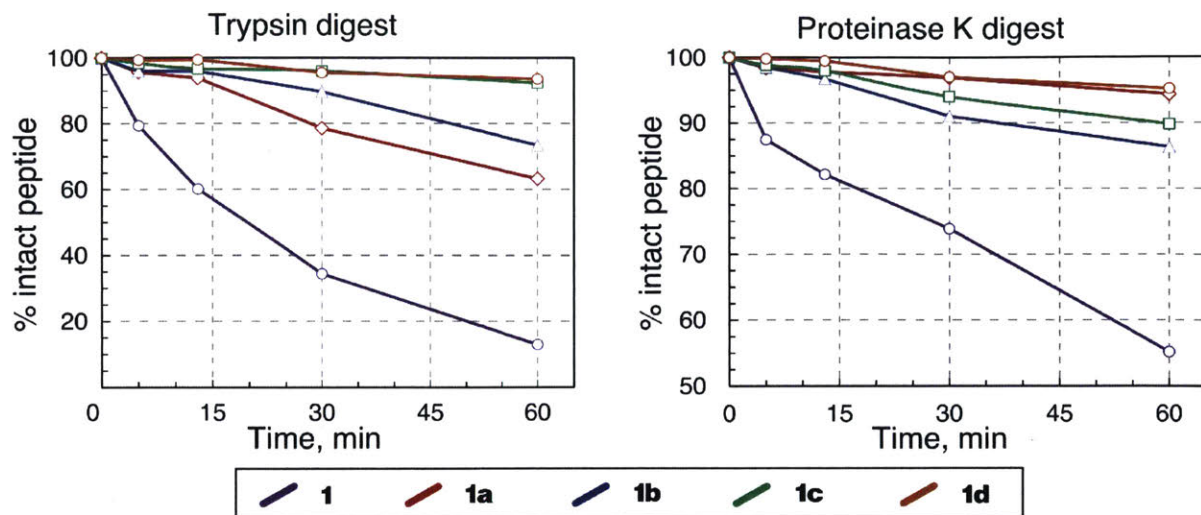
Peptides were oxidized with 20 mM NaNO<sub>2</sub> at -10 °C for 10 minutes, diluted into a 5 mM solution of the nucleophile in water/acetonitrile (1:3, v/v) mixture and incubated at 70 °C for 10 minutes. Numbers in each cell represent HPLC/MS yields of the expected macrocyclized product formed from a corresponding peptide and a linker. Total ion chromatograms were manually integrated and yields were determined from relative peak areas.

## 4.2.2. Properties of Macrocyclic Peptides

Having established the scope of the reaction, we turned to investigating how the biological properties of these unusually cyclized peptides with double N'-acylated semicarbazide tethers are affected. We first studied whether stapled peptides **1a-d** are protected against proteolytic degradation by digestive proteases, a property which some stapled peptide possess.<sup>27</sup> To this end, 200 µg/mL solutions of peptides **1**, **1a-d** were incubated in digesting buffer (50 mM Tris, 100 mM NaCl, 5 mM CaCl<sub>2</sub> buffer, pH 7.4) for 30 minutes. Then, 1 mg/mL stock solutions of trypsin and proteinase K in the same buffer were added to peptides to a final concentration of 200 ng/mL for trypsin and 40 ng/mL for proteinase K (1000: 1 (w/w) ratio of substrate to protease ratio for trypsin, and 5000: 1 (w/w) for proteinase K). The digests were performed at 37 °C for one hour, and three intermediate time points were taken at 5, 13 and 30 minutes. The reactions were quenched by adding excess 50% A/50% B and analyzed by HPLC-MS. To quantitate the extent of proteolysis, chromatograms were manually integrated and the percentage of intact peptide was plotted as a function of time. As shown in Fig. 4.6, all stapled peptides clearly demonstrated improved stability to both trypsin and proteinase K as compared to the unstapled control **1**. For example, after one hour incubation with trypsin, only 13% of unstapled peptide **1** remained undigested, which is in stark contrast with 94% remaining for **1d**. Generally, we found that longer linkers provided better protection against digestion: peptides stapled with adipic dihydrazide linker were most stable towards proteolysis.

Separately, we also investigated the chemical stability of macrocyclized peptides under different conditions. We found that peptides **1a-d** were completely stable to both acidic (pH 2.0) and basic (pH 9.5) conditions as well as to the action of reducing (50 mM TCEP, 50 mM DTT) and oxidizing (50 mM NaNO<sub>2</sub>) agents after 24 h incubation. Taken together, these data suggest that the tethering with unnatural linkers does not compromise the chemical stability of a peptide, and moreover, renders it more resistant to proteolysis.

To study the effects of macrocyclization on biological activity and conformational behavior of a peptide, we prepared peptide NH<sub>2</sub>-ITFXDLLXYYGKKKK(Biotin)-CONH<sub>2</sub> (an NYAD-1 analogue, **4**), a known binder of the C-terminal domain of HIV-1 capsid assembly polyprotein (C-CA).<sup>28</sup> Earlier, this peptide was successfully stapled using olefin metathesis and cysteine perfluoroarylation approaches.<sup>14,29,30</sup> In both studies, researchers found that the peptide cyclization promoted  $\alpha$ -helical conformation and improved binding to C-CA. As illustrated in Fig. 4.7, we cyclized peptide **4** with linkers **a-d** under the standard conditions to obtain a total of four stapled NYAD-1 analogues, **4a-d**. To probe the influence of the stapling on the conformation of the peptides, we turned to far-UV CD spectroscopy (Fig. 4.7a). Although all tested NYAD variants displayed at least some degree of  $\alpha$ -helicity as evident by characteristic minima at 208 and 222 nm, the nature of the linker influenced the peptide conformation. Stapling with succinic di-hydrazide



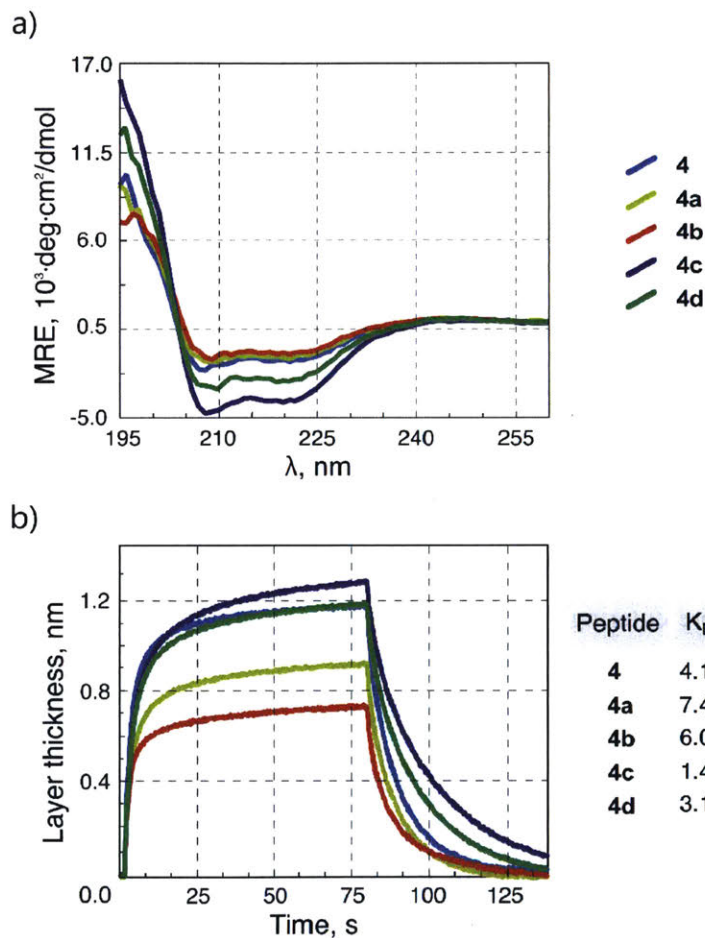
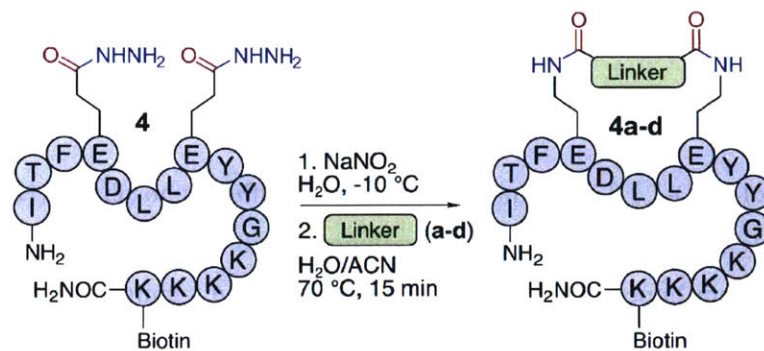
**Figure 4.6. Proteolytic stability of macrocyclic peptides**

Comparative proteolytic degradation of peptides **1** and **1a-d** with trypsin and proteinase K. The plots are the fraction of undigested peptides as a function of digest time.

increased the  $\alpha$ -helical character of the peptide, from 11% for unstapled peptide **4** to 26% for **4c**. NYAD-1 analogue **4d** stapled with longer and more flexible adipic also led to increase in helicity of the peptide, albeit to a lesser extent: the  $\alpha$ -helicity of **4d** was estimated at 19%. In stark contrast, cyclization with short and rigid linkers **a** and **b** resulted in practically no change in the CD spectrum as compared to peptide **4**. Consistent with previous studies,<sup>20</sup> these observations indicate that both linker length and rigidity are important considerations for increasing  $\alpha$ -helical content of a cyclized peptide.

Finally, we evaluated how isocyanate macrocyclization affects the binding affinity of peptide **4** towards C-CA by measuring binding kinetics using bi-layer interferometry on Octet RED96 system (detailed protocols and experimental data are provided in Sections 4.3.5 and 4.3.8). Consistent with previous reports,<sup>14,29</sup> the binding affinity of NYAD-1 analogues directly correlated with the  $\alpha$ -helicity of the peptides. As indicated in Fig 4.7b, we observed approximately threefold improvement in the binding constant for the most helical analogue **4c** over the unstapled control **4** ( $K_D$  (**4c**) = 1.4  $\mu$ M versus  $K_D$  (**4**) = 4.1  $\mu$ M). At the same time, the least helical stapled peptides **4a** and **4b** bound C-CA less tightly than the unstapled peptide with  $K_D$  values of 7.4  $\mu$ M and 6.0  $\mu$ M, respectively. Thus, the five NYAD-1 analogues demonstrated a range of biological properties, such as  $\alpha$ -helicity and binding affinity based on the nature of the dihydrazide linker. These results indicate that it is possible to fine-tune biochemical properties of a peptide by modulating the nature of a dihydrazide linker used for cyclization.

In conclusion, we developed a unique approach to peptide macrocyclization, which relies on conjugating peptide isocyanates with commercially available dicarboxylic acid hydrazides. Fully unprotected peptide isocyanates are readily available via a Curtius rearrangement of corresponding acyl azides, and react with bifunctional nucleophiles in aqueous solvents in a chemo- and regioselective fashion. Our method complements existing tools by providing a two-component macrocyclization platform that does not rely on cysteine modification or expensive customized amino acids, and provides an access to abiotic, hydrophilic semicarbazide-bearing constructs unobtainable by existing approaches. This methodology allows the construction of rings of variable size, rigidity and topology, including side-chain to side-chain and side-chain to C-terminus cyclizations. Resulting stapled peptides proved to be chemically stable and showed improved proteolytic stability compared to their linear analogues. We also demonstrated that the biological activity of a peptide can be enhanced and fine-tuned using our approach. We believe that this macrocyclization platform will expand the design scope for developing novel biologically active macrocyclized peptides.



**Figure 4.7. The effects of the macrocyclization on the biological properties of NYAD-1 and its macrocyclized analogues**

The peptide was oxidized with 20 mM  $\text{NaNO}_2$  at  $-10\text{ }^\circ\text{C}$  for 10 minutes, diluted into a 5 mM solution of the nucleophile in water/acetonitrile (1:3, v/v) mixture and incubated at  $70\text{ }^\circ\text{C}$  for 10 minutes to prepare the cyclic constructs **4a-d**. a) CD spectra of peptides **4**, **4a-d**. b) Bio-layer interferometry binding sensograms of immobilized **4**, **4a-d** with  $10\text{ }\mu\text{M}$  C-CA.  $K_D$  values in  $\mu\text{M} \pm$  one standard deviation are displayed.

## 4.3. Experimental

### 4.3.1. General

2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-(7-aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU), and Fmoc-L-Glu(All)-OH were purchased from Chem-Impex International. N<sup>α</sup>-Fmoc protected L-amino acids were obtained through Advanced ChemTech, and Novabiochem. Boc-Gly-OH and Boc-Ile-OH were from Midwest Bio-Tech.

H-Rink Amide-ChemMatrix resin (100-200 mesh, 0.45 mmol/g loading) was purchased from PCAS BioMatrix Inc. 2-Chlorotrityl chloride resin (200-400mesh, 1.2 mmol/g) was purchased from Chem-Impex International and was subsequently used to prepare 2-chlorotrityl hydrazine resin adhering to the published protocols.<sup>31</sup> The loading of the hydrazide resin was determined to be 0.75 mmol/g. *N,N*-Dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile were from EMD Millipore. Triisopropyl silane (TIPS), and 1,2-ethanedithiol (EDT) were from Alfa Aesar. Solvents for HPLC-MS were purchased from Fluka. All other reagents were purchased from Sigma-Aldrich.

### 4.3.2. Fmoc Solid Phase Peptide Synthesis

Peptides were synthesized with the previously described<sup>25</sup> flow-based peptide synthesis platform. H-Rink Amide ChemMatrix resin (200 mg) was used to synthesize all peptides except for peptide 3, for which 200 mg 2-chlorotrityl hydrazide resin was used. During synthesis, all solvents and reagents were preheated to 60 °C immediately before entering the synthesis vessel. The following 120-second cycle was used to synthesize all peptides: amide bond formation (coupling) was performed over 40 seconds at 6 mL/min. Removal of coupling reagent (wash) was done in 30 s at 20 mL/min, followed by 20 s at 20 mL/min to carry out N<sup>α</sup>-Fmoc removal (deprotection). Finally, the second was to remove the deprotection reagent and products was achieved in 30 s at 20 mL/min.

*Coupling.* Coupling solution, consisting of 0.4 M N<sup>α</sup>-Fmoc and side chain protected amino acid dissolved in 0.38 M HATU in DMF, was delivered at 6 mL/min for 30 s and left in the reactor for an additional 10 s. The amino acid was activated with 500 μL of DIEA two minutes prior to delivery. When coupling any valine or isoleucine residues, synthesis was paused for 10 min following delivery. Increased reaction time was necessary to quantitatively couple these sterically hindered residues. Side chain protection was as follows: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Trt), Gln(Trt), Glu(OAll), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu). In addition, the N-terminus amino acid of each peptide was N<sup>α</sup>-Boc protected to prevent guanidinylation by HBTU during side-chain hydrazide formation.

*Wash.* Excess reagent and reaction by-products were washed out from the synthesis vessel with 10 mL of DMF delivered at 20 mL/min over 30 s.

*Deprotection.* N<sup>α</sup>-Fmoc protecting groups were removed with 6.6 mL of 20% (v/v) piperidine in DMF delivered at 20 mL/min over 20 s.

All peptides were cleaved from the resin and side chain deprotected with a standard cleavage cocktail of 2.5% (v/v) EDT, 2.5% (v/v) H<sub>2</sub>O, and 1% (v/v) TIPS in TFA for 7 minutes at 60 °C. In all cases, compressed nitrogen was used to evaporate the cleavage solution to dryness after filtering the resin. The resulting mixtures were washed three times with cold diethyl ether, dissolved in 50% water / 50% acetonitrile (0.1% TFA added), and lyophilized.

All peptides were purified by RP-HPLC on a Waters 600 HPLC system with a Waters 484 or 486 UV detector using water with 0.1% TFA added (solvent A) and acetonitrile with 0.1% TFA added (solvent B) as solvents. An Agilent Zorbax 300SB preparative C3 column (300 Å, 7 μm, 21.2 x 250 mm) was used for purification of preparative scale syntheses of peptide **1** and peptides **1a-1d**. The following gradient was used at a flow rate of 20 mL/min: 5% B in A for 10 minutes, followed by 5%-10% B ramping linearly over 5 minutes, and 10%-50% B ramping linearly over 120 minutes. Fractions were collected and screened for the desired material using HPLC-ESI-TOF. An Agilent Zorbax 300SB semi-preparative C3 column (300 Å, 5 μm, 9.4 x 250 mm) was used for purification of all other peptides with the following gradient at a flow rate of 5 mL/min: 5% B in A for 5 minutes, then 5%-35% B ramping linearly over 120 minutes. Fractions were collected and screened for the desired material using HPLC-ESI-TOF.

#### 4.3.3. Peptide Macrocyclization

Unless noted, macrocyclization reactions were performed under the following standard conditions: 3.07 mM peptide hydrazide solution in oxidation buffer (200 mM Na<sub>2</sub>HPO<sub>4</sub> and 6 M Gn·HCl in water, pH 3.0-3.2) was incubated in an ice-salt bath at -15 °C for 5 min. Then, 200 mM NaNO<sub>2</sub> solution in water was carefully added to the peptide to a final NaNO<sub>2</sub> concentration of 18 mM. The reaction was allowed to proceed for 10 minutes at -10 °C. Then, a 5 mM solution of bifunctional nucleophile in water/acetonitrile (1: 3, v/v, 0.1% TFA added, pH ~ 2.5-3.5 depending on the nucleophile) was added to the reaction mixture to a final peptide concentration of 80-100 μM and the mixture was thermostated at 70 °C for 10 minutes.

Preparative scale reactions were performed in a similar manner unless otherwise noted. Crude reaction mixtures were frozen in liquid nitrogen, lyophilized, redissolved in 95% A/5% B and purified essentially as described above.

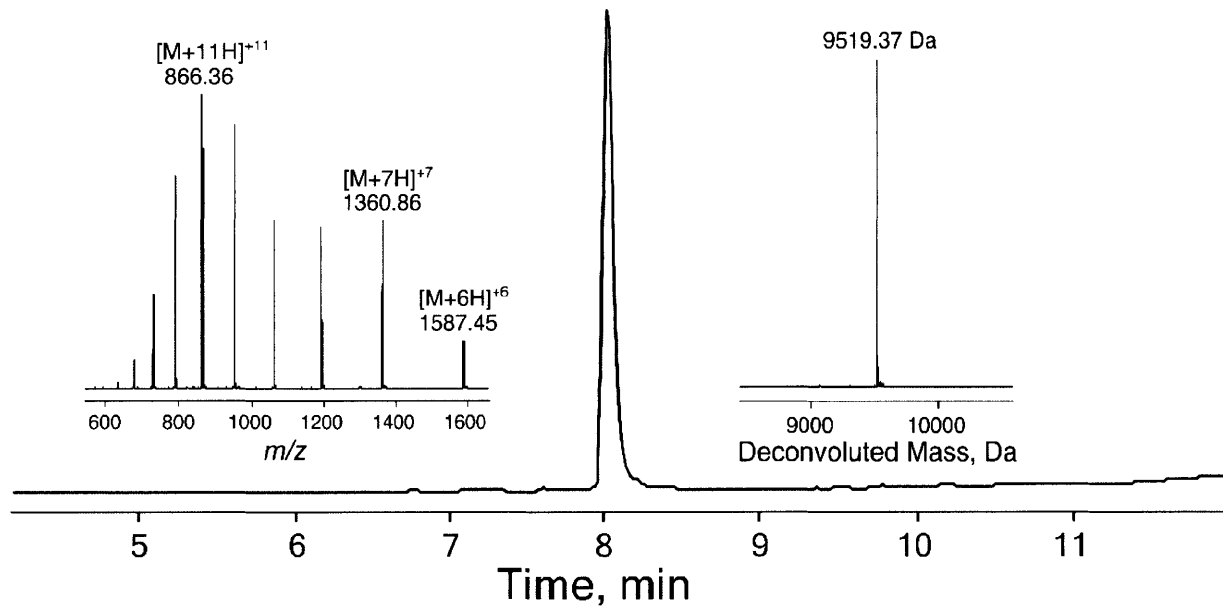
#### 4.3.4. Protein Expression

Full-length HIV1 Core Antigen was obtained from Invitrogen. CCA sequence was cloned into pET-SUMO plasmid (Thermo Fisher) and transformed into *E. coli* DH5 $\alpha$  cells. The protein was expressed in *E. coli* BL21 DE3 carrying the pET-SUMO plasmid from LB medium supplied with 30  $\mu$ g/mL kanamycin. The culture was induced with 0.4 M IPTG upon reaching OD600 = 0.6.

Following overnight expression at 37 °C, cells were pelleted and resuspended in 50 mL Tris buffer (50 mM Tris, 150 mM NaCl, pH 7.4) containing 20 mg lysozyme, 2 mg DNase I, and a half tablet cComplete Mini protease inhibitor cocktail (Roche Life Science). After sonication, the suspension was pelleted (25 minutes, 17000 rpm) and supernatant loaded onto a 5mL GE HisTrap FF crude Ni-NTA column pre-equilibrated with binding buffer (20 mM Tris, 150 mM NaCl, pH 8.5). The column was washed with 30 mL binding buffer and then 30 mL binding buffer containing 40 mM imidazole. His<sub>6</sub>-SUMO-CCA fusion protein was eluted with 50 mL 500 mM imidazole in the binding buffer and buffer exchanged to remove imidazole using a HiPrep 26/10 Desalting column (GE Healthcare, UK) on ÄKTAprime plus FPLC system (GE Healthcare, UK).

Protein concentration was determined by measuring absorbance at 280 nm (MW = 22916.9 Da,  $\epsilon$  = 9970 as found on ExPASy ProtParam (<http://web.expasy.org/protparam/>)); His<sub>6</sub>-SUMO cleavage was subsequently performed by incubating SUMO-CCA overnight at 4 °C with 1 ng SUMO protease per 1 mg protein. The resulting mixture was loaded onto a pre-equilibrated 5mL GE HisTrap FF crude Ni-NTA column, and washed with 10 mL 40 mM imidazole in Tris buffer (pH 8.0) to isolate CCA. The protein was further buffer exchanged to remove imidazole using the HiPrep 26/10 Desalting column into Tris buffer at pH 7.4, concentrated over a 3000 Da Amicon Ultra-15 Centrifugal Filter Unit (EMD Millipore), analyzed by HPLC-MS (Fig. 4.8), and flash frozen in liquid nitrogen for long-term storage at -80 °C.





**Figure 4.8. HPLC-MS analysis of expressed C-CA**

HPLC-MS (TIC chromatogram with MS insets and MS deconvolution result) characterization of recombinant C-CA. Calculated molecular weight: 9518.9 Da.

### 4.3.5. Analytical Techniques

*HPLC-MS analysis.* Unless noted, all peptides, proteins and reaction mixtures were analyzed on an Agilent 6520 Accurate Mass Q-TOF LC-MS using an Agilent Zorbax 300SB C3 column (300 Å, 5 µm, 2.1 x 150 mm) with the following method. At 40 °C and a flow rate of 0.8 mL/min, the following gradient was used: 1% acetonitrile with 0.1% formic acid added (FA, solvent B') in water with 0.1% FA (solvent A') for 2 min, 1-61% B' in A' ramping linearly over 11 min, 61% B' in A' for 1 minute. Typically, 100 – 1000 µg/mL solutions of peptides and proteins were subject to analysis. Unless noted, all chromatograms shown in this work are plots of total ion current (TIC) versus time. Yields for each product were calculated as an area fraction of the total integrated area. Performed in this way, such analysis agreed well with manual integration of UV<sub>214</sub> traces (<5% relative difference). The data were analyzed using Agilent MassHunter Qualitative analysis software. MS deconvolution spectra were obtained using the maximum entropy algorithm.

*Circular dichroism spectroscopy.* All CD spectra were acquired on an Aviv 202 Circular Dichroism Spectrometer in a 1 mm quartz cuvette from 260 to 190 nm with fifteen-second averaging times at each wavelength. Peptide solutions (0.5 mg/mL) were prepared by dissolving solid samples in a 25% acetonitrile / 75% phosphate buffer (50 mM Na<sub>2</sub>SO<sub>4</sub>, 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4). Data processing included solvent background correction (subtraction) and adjustment for pathlength and concentration (*mean residue ellipticity*,  $MRE = [\theta]\lambda = \frac{\theta_{obs}}{10lcn}$ , where  $\theta_{obs}$  is measured ellipticity,  $\theta$  — mean residual ellipticity in deg·cm<sup>2</sup>·dmol<sup>-1</sup>,  $l$  — pathlength in cm;  $c$  — concentration of peptide in M;  $n$  — # of amino acids).  $\alpha$ -helicities of the peptides were estimated using previously established methods.<sup>17</sup>

*Bio-layer interferometry.* Binding kinetics for the C-CA/NYAD-1 interaction were measured on Octet RED96 instrument from ForteBio. All measurements were performed at 30° C in a 96-well plate shaking at 1000 rpm. In general, streptavidin-coated tips (from ForteBio) were dipped in binding buffer (1x phosphate buffered saline, 1% w/v bovine serum albumin, 0.02% tween-20, pH 7.4) for 10 minutes to remove sucrose coating. A typical experiment consisted of five steps. First, base line reading was performed by dipping the tips in a well containing fresh binding buffer for 60 sec. Next, ligand loading was performed by associating ~100 µM solutions of biotinylated NYAD-1 analogues in binding buffer onto the tips for 90-300 seconds. Second base line reading was achieved by dipping the tips in a well containing fresh binding buffer for 90-120 sec. For the association step, NYAD-1 loaded tips were dipped into well containing variable concentrations of C-CA (100 nM – 10 µM) in binding buffer for 90-300 seconds, and finally, to measure dissociation, the tips were dipped back into wells containing binding buffer only for 300 seconds or until the response decreased back to the base line level.

Double referencing using wells containing no C-CA and tips loaded with no ligands was used to subtract the background. Reference readings were subtracted universally from both association and dissociation curves. Association curves were aligned to the last five seconds of the base line, and dissociation curves were aligned to the last value of association curves. No curve-smoothing filtering was performed. The resulting curves were fit to the following equations:

$$\text{For association: } \text{response} = R_{max} \frac{1}{1 + \frac{k_d}{k_a [CCA]}} (1 - e^{-(k_a [CCA] + k_d)t})$$

where  $R_{max}$  is a fitting parameter,  $k_a$  and  $k_d$  are association and dissociation constants respectively,  $[CCA]$  is the concentration of C-CA protein, and  $t$  is aligned time.

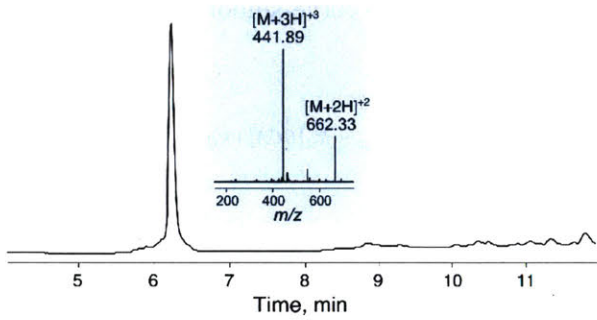
$$\text{For dissociation: } \text{response} = R_0 e^{-k_d t'}$$

where  $R_0$  is the last response value obtained from the first equation,  $t'$  is aligned time for the dissociation step. A BFGS Quasi-Newton minimization algorithm on MATLAB (fminunc) was used to find parameters  $k_d$ ,  $k_a$ , and  $R_{max}$  that minimized the squared difference between calculated and observed responses during association and dissociation. The binding constant,  $K_D$ , was calculated from the fitting parameters  $k_a$  and  $k_d$  as the ratio of the two. Reported  $K_D$  values are average of five measurements and reported errors are standard deviations from these values. Section 4.3.8 shows sensograms and fitting for each tested construct.

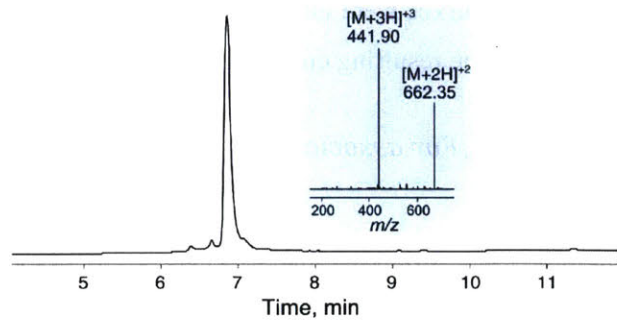
*Nuclear magnetic resonance.*  $^1\text{H}$  NMR spectra were recorded on a Bruker Avance 600 MHz spectrometer and are reported in ppm using solvent as an internal standard (DMSO- $d_6$  at 2.54 ppm). Data are reported as (s = singlet, d = doublet, dd = doublet of doublets, t = triplet, q = quartet, m = multiplet, br = broad; coupling constant(s) in Hz; integration). Proton-decoupled  $^{13}\text{C}$  NMR spectra were recorded on a Bruker Avance 600 MHz (150 MHz) spectrometer and are reported in ppm using solvent as an internal standard (DMSO- $d_6$  at 39.52 ppm).

### 4.3.6. Peptide Synthesis: Analytical Data

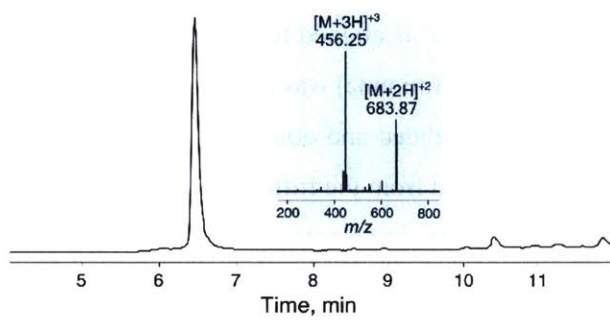
H<sub>2</sub>N-GALPYE(CONHNH<sub>2</sub>)VKSE(CONHNH<sub>2</sub>)FG-  
CONH<sub>2</sub> (1)



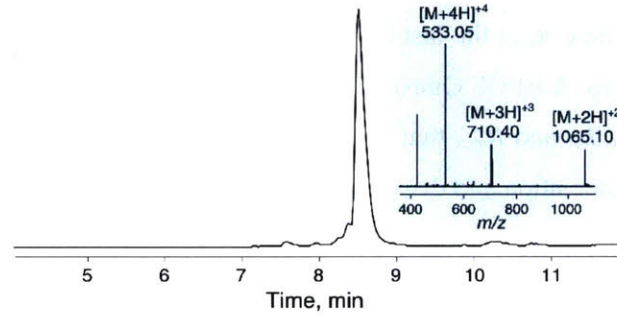
H<sub>2</sub>N-GAE(CONHNH<sub>2</sub>)PYLVKSE(CONHNH<sub>2</sub>)FG-  
CONH<sub>2</sub> (2)



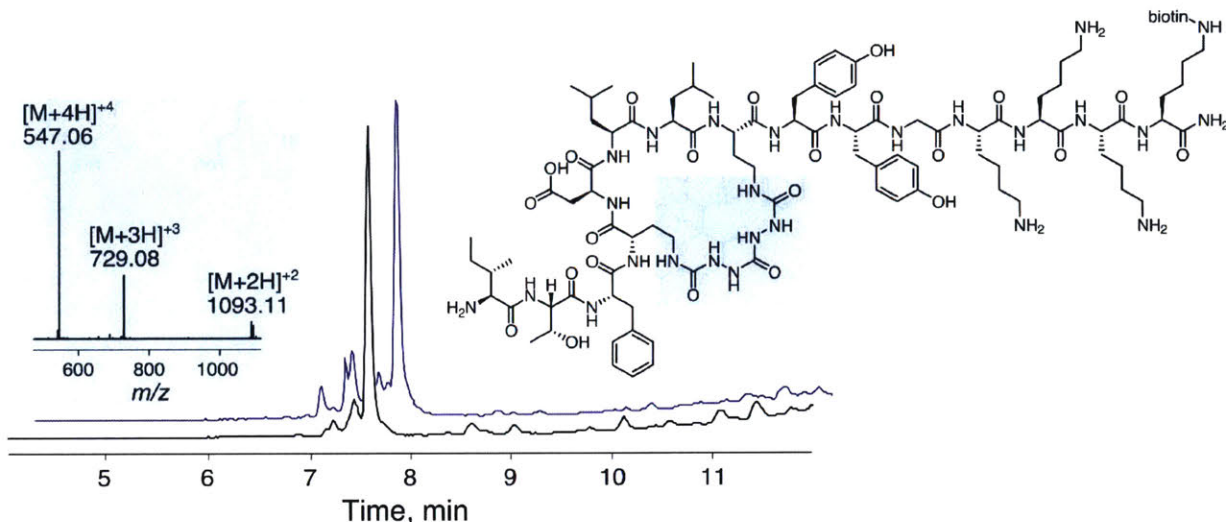
H<sub>2</sub>N-GALPYE(CONHNH<sub>2</sub>)VKSE(COOH)FG-  
CONHNH<sub>2</sub> (3)



H<sub>2</sub>N-ITFE(CONHNH<sub>2</sub>)DLLE(CONHNH<sub>2</sub>)YYGK-  
KKK(biotin)-CONH<sub>2</sub> (4)

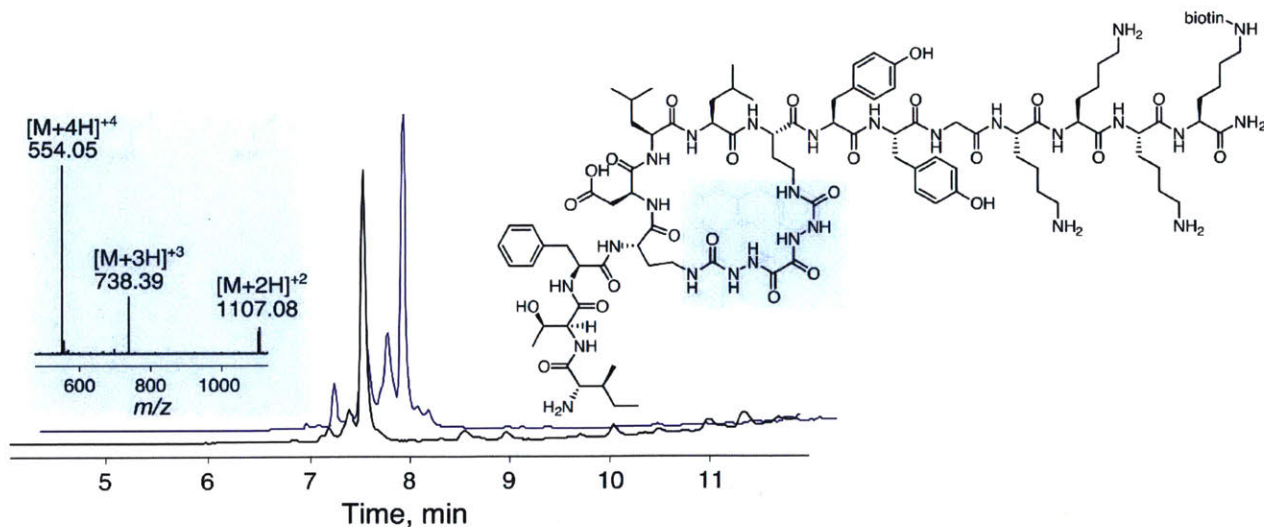


### 4.3.7. Peptide Macrocyclization: Analytical Data



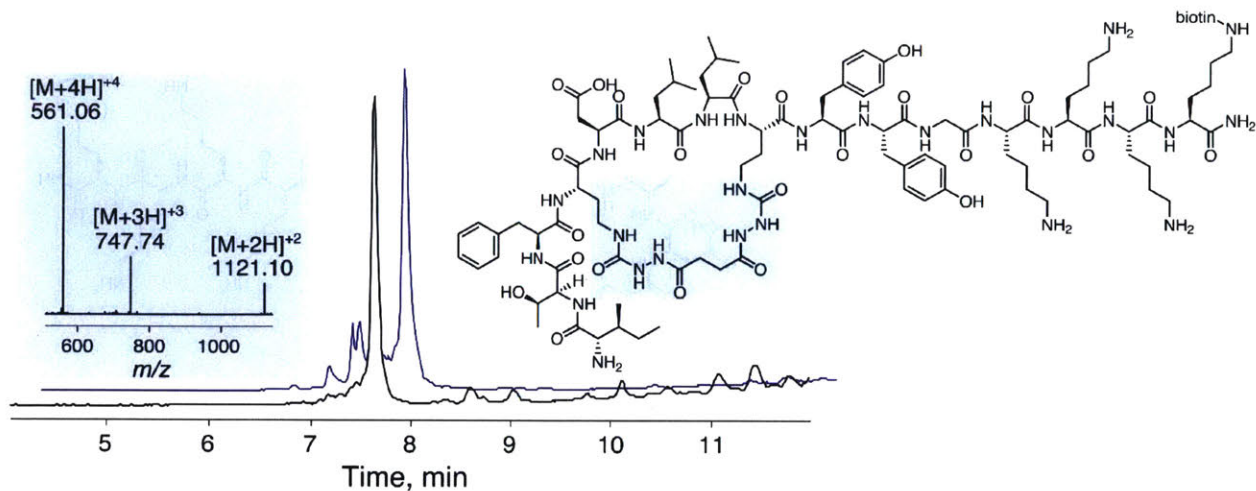
**Figure 4.9. HPLC-MS analysis of macrocyclized peptide 4a**

9.4 mg of peptide **4** were oxidized with sodium nitrite and cyclized with 5mM carbohydrazide as described in section 4.3.3. RP-HPLC purification of the crude reaction mixture yielded 1.3 mg of pure **4a** after lyophilization. HPLC-MS (TIC) chromatogram for the crude and purified cyclized peptide **4a** with MS inset on the left and the peptide structure is shown on the right. The semicarbazide tether is highlighted. Crude macrocyclization reaction mixture chromatogram is shown in blue; purified product — in black. Calc. monoisotopic mass = 2183.19 Da.



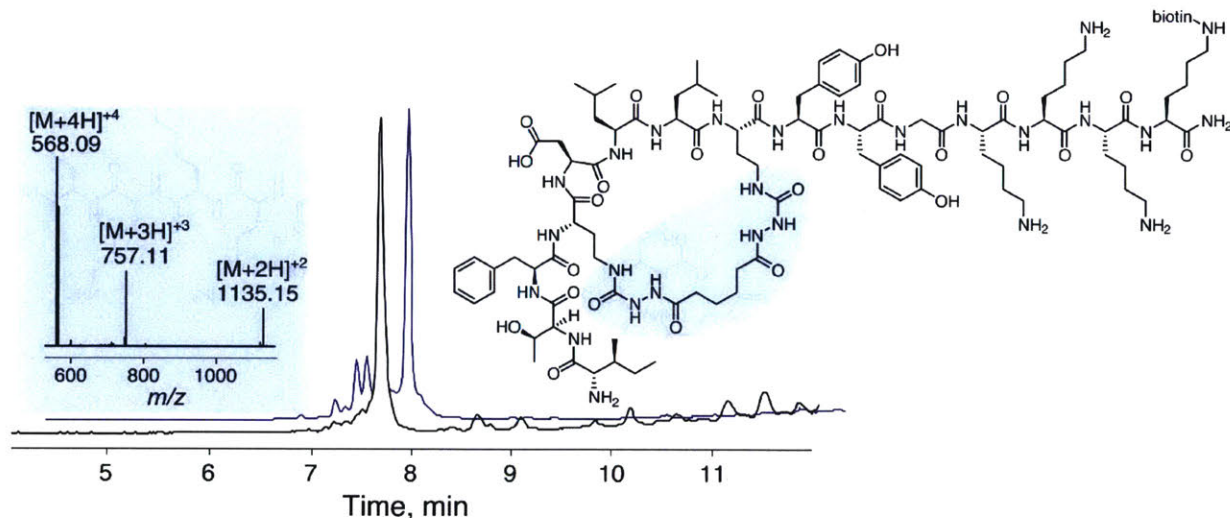
**Figure 4.10. HPLC-MS analysis of macrocyclized peptide 4b**

9.4 mg of peptide **4** were oxidized with sodium nitrite and cyclized with 5 mM oxalyldihydrazide as described in section 4.3.3. RP-HPLC purification of the crude reaction mixture yielded 1.0 mg of pure **4b** after lyophilization. HPLC-MS (TIC) chromatogram for the crude and purified cyclized peptide with MS inset on the left and the peptide structure is shown on the right. The semicarbazide tether is highlighted. Crude macrocyclization reaction mixture chromatogram is shown in blue; purified product — in black. Calc. monoisotopic mass = 2211.16 Da.



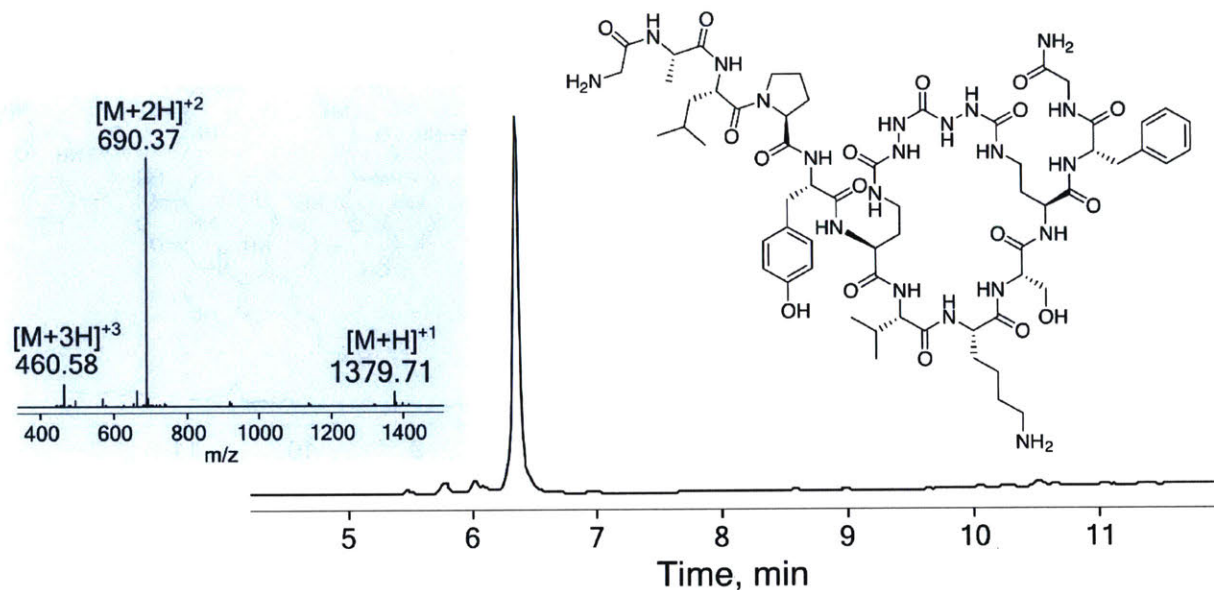
**Figure 4.11. HPLC-MS analysis of macrocyclized peptide 4c**

9.4 mg of peptide **4** were oxidized with sodium nitrite and cyclized with 5 mM succinic dihydrazide as described in section 4.3.3. RP-HPLC purification of the crude reaction mixture yielded 1.6 mg of pure **4c** after lyophilization. HPLC-MS (TIC) chromatogram for the crude and purified cyclized peptide **4c** with MS inset on the left and the peptide structure is shown on the right. The semicarbazide tether is highlighted. Crude macrocyclization reaction mixture chromatogram is shown in blue; purified product — in black. Calc. monoisotopic mass = 2239.19 Da.



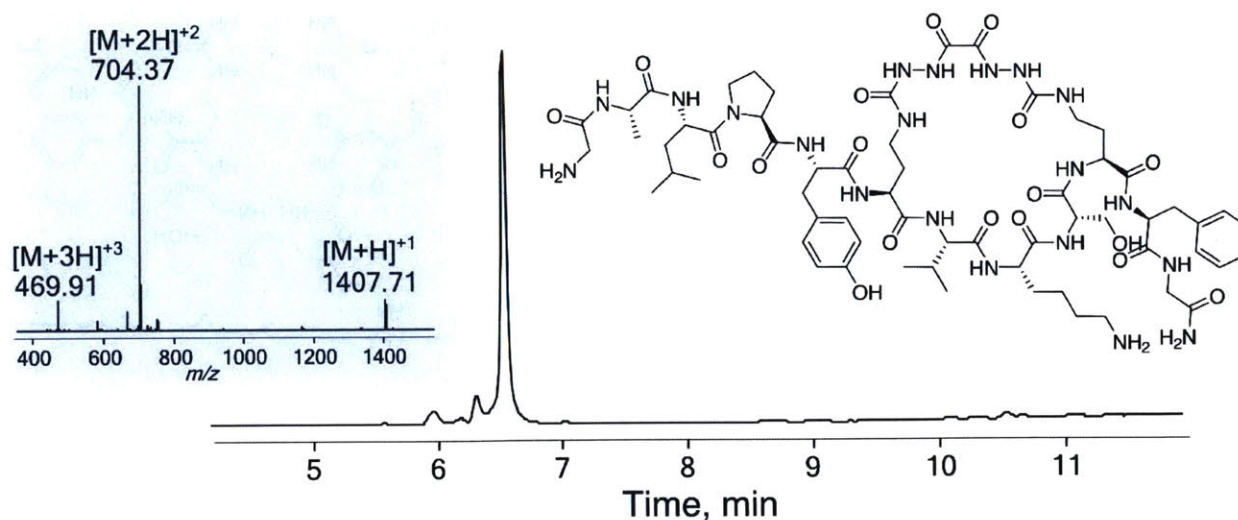
**Figure 4.12. HPLC-MS analysis of macrocyclized peptide 4d**

9.4 mg of peptide **4** were oxidized with sodium nitrite and cyclized with 5 mM adipic acid dihydrazide as described in section 4.3.3. RP-HPLC purification of the crude reaction mixture yielded 1.3 mg of pure **4d** after lyophilization. HPLC-MS (TIC) chromatogram for the crude and purified cyclized peptide **4d** with MS inset on the left and the peptide structure is shown on the right. The semicarbazide tether is highlighted. Crude macrocyclization reaction mixture chromatogram is shown in blue; purified product — in black. Calc. monoisotopic mass = 2267.22 Da.



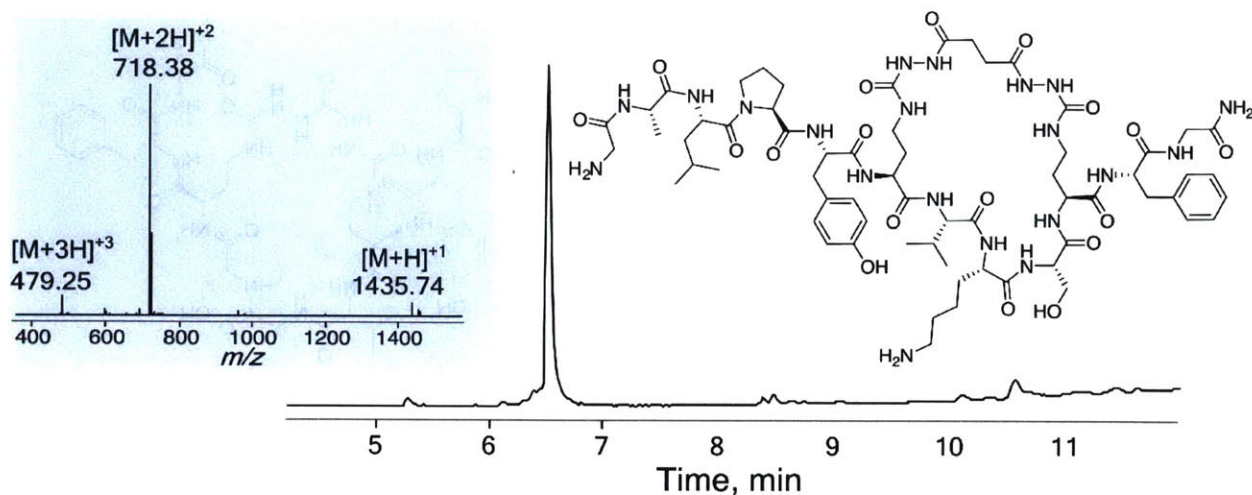
**Figure 4.13. HPLC-MS analysis of macrocyclized peptide 1a**

60.5 mg of peptide **1** were dissolved in 14.9 mL of oxidation buffer, oxidized with 1.49 mL of 200 mM sodium nitrite, and cyclized with 540.1 mL of 5 mM carbonyl dihydrazide as described in section 4.3.3. 17.0 g citrate was added to the reaction mixture and excess acetonitrile was removed with rotary evaporation. The cyclized peptide **1a** was purified by RP-HPLC, yielding 15.4 mg of the desired peptide as a white powder after lyophilization. The peptide was isolated as a trifluoroacetic acid salt. HPLC-MS (TIC) chromatogram for the purified cyclized peptide **1a** with MS inset is shown on the left and the peptide structure on the right. Calc. monoisotopic mass = 1378.71 Da



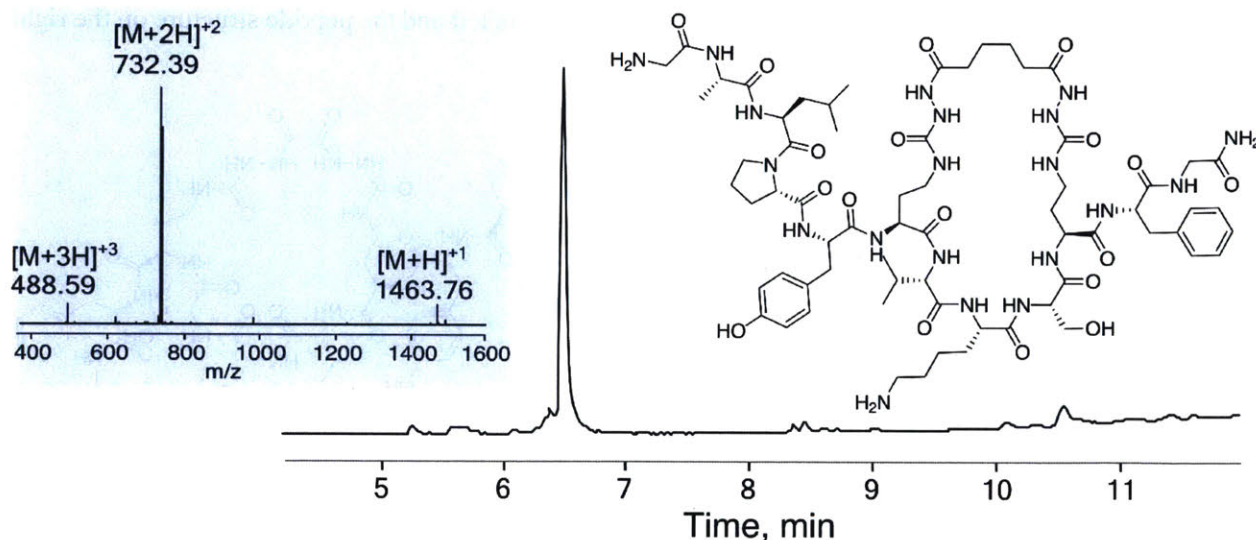
**Figure 4.14. HPLC-MS analysis of macrocyclized peptide 1b**

62.5 mg of peptide **1** were dissolved in 15.4 mL of oxidation buffer, oxidized with 1.54 mL of 200 mM sodium nitrite, and cyclized with 558 mL of 5 mM oxalyl dihydrazide as described in section 4.3.3. 18.5 g citric acid was added to the reaction mixture and excess acetonitrile was removed with rotary evaporation. The cyclized peptide **1b** was purified by RP-HPLC, yielding 17.6 mg of the desired peptide as a white powder after lyophilization. The peptide was isolated as a trifluoroacetic acid salt. HPLC-MS (TIC) chromatogram for the purified cyclized peptide **1b** with MS inset on the left and the peptide structure is shown on the right. Calc. monoisotopic mass = 1406.71 Da.



**Figure 4.15. HPLC-MS analysis of macrocyclized peptide 1c**

41.36 mg of peptide **1** were dissolved in 16.2 mL of oxidation buffer, oxidized with 1.62 mL of 200 mM sodium nitrite, and cyclized with 369 mL of 5 mM succinic dihydrazide as described in section 4.3.3. 11.6 g citric acid monohydrate was added to the reaction mixture and excess acetonitrile was removed by rotary evaporation. The cyclized peptide **1b** was purified with RP-HPLC, yielding 20.5 mg of the desired peptide as a white powder after lyophilization. The peptide was isolated as a trifluoroacetic acid salt. HPLC-MS (TIC) chromatogram for the purified cyclized peptide **1c** with MS inset on the left and the peptide structure is shown on the right. Calc. monoisotopic mass = 1434.74 Da.

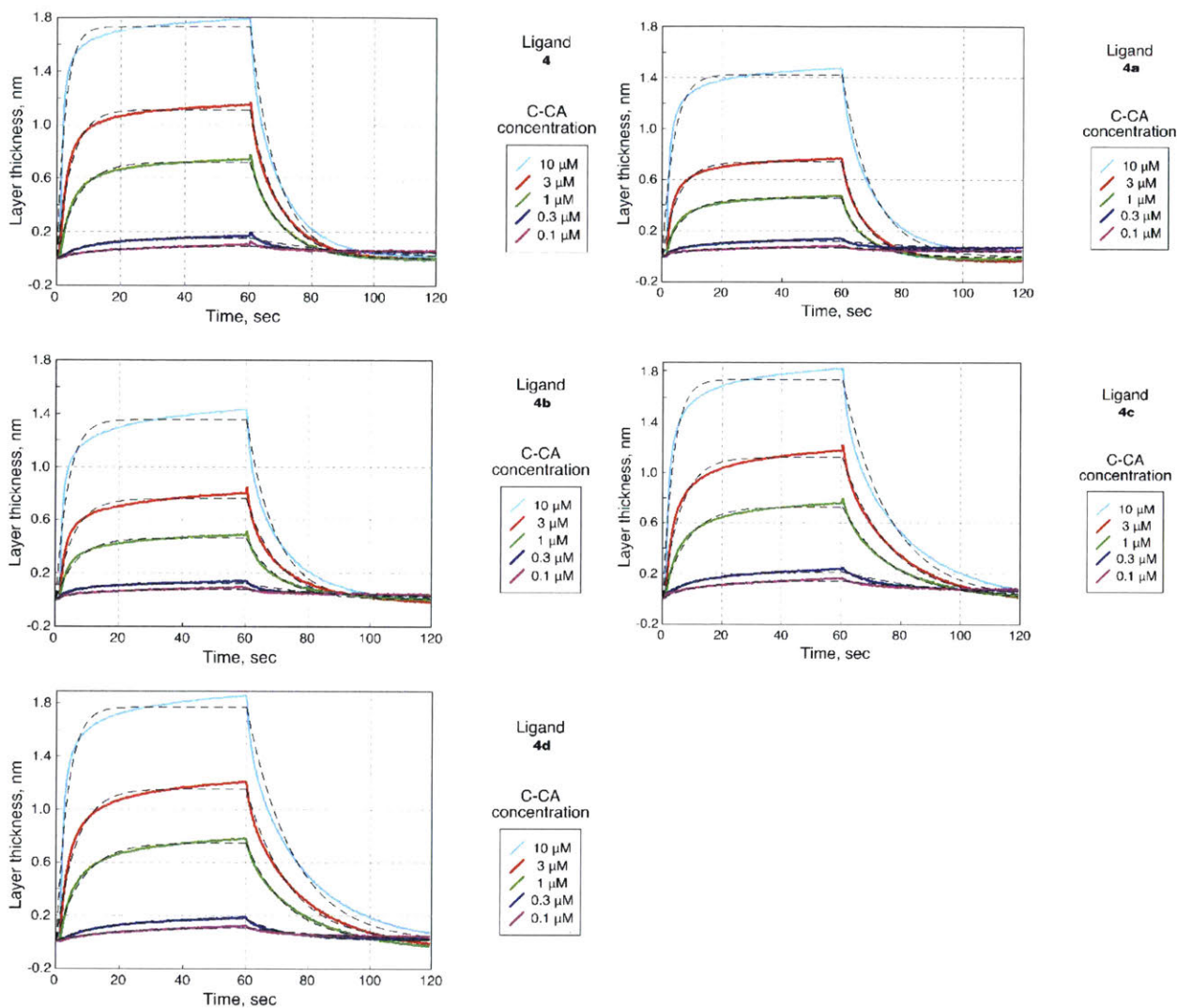


**Figure 4.16. HPLC-MS analysis of macrocyclized peptide 1d**

44.53 mg of peptide **1** were dissolved in 11.0 mL of oxidation buffer, oxidized with 1.1 mL of 200 mM sodium nitrite, and cyclized with 400 mL of 5 mM adipic acid dihydrazide as described in section 4.3.3. 12.5 g citric acid monohydrate was added to the reaction mixture and excess acetonitrile was removed by rotary evaporation. The cyclized peptide was purified by RP-HPLC, yielding 15.5 mg of the desired peptide after lyophilization as white powder. The peptide was isolated as a trifluoroacetic acid salt. HPLC-MS (TIC) chromatogram for the purified cyclized peptide **1d** with MS inset on the left and the peptide structure is shown on the right. Calc. monoisotopic mass = 1462.76 Da.



### 4.3.8. Bio-Layer Interferometry: Analytical Data



### 4.3.9. NMR Spectra

Peptide 1.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  10.52 (s, 2H), 8.53 (d,  $J = 7.4$  Hz, 1H), 8.09 – 8.20 (m, 8H), 7.98 (d,  $J = 6.6$  Hz, 1H), 7.86 – 7.89 (m, 4H), 7.79 (d,  $J = 7.2$  Hz, 1H), 7.23 – 7.25 (m, 5H), 7.19 (d,  $J = 6.8$  Hz, 1H), 7.13 (s, 1H), 7.01 (d,  $J = 8.1$  Hz, 2H), 7.01 (d,  $J = 8.1$  Hz, 2H), 6.62 (d,  $J = 8.1$  Hz, 2H), 4.51 – 4.56 (m, 2H), 4.36 – 4.42 (m, 2H), 4.28 – 4.34 (m, 4H), 4.18 – 4.22 (m, 2H), 3.70 (dd,  $J = 16.8, 6.0$  Hz, 1H), 3.65 (dd,  $J = 10.3, 5.7$  Hz, 1H), 3.60 (d,  $J = 5.1$  Hz, 2H), 3.57 (s, 2H), 3.54 (m, 1H), 3.41 – 3.49 (s, 1H), 3.06 (dd,  $J = 13.6, 4.4$  Hz, 1H), 2.90 (d,  $J = 10.5$  Hz, 1H), 2.71 – 2.82 (m, 4H), 2.22 (m, 2H), 2.15 (m, 2H), 2.01 (m, 1H), 1.89 – 1.97 (m, 2H), 1.83 (m, 4H), 1.71 – 1.78 (m, 2H), 1.69 (m, 1H), 1.62 (m, 1H), 1.54 (t,  $J = 6.9$  Hz, 3H), 1.47 (m, 1H), 1.21 – 1.40 (m, 3H), 1.20 (d,  $J = 6.8$  Hz, 3H), 0.90 (d,  $J = 6.6$  Hz, 3H), 0.88 (d,  $J = 6.3$  Hz, 3H), 0.84 (d,  $J = 6.6$  Hz, 3H).

$^{13}\text{C}$  NMR (150 MHz, DMSO- $d_6$ ):  $\delta$  172.03, 171.72, 171.67, 171.65, 171.58, 171.54, 171.49, 171.43, 171.38, 171.31, 171.11, 170.70, 165.83, 158.88 (d,  $J = 30.8$  Hz), 156.30, 138.08, 130.56, 129.57, 128.55, 127.97, 126.77, 118.50 (q,  $J = 299$  Hz), 115.31, 62.17, 59.90, 58.10, 55.32, 54.63, 54.59, 52.87, 52.43, 49.32, 48.53, 47.12, 42.12, 40.58, 39.14, 37.64, 36.09, 31.59, 30.95, 30.03, 29.77, 29.13, 28.24, 27.73, 27.02, 24.83, 24.54, 23.61, 22.64, 21.90, 19.70, 18.97, 18.45.

Peptide **1a**.  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ ):  $\delta$  10.38 (s, 1H), 10.35 (s, 1H), 8.51 (d,  $J = 7.2$  Hz, 1H), 8.17 – 8.22 (m, 2H), 8.08 – 8.12 (m, 2H), 8.03 (s, 5H), 7.90 (s, 1H), 7.75 – 7.79 (m, 3H), 7.71 (s, 1H), 7.25 (m, 4H), 7.19 (m, 1H), 7.14 (d,  $J = 11.9$  Hz, 2H), 7.09 (m, 1H), 7.00 (d,  $J = 6.8$  Hz, 2H), 6.63 (d,  $J = 6.9$  Hz, 2H), 6.45 (s, 1H), 4.43 – 4.54 (m, 2H), 4.36 – 4.42 (m, 3H), 4.33 (m, 2H), 4.22 – 4.27 (m, 2H), 4.08 – 4.15 (m, 1H), 3.68 (dd,  $J = 16.4, 5.4$  Hz, 1H), 3.55 – 3.60 (m, 6H), 3.45 (m, 1H), 3.05 – 3.12 (m, 2H), 2.92 – 3.02 (m, 2H), 2.78 – 2.88 (m, 3H), 2.75 (m, 1H), 2.63 – 2.70 (m, 2H), 2.08 – 2.10 (m, 1H), 1.97 (s, 1H), 1.84 (m, 4H), 1.77 (m, 1H), 1.68 – 1.72 (m, 1H), 1.58 – 1.67 (m, 2H), 1.50 – 1.57 (m, 3H), 1.45 – 1.48 (m, 2H), 1.34 – 1.40 (m, 1H), 1.24 – 1.32 (m, 2H), 1.20 (d,  $J = 6.6$  Hz, 3H), 0.87 – 0.91 (m, 12H).

$^{13}\text{C}$  NMR (150 MHz, DMSO- $d_6$ ):  $\delta$  172.08, 172.01, 171.77, 171.63, 171.58, 171.50, 171.37, 171.23, 171.16, 171.04, 170.49, 165.81, 159.50, 159.43, 159.36, 158.66 (d,  $J = 31.8$  Hz), 158.17, 158.07, 156.26, 138.15, 130.57, 129.63, 128.56, 129.97, 126.76, 117.47 (d,  $J = 295.3$  Hz), 115.29, 62.18, 61.91, 61.29, 59.88, 55.70, 55.24, 54.86, 54.59, 51.73, 51.44, 51.12, 50.99, 49.28, 48.51, 47.11, 42.49, 42.45, 40.60, 39.24, 37.67, 37.08, 36.89, 36.68, 29.18, 27.06, 27.03, 24.83, 24.54, 23.65, 22.43, 21.89, 19.88, 19.76, 19.70, 19.02, 18.45.

Peptide **1b**.  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ ):  $\delta$  10.38 (s, 1H), 10.34 (s, 1H), 8.51 (d,  $J = 7.4$  Hz, 1H), 8.20 (d,  $J = 7.6$  Hz, 1H), 8.17 (t,  $J = 5.7$  Hz, 1H), 8.07 – 8.11 (m, 2H), 8.03 (m, 5H), 7.81 – 7.91 (m, 2H), 7.71 – 7.76 (m, 3H), 7.23 – 7.27 (m, 4H), 7.13 – 7.20 (m, 3H), 7.09 (s, 1H), 7.00 (d,  $J = 8.1$  Hz, 1H), 6.62 (d,  $J = 8.4$  Hz, 1H), 6.45 (s, 1H), 5.10 (br. s, 1H), 4.44 – 7.50 (m, 2H), 4.36 – 4.42 (m, 3H), 4.31 – 4.34 (m, 2H), 4.26 – 4.28 (m, 1H), 4.17 – 4.24 (m, 1H), 4.13 (t,  $J = 7.0$  Hz, 1H), 3.68 (dd,  $J = 16.6, 5.4$  Hz, 1H), 3.50 – 3.65 (m, 6H), 3.45 (s, 1H), 3.10 – 3.13 (m, 1H), 3.06 (dd,  $J = 13.8, 4.3$  Hz, 1H), 2.81 (m, 1H), 2.68 – 2.78 (m, 3H), 2.09 (q,  $J = 6.6$  Hz, 1H), 1.97 (m, 1H), 1.81 – 1.86 (m, 3H), 1.75 – 1.81 (m, 2H), 1.68 – 1.74 (m, 1H), 1.59 – 1.66 (m, 1H), 1.47 – 1.59 (m, 4H), 1.46 – 1.48 (m, 1H), 1.36 – 1.38 (m, 1H), 1.27 – 1.32 (m, 2H), 1.20 (d,  $J = 7.0$  Hz, 3H), 0.90 (d,  $J = 7.4$  Hz, 3H), 0.98 (d,  $J = 6.8$  Hz, 6H), 0.85 (d,  $J = 6.6$  Hz, 3H).

$^{13}\text{C}$  NMR (150 MHz, DMSO- $d_6$ ):  $\delta$  172.01, 171.91, 171.79, 171.64, 171.59, 171.51, 171.42, 171.27, 171.23, 171.07, 170.50, 165.81, 159.66, 159.60, 158.72 (d,  $J = 30.7$  Hz), 158.18, 158.10, 156.26, 138.13, 130.59, 129.59, 128.56, 127.97, 126.77, 117.42 (d,  $J = 299.6$  Hz), 115.29, 61.92, 59.90, 58.71, 55.26, 54.78, 54.46, 52.92, 51.00, 49.28, 48.51, 47.12, 42.46, 40.60, 39.24, 37.69, 36.88, 36.68, 32.46, 31.19, 30.22, 29.16, 27.03, 24.83, 24.54, 23.65, 22.43, 21.90, 19.87, 19.76, 19.70, 19.01, 18.43.

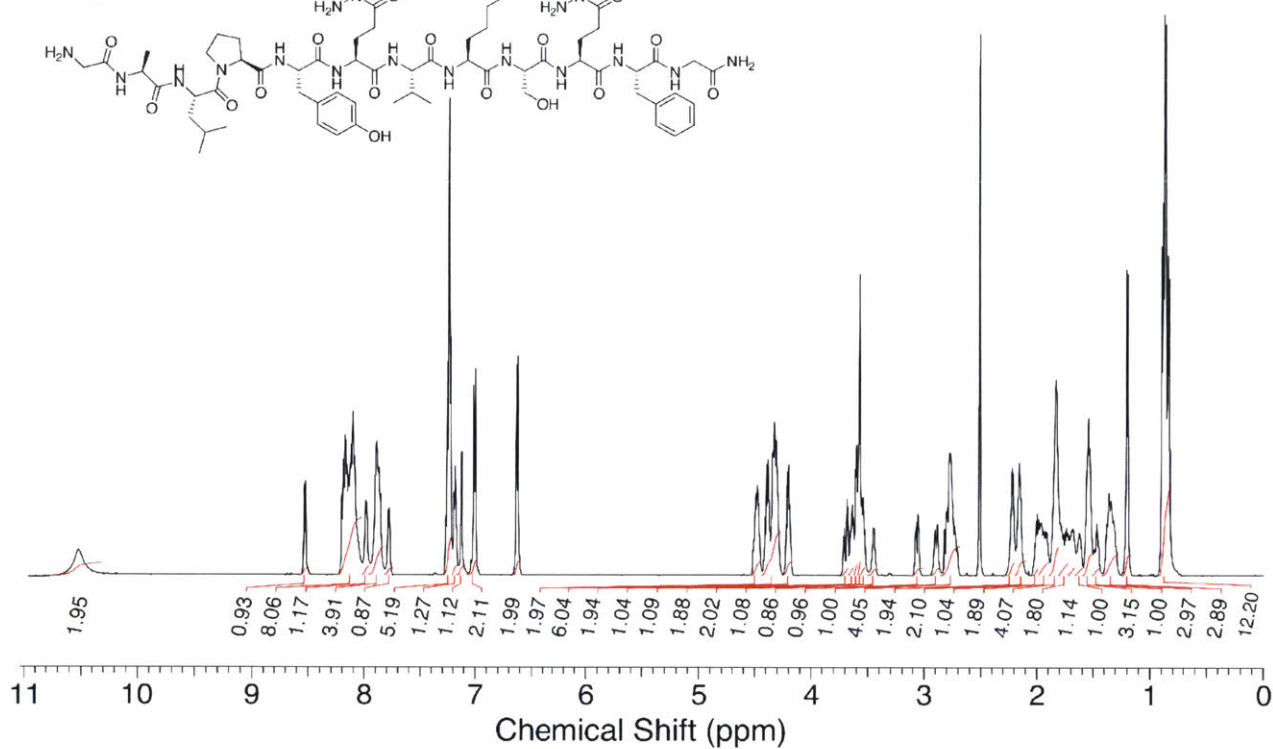
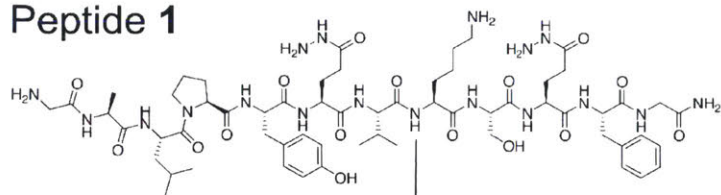
Peptide **1c**.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  9.69 (s, 1H), 9.66 (s, 1H), 8.50 (d,  $J = 7.3$  Hz, 1H), 8.22 (d,  $J = 7.5$  Hz, 1H), 8.15 – 8.18 (m, 1H), 8.06 (d,  $J = 8.3$  Hz, 1H), 8.04 (d,  $J = 8.0$  Hz, 1H), 7.99 (m, 3H), 7.93 (s, 1H), 7.81 – 7.83 (m, 1H), 7.72 – 7.79 (m, 2H), 7.68 (m, 2H), 7.23 – 7.28 (m, 4H), 7.20 (t,  $J = 6.2$  Hz, 1H), 7.13 (m, 1H), 7.09 (s, 1H), 7.05 (s, 3H), 7.00 (d,  $J = 8.1$  Hz, 2H), 6.61 (d,  $J = 8.3$  Hz, 2H), 6.39 (s, 1H), 4.47 – 7.52 (m, 1H), 4.43 – 7.46 (m, 1H), 4.36 – 7.42 (m, 2H), 4.28 – 7.34 (m, 3H), 4.17 – 7.27 (m, 2H), 4.12 – 7.14 (m, 1H), 3.68 (dd,  $J = 16.7, 5.8$  Hz, 1H), 3.54 – 3.63 (m, 6H), 3.45 (s, 1H), 2.98 – 3.12 (m, 4H), 2.96 (m, 1H), 2.88 (d,  $J = 9.9$  Hz, 1H), 2.81 (dd,  $J = 13.2, 10.8$  Hz, 1H), 2.68 – 2.76 (m, 3H), 2.41 (m, 3H), 2.08 – 2.15 (m, 1H), 1.97 (s, 1H), 1.84 (m, 3H), 1.74 – 1.81 (m, 2H), 1.66 – 1.71 (m, 2H), 1.58 – 1.65 (m, 2H), 1.48 – 1.56 (m, 4H), 1.46 (m, 1H), 1.35 – 1.39 (m, 1H), 1.28 – 1.33 (m, 2H), 1.20 (d,  $J = 7.0$  Hz, 3H), 0.85 – 0.91 (m, 12H).

$^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  172.62, 172.53, 172.25, 172.13, 172.03, 171.94, 171.82, 171.70, 171.55, 171.49, 171.37, 171.24, 170.76, 165.60, 159.04, 158.90, 158.87, 156.05, 137.98, 130.58, 129.58, 128.61, 127.91, 126.85, 117.52, 115.26, 61.32, 59.98, 59.26, 55.77, 54.79, 54.57, 53.03, 51.49, 51.19, 49.33, 48.46, 47.14, 42.29, 40.31, 38.96, 37.51, 36.68, 36.58, 32.91, 32.13, 31.41, 30.06, 29.14, 29.58, 26.76, 26.71, 24.82, 24.54, 23.58, 22.27, 21.81, 19.69, 18.74, 18.32.

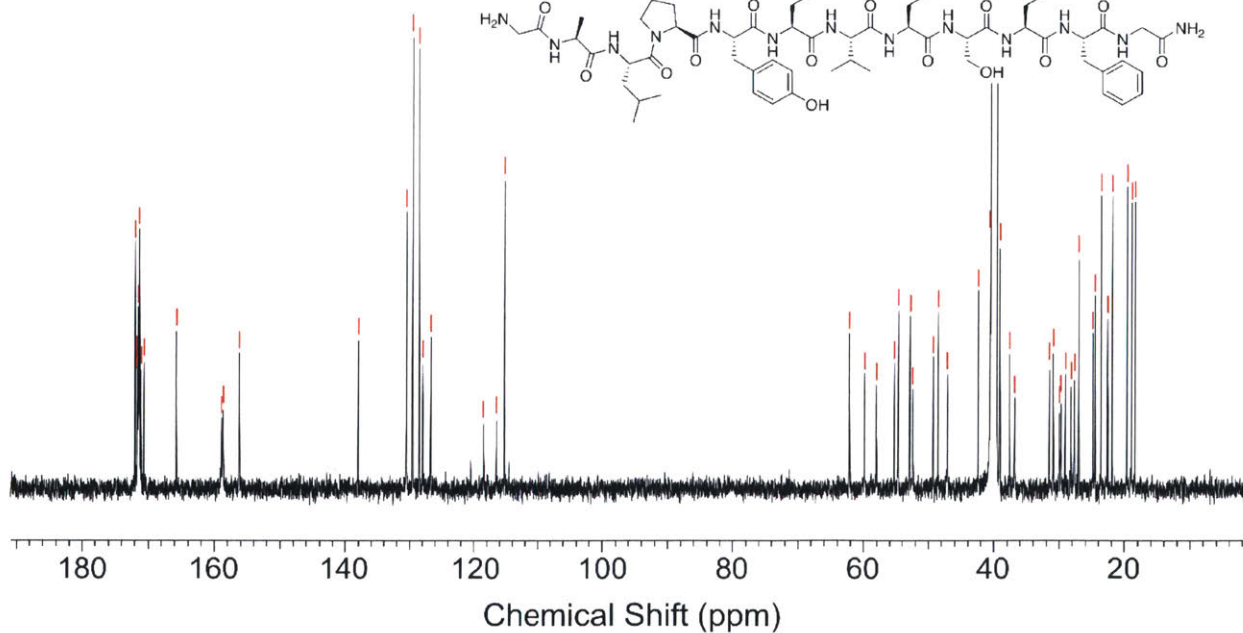
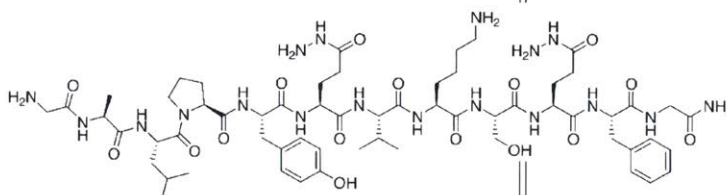
Peptide **1d**.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  9.53 (s, 1H), 9.49 (s, 1H), 8.51 (d,  $J = 7.2$  Hz, 1H), 8.20 (d,  $J = 7.2$  Hz, 1H), 8.15 (s, 1H), 8.01 – 8.10 (m, 5H), 7.96 (m, 1H), 7.92 (s, 1H), 7.84 (s, 1H), 7.73 – 7.81 (m, 4H), 7.23 – 7.27 (m, 4H), 7.17 – 7.23 (m, 3H), 7.14 (s, 1H), 7.09 (s, 1H), 6.99 (d,  $J = 7.9$  Hz, 2H), 6.63 (d,  $J = 7.9$  Hz, 2H), 6.46 (s, 1H), 6.37 (s, 1H), 4.49 (m, 1H), 4.43 – 4.47 (m, 1H), 4.36 – 4.41 (m, 2H), 4.25 – 4.35 (m, 4H), 4.18 – 4.23 (m, 1H), 4.15 (t,  $J = 6.6$  Hz, 1H), 3.69 (dd,  $J = 16.6, 5.4$  Hz, 1H), 3.52 – 3.64 (m, 6H), 3.45 (s, 1H), 2.96 – 3.14 (m, 4H), 2.94 (m, 1H), 2.88 (d,  $J = 10.7$  Hz, 1H), 2.79 – 2.84 (m, 1H), 2.67 – 2.83 (m, 3H), 2.12 (m, 4H), 2.03 – 2.08 (m, 1H), 1.97 (s, 1H), 1.83 (m, 3H), 1.70 – 1.77 (m, 2H), 1.58 – 1.68 (m, 2H), 1.48 – 1.57 (m, 8H), 1.41 – 1.46 (m, 1H), 1.37 (d,  $J = 9.8$  Hz, 1H), 1.28 – 1.35 (m, 2H), 1.20 (d,  $J = 6.8$  Hz, 3H), 0.85 – 0.92 (m, 12H).

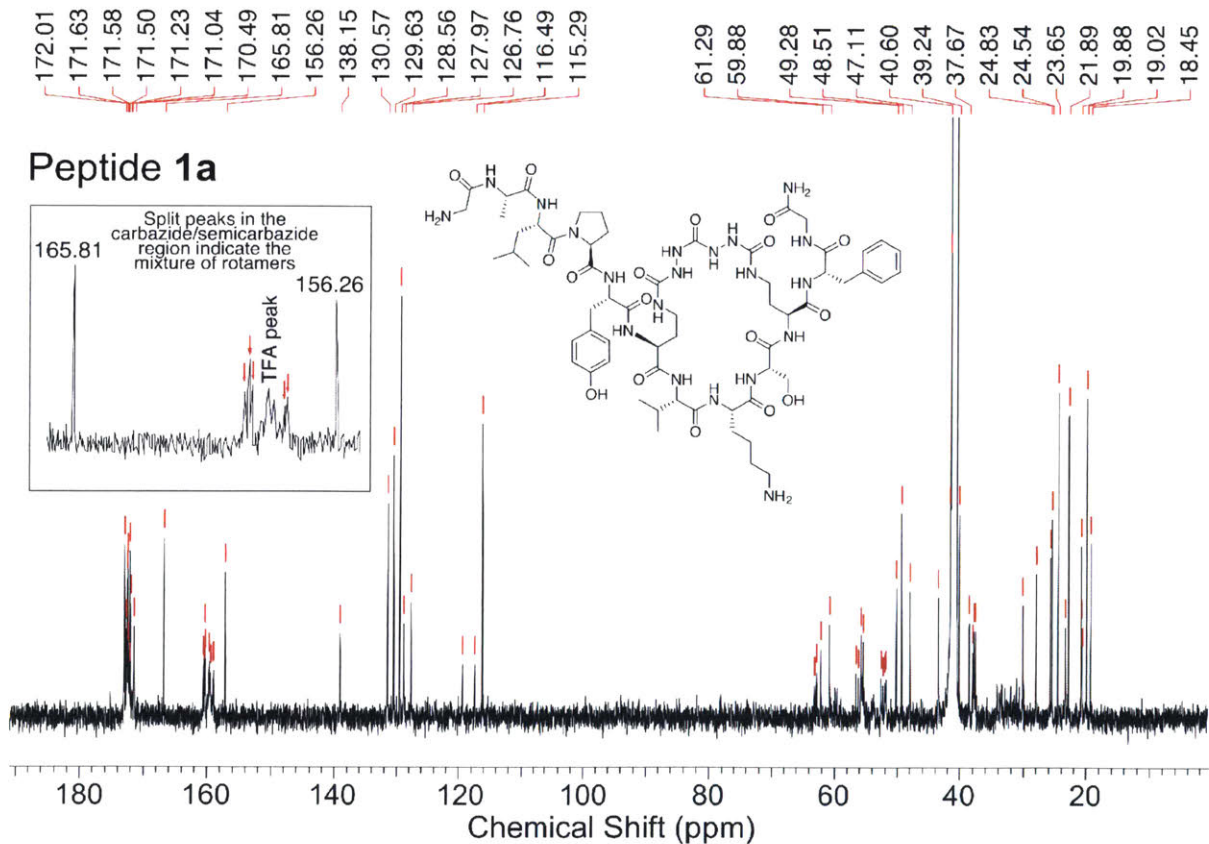
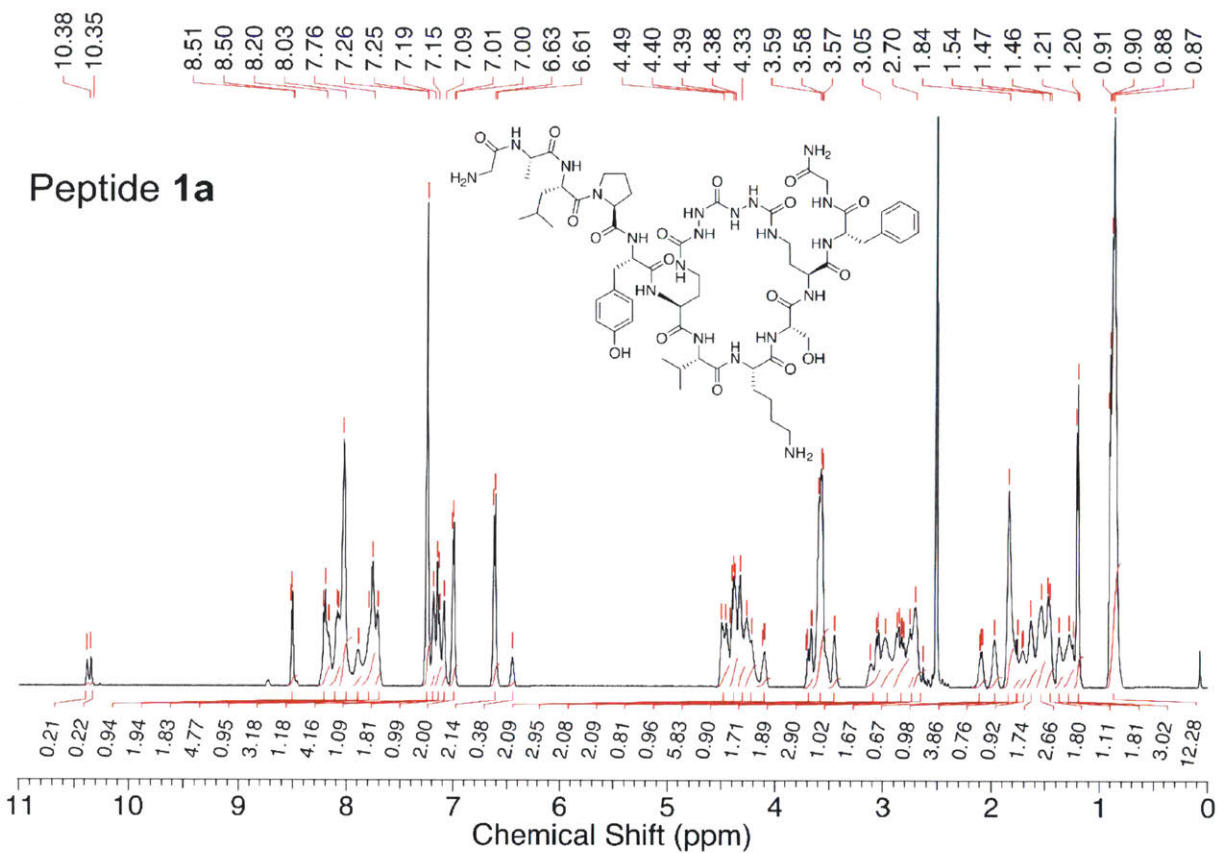
$^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO-}d_6$ ):  $\delta$  172.85, 172.80, 172.01, 171.87, 171.75, 171.69, 171.63, 171.61, 171.51, 171.33, 171.27, 171.07, 170.67, 165.81, 158.80 (d,  $J = 23.1$  Hz), 158.82, 156.26, 138.12, 130.58, 129.61, 128.56, 127.94, 126.77, 117.44 (q,  $J = 303$  Hz), 115.29, 61.91, 59.90, 58.73, 55.40, 54.80, 54.55, 52.91, 51.48, 51.06, 49.28, 48.51, 47.11, 42.47, 40.60, 39.24, 37.67, 36.85, 36.59, 33.50, 33.15, 33.08, 32.67, 31.50, 30.32, 29.18, 26.93, 24.83, 24.54, 23.63, 22.34, 21.89, 19.76, 19.00, 18.42.

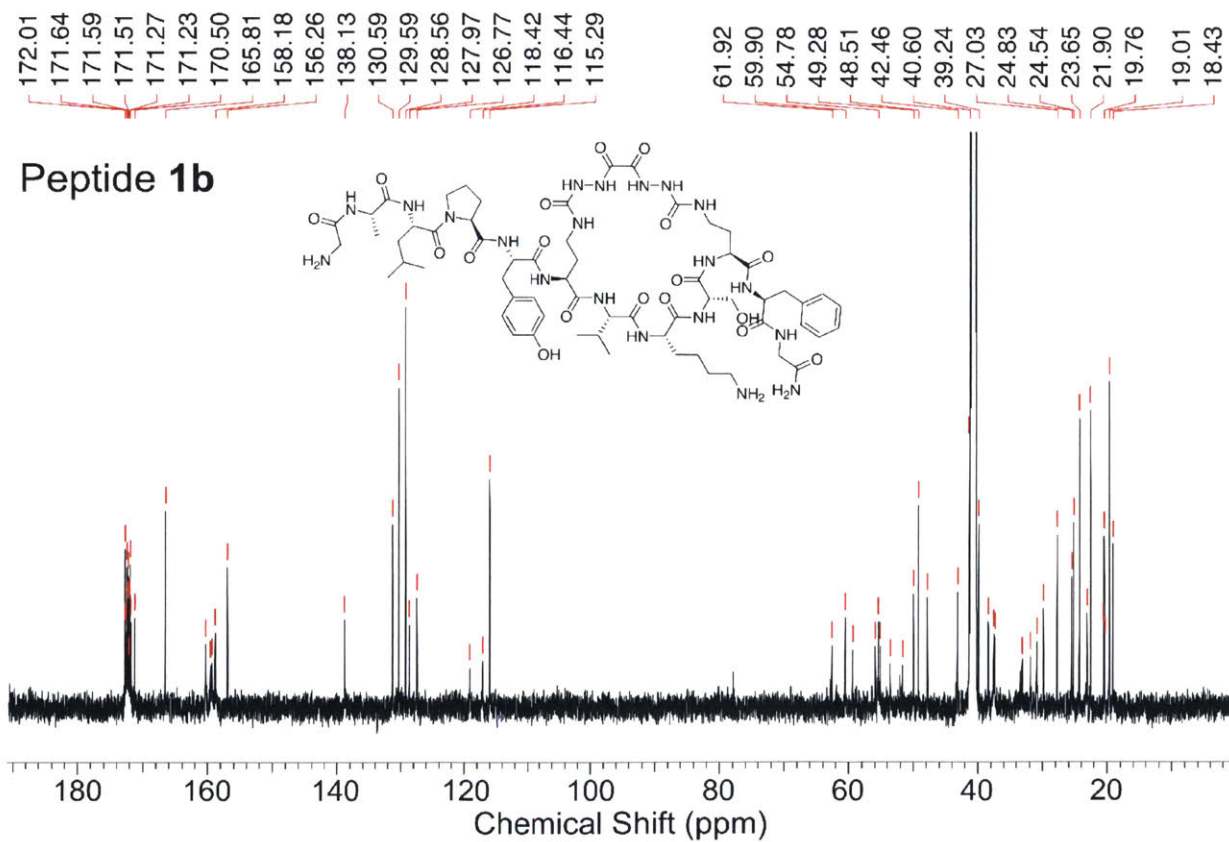
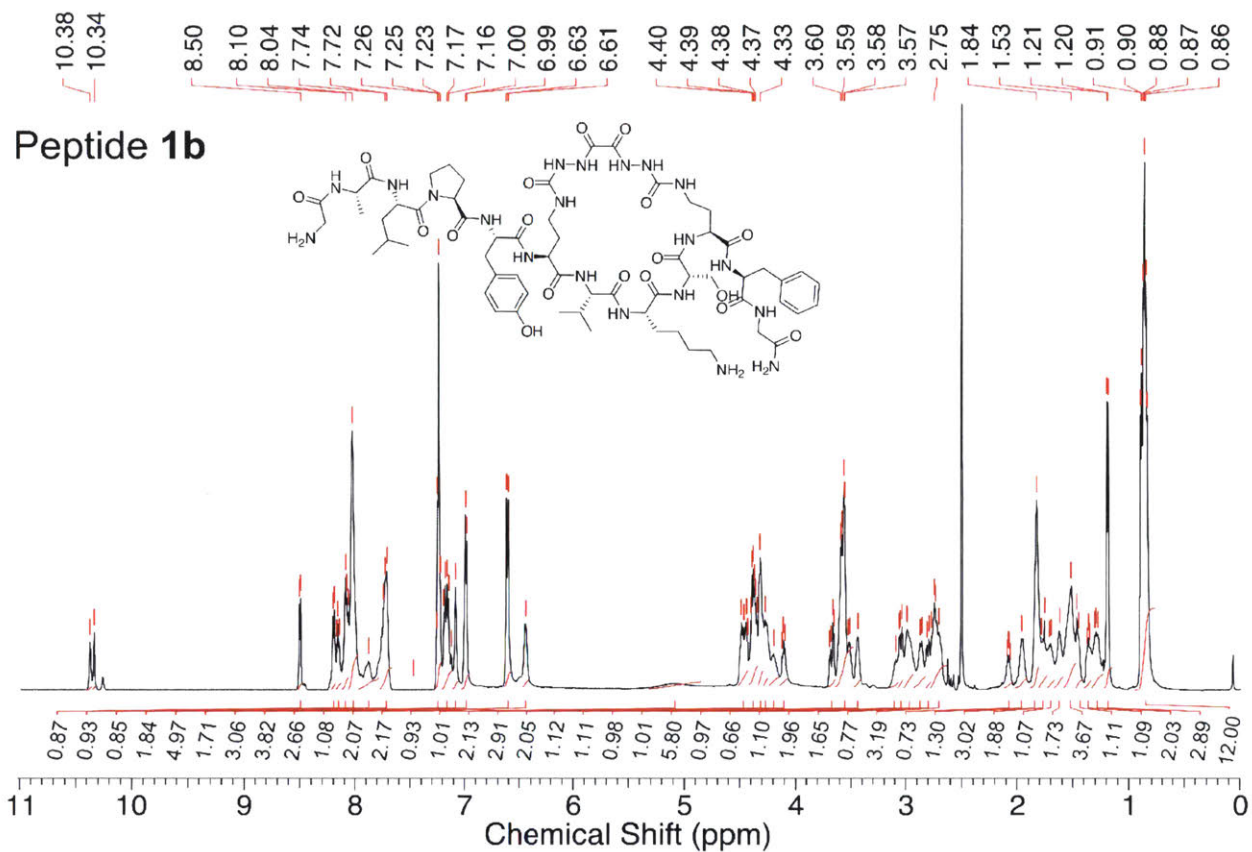
# Peptide 1

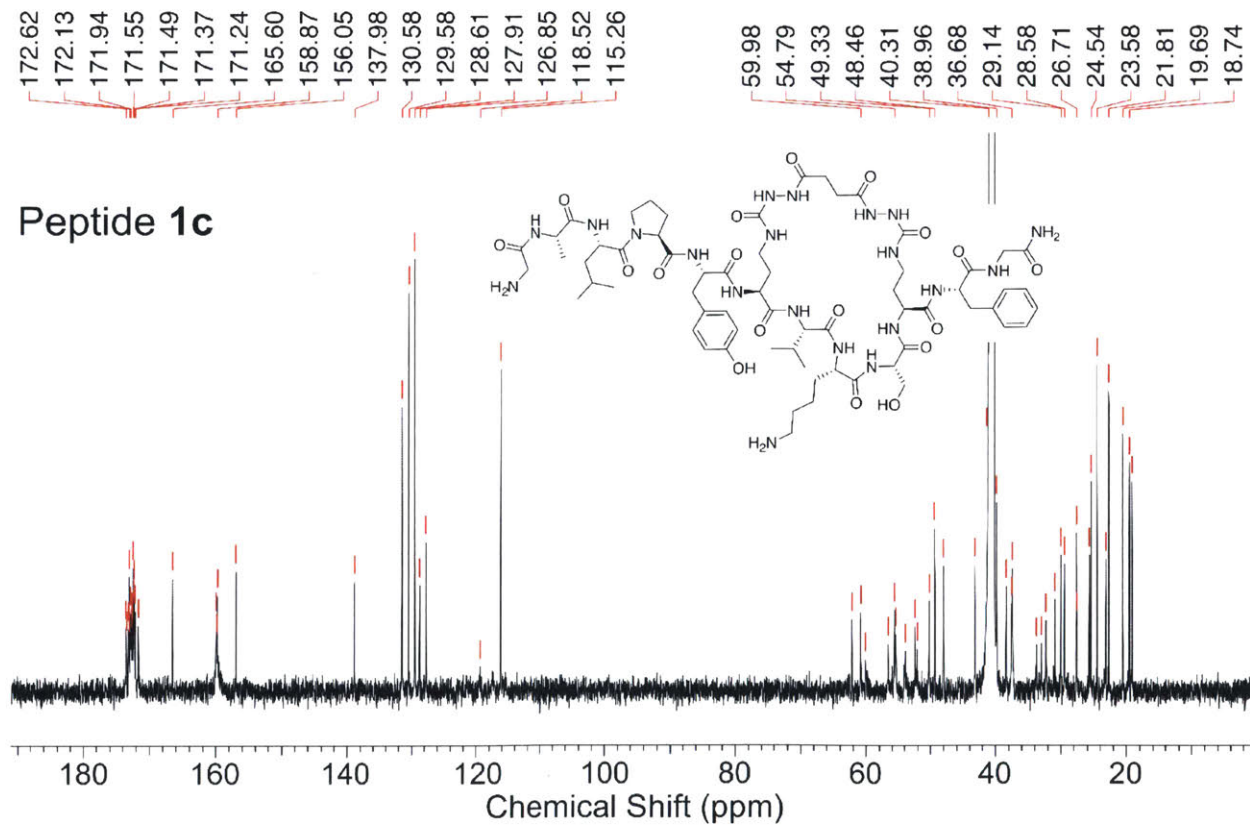
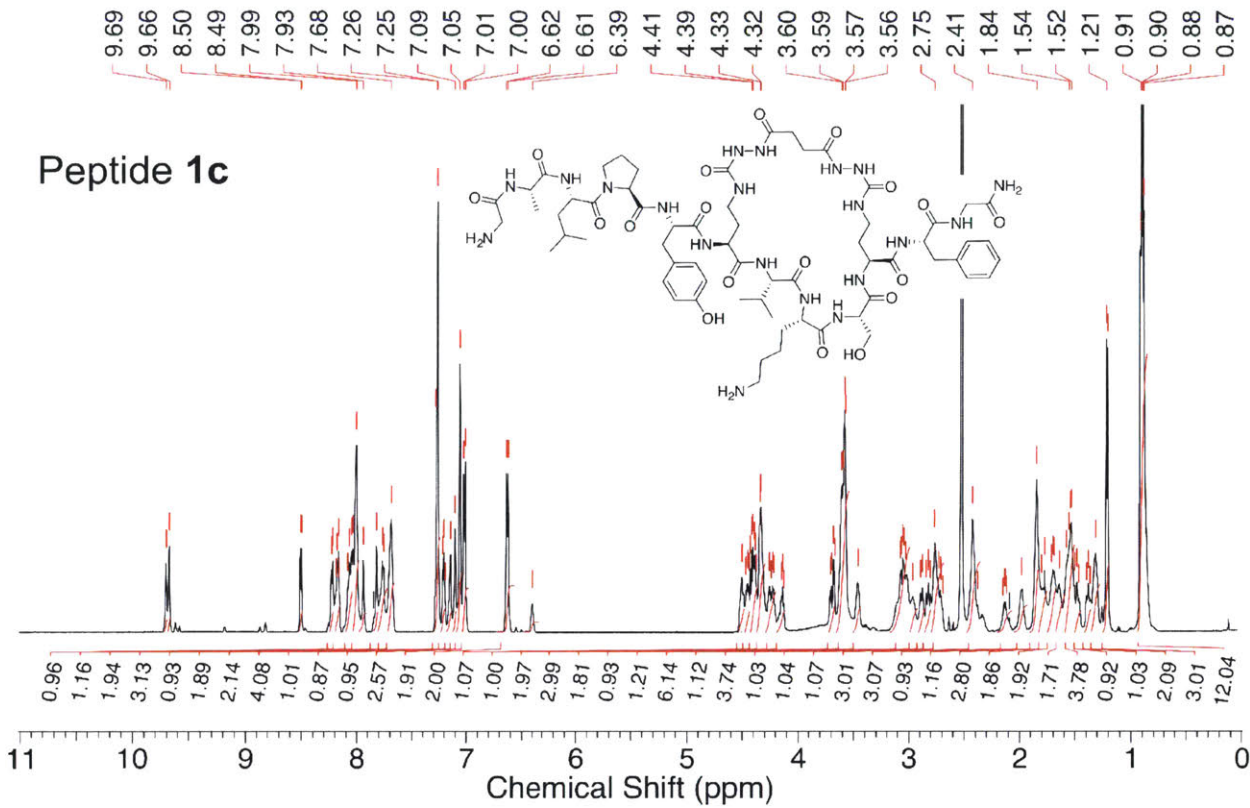


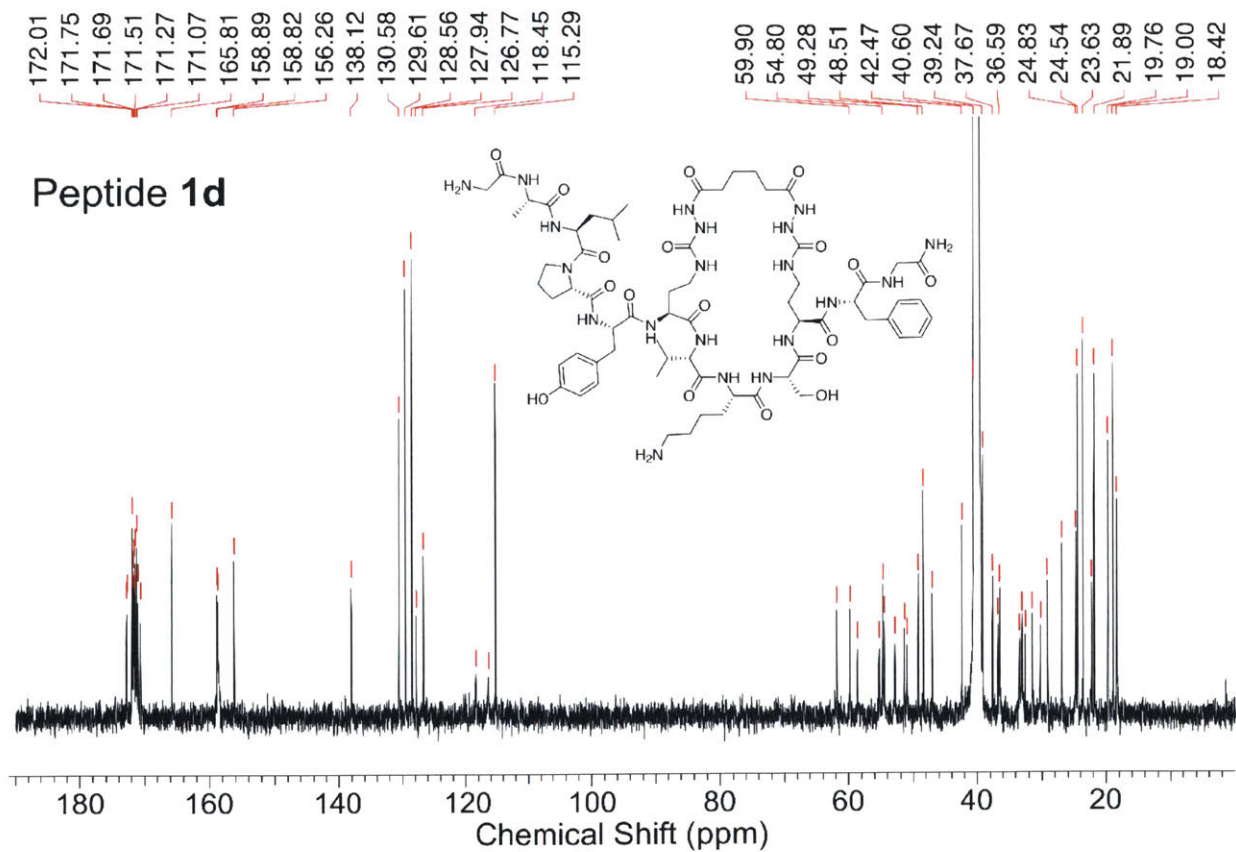
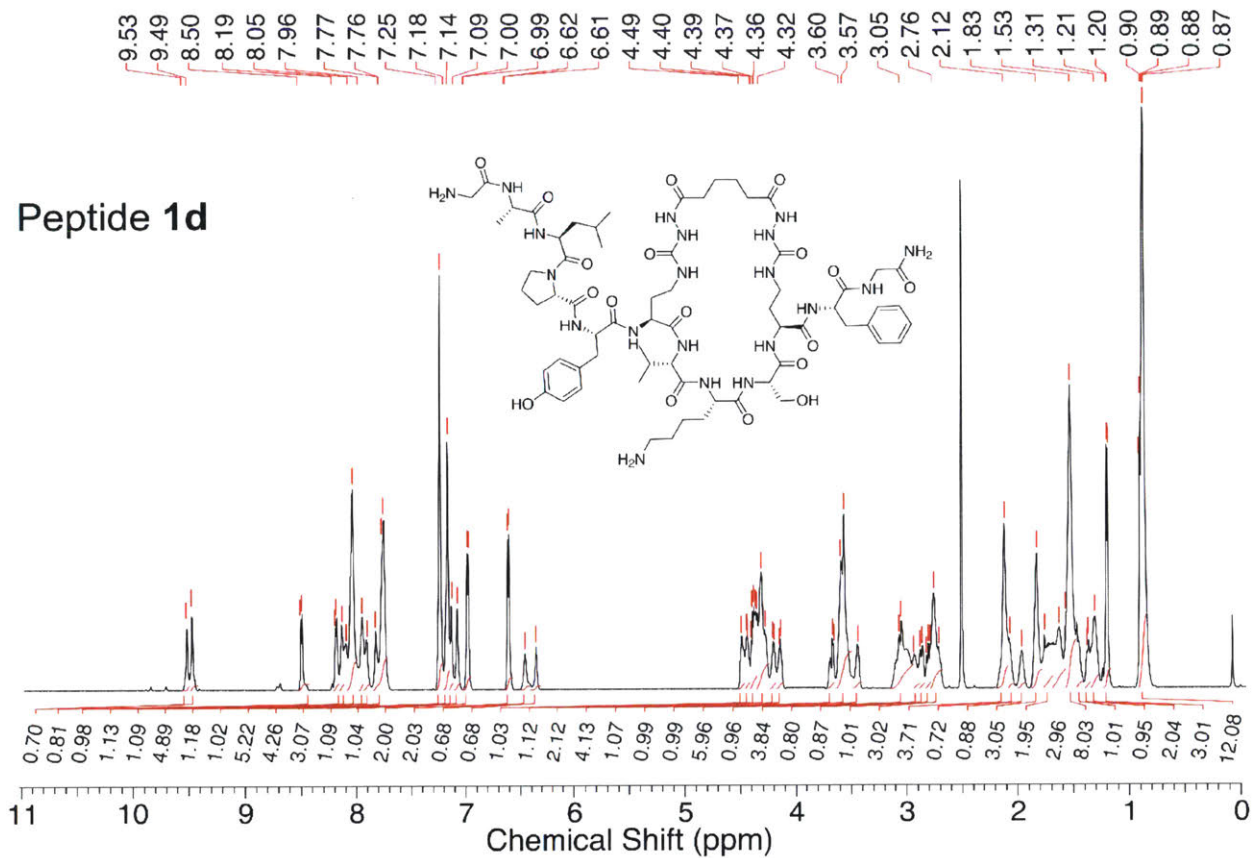
# Peptide 1













## 4.4. Acknowledgements

We are thankful to Dr. Amy Rabideau, Dr. Daniel Cohen, Justin Wolfe, and Ethan Evans (all — MIT, Chemistry Department) for technical assistance and fruitful discussions throughout the course of the work. This research was supported by Amgen summer graduate fellowship for A. V., MIT UROP program for Z. C. and funded by MIT start-up funds, MIT Reed Fund, and DARPA (Award #023504-001) for B.L.P. We acknowledge the Biological Instrument Facility of MIT for providing access to the Octet Bio-layer Interferometry System (NIH S10 OD016326) and to CD spectrometry (NSF-0070319).

## 4.5. References

- (1) Schafmeister, C.; Po, J.; Verdine, G. *J. Am. Chem. Soc.* **2000**, *122*, 5891–5892.
- (2) Verdine, G.; Hilinski, G. *J. Methods Enzymol.* **2012**, *503*, 3.
- (3) Bernal, F.; Tyler, A.; Korsmeyer, S.; Walensky, L.; Verdine, G. *J. Am. Chem. Soc.* **2007**, *129*, 2456–2457.
- (4) Bautista, A.; Appelbaum, J.; Craig, C.; Michel, J.; Schepartz, A. *J. Am. Chem. Soc.* **2010**, *132*, 2904–2906.
- (5) Chang, Y.; Graves, B.; Guerlavais, V.; Tovar, C.; Packman, K.; To, K.-H.; Olson, K.; Kesavan, K.; Gangurde, P.; Mukherjee, A.; Baker, T.; Darlak, K.; Elkin, C.; Filipovic, Z.; Qureshi, F.; Cai, H.; Berry, P.; Feyfant, E.; Shi, X.; Horstick, J.; Annis, D.; Manning, A.; Fotouhi, N.; Nash, H.; Vassilev, L.; Sawyer, T. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E3445–E3454.
- (6) Walensky, L.; Kung, A.; Escher, I.; Malia, T.; Barbuto, S.; Wright, R.; Wagner, G.; Verdine, G.; Korsmeyer, S. *Science*. **2004**, *305*, 1466–1470.
- (7) Walensky, L.; Pitter, K.; Morash, J.; Oh, K.; Barbuto, S.; Fisher, J.; Smith, E.; Verdine, G.; Korsmeyer, S. *Mol. Cell* **2006**, *24*, 199–210.
- (8) Gavathiotis, E.; Reyna, D.; Davis, M.; Bird, G.; Walensky, L. *Mol. Cell* **2010**, *40*, 481–492.
- (9) Bird, G.; Gavathiotis, E.; Labelle, J.; Katz, S.; Walensky, L. *ACS Chem. Biol.* **2014**, *9*, 831–837.
- (10) Takada, K.; Zhu, D.; Bird, G.; Sukhdeo, K.; Zhao, J.-J.; Mani, M.; Lemieux, M.; Carrasco, D.; Ryan, J.; Horst, D.; Fulciniti, M.; Munshi, N.; Xu, W.; Kung, A.; Shivdasani, R.; Walensky, L.; Carrasco, D. *Sci. Transl. Med.* **2012**, *4*, 148ra117.
- (11) Grossmann, T.; Yeh, J.; Bowman, B.; Chu, Q.; Moellering, R.; Verdine, G. *Proc. Natl. Acad. Sci.* **2012**, *109*, 17942–17947.
- (12) Cromm, P.; Spiegel, J.; Grossmann, T. *ACS Chem. Biol.* **2015**, *10*, 1362–1375.
- (13) Walensky, L.; Bird, G. *J. Med. Chem.* **2014**, *57*, 6275–6288.
- (14) Spokoiny, A.; Zou, Y.; Ling, J.; Yu, H.; Lin, Y.; Pentelute, B. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.
- (15) Brown, S.; Smith, A. *J. Am. Chem. Soc.* **2015**, *137*, 4034–4037.
- (16) Assem, N.; Ferreira, D.; Wolan, D.; Dawson, P. *Angew. Chem.; Int. Ed.* **2015**, *54*, 8665–8668.
- (17) Muppidi, A.; Doi, K.; Edwardraja, S.; Drake, E.; Gulick, A.; Wang, H.; Lin, Q. *J. Am. Chem. Soc.* **2012**, *134*, 14734–14737.
- (18) Wang, Y.; Chou, D. *Angew. Chem.; Int. Ed.* **2015**, *54*, 10931–10934.
- (19) Jackson, D.; King, D.; Chmielewski, J.; Singh, S.; Schultz, P. *J. Am. Chem. Soc.* **1991**, *113*, 9391–9392.
- (20) De Araujo, A.; Hoang, H.; Kok, W.; Diness, F.; Gupta, P.; Hill, T.; Driver, R.; Price, D.; Liras, S.; Fairlie, D. *Angew. Chem.; Int. Ed.* **2014**, *53*, 6965–6969.
- (21) Bracken, C.; Gulyas, J.; Taylor, J.; Baum, J. *J. Am. Chem. Soc.* **1994**, *116*, 6431–6432.
- (22) Haney, C.; Loch, M.; Horne, W. *Chem. Commun.* **2011**, *47*, 10915–10917.

- (23) Lau, Y.; de Andrade, P.; Quah, S.-T.; Rossmann, M.; Laraia, L.; Sköld, N.; Sum, T.; Rowling, P.; Joseph, T.; Verma, C.; Hyvönen, M.; Itzhaki, L.; Venkitaraman, A.; Brown, C.; Lane, D.; Spring, D. *Chem. Sci.* **2014**, *5*, 1804–1809.
- (24) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7645–7649.
- (25) Simon, M.; Heider, P.; Adamo, A.; Vinogradov, A.; Mong, S.; Li, X.; Berger, T.; Policarpo, R.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A.; Jensen, K.; Pentelute, B. *Chembiochem* **2014**, *15*, 713–720.
- (26) Mong, S.; Vinogradov, A.; Simon, M.; Pentelute, B. *Chembiochem* **2014**, *15*, 721–733.
- (27) Bird, G.; Bernal, F.; Pitter, K.; Walensky, L. *Methods Enzymol.* **2008**, *446*, 369–386.
- (28) Sticht, J.; Humbert, M.; Findlow, S.; Bodem, J.; Müller, B.; Dietrich, U.; Werner, J.; Kräusslich, H.-G. *Nat. Struct. Mol. Biol.* **2005**, *12*, 671–677.
- (29) Zhang, H.; Zhao, Q.; Bhattacharya, S.; Waheed, A.; Tong, X.; Hong, A.; Heck, S.; Curreli, F.; Goger, M.; Cowburn, D.; Freed, E.; Debnath, A. *J. Mol. Biol.* **2008**, *378*, 565–580.
- (30) Zhang, H.; Curreli, F.; Zhang, X.; Bhattacharya, S.; Waheed, A.; Cooper, A.; Cowburn, D.; Freed, E.; Debnath, A. *Retrovirology* **2011**, *8*, 28–46.
- (31) Stavropoulos, G.; Gatos, D.; Magafa, V.; Barlos, K. *Lett. Pept. Sci.* **1995**, *2*, 315–318.

## Chapter 5. Strategies for High-Throughput Sequencing of Synthetic Peptide Libraries

The work presented in this chapter has been submitted for publication:

Vinogradov, A.; Gates, Z.; Zhang, C.; Quartararo, A.; Halloran, K.; Pentelute, B. “Strategies for high-throughput sequencing of synthetic peptide libraries”

## 5.1. Introduction

In this Chapter we significantly diverge from the established chemistry-intensive narrative presented in Chapters 1 through 4. *The focus of the current work is the development of reliable high-throughput methods for the analysis of complex (hundreds to thousands) peptide mixtures using a combination of liquid chromatography and tandem mass spectrometry.* By analysis in this context we understand the combination of automated procedures for the separation, identification and *de novo* sequence determination of as many peptides comprising corresponding library mixtures as possible. Such a methodology can be utilized in various contexts and for different purposes. For instance, these methods can be used to identify the structures of peptides isolated from screens of combinatorial libraries, or, as we show in Chapter 6, to perform direct reactivity screening on peptide mixtures to discover new chemistries.

---

Decoding the structure of identified “hits” is a necessary step in most embodiments of the combinatorial discovery process. With regard to peptide libraries, several approaches to sequence determination of screening hits are established (Fig. 5.1). DNA/RNA encoding of peptide libraries is one the most commonly used strategies, because the DNA sequencing technology is reliable, fast and cheap.<sup>1</sup> This approach is commonly utilized in phage,<sup>2,3</sup> yeast<sup>4,5</sup> and mRNA<sup>6,7</sup> display strategies. Genetic encoding of peptide libraries is limited, however, in its ability to successfully incorporate a variety of “privileged” non-proteogenic amino acids and their analogues ( $\beta$ -,  $\gamma$ -, D-amino acids and peptoids),<sup>8-11</sup> although this issue can be circumvented to a certain degree in mRNA display.<sup>12,13</sup> Another limitation of this approach is a number of experimental difficulties associated with encoding peptide library members in some combinatorial approaches. For instance, although one-bead one-compound (OBOC) peptide libraries were shown to be DNA-encodable,<sup>14</sup> such a strategy did not gain popularity due to associated synthetic challenges. In the context of OBOC libraries, much more common decoding strategies are either derived from Edman degradation<sup>15-18</sup> or rely on MALDI-coupled<sup>19-21</sup> tandem mass spectrometry-based (MSMS) *de novo* peptide sequencing. Edman degradation, while reliable, is slow,<sup>22</sup> not easily amenable to multiplexing, and, like genetic encoding, is generally limited to  $\alpha$ -amino acids.<sup>23</sup> Mass spectrometry-based strategies are able to decode amino acids of great structural diversity,<sup>24-26</sup> but the techniques remain fairly slow, and are less reliable than either of the two strategies described above.<sup>20</sup> Taken together, the considerations listed above suggest that there is still a need for a robust, high-throughput decoding technique applicable to OBOC libraries. The development of such a strategy is the focus of the present work.

The idea of our approach is summarized in Figure 5.2. We reasoned that high-throughput peptide sequencing could be achieved via LC/MSMS analysis of samples containing complex library bead mixtures.



**Nucleic acid encoding**



Fast  
Reliable  
Cheap

Proteogenic amino acids

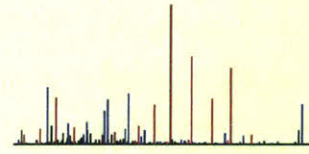
**Edman degradation**



Slow  
Reliable

$\alpha$ -amino acids

**Mass spectrometry**



Probabilistic  
Cheap

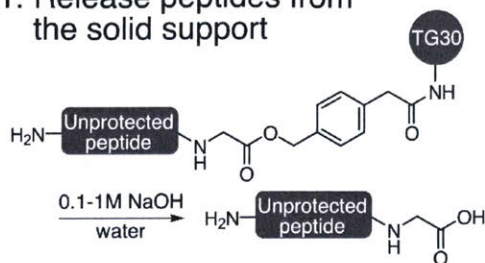
Any amino acids

**Figure 5.1. Comparison of various decoding strategies**

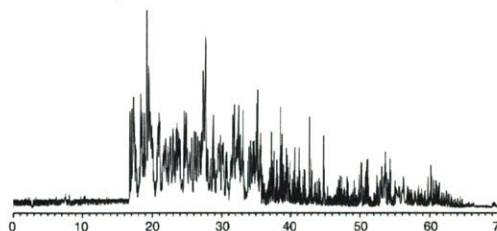
Indeed, HPLC and nano-LC (nLC) are capable of separating mixtures containing thousands of peptides, as recent advances from the field of proteomics indicate.<sup>27-29</sup> At first glance, the task of decoding a sample of several hundred to a few thousand OBOC beads is not much different from the sequencing of a subset of human proteome peptides. The key difference, however, lies in the fact that databases, commonly used in the proteomics peptide identification process, for combinatorial libraries are not helpful in the task of producing high accuracy *de novo* spectral assignments. On the other hand, synthetic peptides are a subject to *a priori* library design considerations, which can be leveraged to increase peptide “sequencability” and the resulting confidence in obtained results. Provided a sufficiently constraining library design, verification of *de novo* peptides becomes a lot like database matching in proteomics. In this case matching sequencing outcomes back to the original library design can be intuitively expected to increase the overall decoding accuracy.

Here, we investigate the utility of library design rules for improving the accuracy of automated *de novo* sequence assignments of MSMS spectra generated from LC/MS analysis of synthetic peptide mixtures. We demonstrate the benefits of constraining library peptide sequences for high-throughput decoding of libraries comprised of proteogenic and non-proteogenic amino acid monomers.

1. Release peptides from the solid support



2. Analyze by nano-LC/MSMS



3. De novo sequence with PEAKS 8

Generate a list of candidate sequences for each scan



4. Data filtration based on library design criteria

Remove noise and incorrect assignments



5. Data refinement

Remove non-unique sequences  
Identify and remove side-products



**Figure 5.2. The overall experimental workflow for the proposed sequencing strategy**

## 5.2. Results and Discussion

### 5.2.1. Library Design

We began by identifying the features of a library sequence that might be used to distinguish a potentially correct assignment from an obviously incorrect one (that is, one inconsistent with library design rules). In most cases, all intended library members possess some common features, which allow the discrimination of actual library peptides from noise, nonsense *de novo* assignments and synthetic impurities such as deletions and truncations. Two such features are peptide length and the presence of a fixed subsequence (constant region). For instance, in our case the use of glycyl-*OCH*<sub>2</sub>-phenylacetamidomethyl (PAM) ester as a TFA-stable/base-labile cleavable linker<sup>30,31</sup> resulted in a fixed C-terminal Gly residue for all library peptides.

In addition to these features, we introduced an alternating monomer subset (AMS) design in an attempt to a) increase *de novo* sequencing assignment confidence in the peptide variable region, and b) resolve assignment ambiguities for fragmentation spectra with incomplete ion ladders, especially for those cases where both complimentary b and y ions are missing. Briefly, in this approach a library is constructed using standard SPPS procedures using two non-intersecting amino acid subsets, alternating them positionally, i.e. subset<sub>1</sub>-subset<sub>2</sub>-subset<sub>1</sub>-subset<sub>2</sub>-etc. Monomer subsets are constructed by dividing the total monomer pool into two subsets of equal size:

$$\mathcal{M}_1 \in \mathcal{M}, \mathcal{M}_2 = \mathcal{M} \cap \mathcal{M}_1 \quad (1)$$

such that

$$\forall p_1, p_2 \in \mathcal{M}_1, \forall q_1, q_2 \in \mathcal{M}_2: ||m(p_1) - m(p_2)| - |m(q_1) - m(q_2)|| > \varepsilon, \quad (2)$$

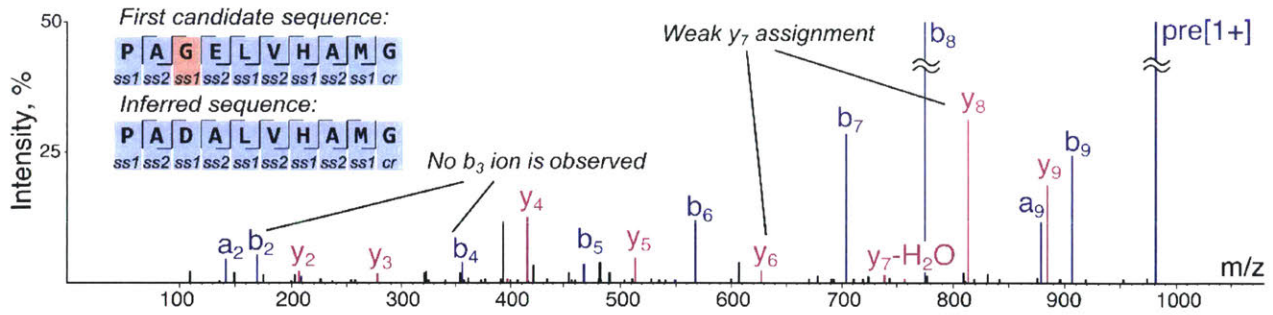
where  $m(p)$  is molecular weight of a monomer  $p$  in Da, and  $\varepsilon$  is a constant (in Da) determined by the accuracy of the mass spectrometer used for sequencing. We used  $\varepsilon=0$  to construct our monomer subsets, but the values of up to 0.01 Da yield identical results for proteogenic amino acids. We anticipated that the AMS approach would improve the overall decoding confidence by decreasing the probability of randomly assigning a subsequence allowed by the library design for spectra where definitive assignments may not be possible. More importantly, we hoped to resolve any potential ambiguous isomeric dipeptide arrangements, due to the fact that every subset<sub>1</sub>-subset<sub>2</sub> dipeptide has unique (to an extent of  $\varepsilon$ ) molecular weight (Table 5.1). Additionally, this approach is straightforward in its experimental implementation as it does not require any extra chemical manipulations during or after library assembly, which can become an important consideration for libraries of great size. One obvious disadvantage of working with AMS peptide libraries is, of course, their artificially limited diversity — both in terms of the theoretical library size and in terms of limited structural-positional variance of library members. We expect that this drawback can be mitigated



**Table 5.1. Monoisotopic masses in Da of all dipeptides encodable with the alternating monomer set library design**

ss <sub>1</sub> \ ss <sub>2</sub>	A	E	G	Q	S	T	V	Y
D	186.06	244.07	172.05	243.09	202.06	216.07	214.10	278.09
F	218.11	276.11	204.09	275.13	234.10	248.12	246.14	310.13
H	208.10	266.10	194.08	265.12	224.09	238.11	236.13	300.12
K	199.13	257.14	185.12	256.15	215.13	229.14	227.16	291.16
L	184.12	242.13	170.11	241.14	200.12	214.13	212.15	276.15
M	202.08	260.08	188.06	259.10	218.07	232.09	230.11	294.10
P	168.09	226.10	154.07	225.11	184.08	198.10	196.12	260.12
W	257.12	315.12	243.10	314.14	273.11	287.13	285.15	349.14

ss<sub>1</sub>: amino acid subset 1, ss<sub>2</sub>: amino acid subset 2.



**Figure 5.3. Recovery of sequences with incomplete fragmentation ladders using the AMS principle**  
 Displayed are merged, pre-processed collision-induced dissociation and higher-energy collisional dissociation spectra for 491.243 Da/e precursor ion. The first candidate sequence does not match the library design pattern due to the fact that the pair of  $b_3/y_7$ -ions is not observed in the spectrum, and consecutively, an unreliable assignment is made. Molecular weight of 186.064 Da corresponds to four different dipeptides (Gly-Glu, Glu-Gly, Ala-Asp and Asp-Ala), but only one of them (Asp-Ala) matches the parent design, which makes an unambiguous assignment possible. *ss1*: monomer subset 1, *ss2*: monomer subset 2, *cr*: constant region.

by increasing the size and the structural diversity of the parent monomer set, such that structural differences in alternating subsets become less prominent.

### 5.2.2. Decoding of a Model Library

With these considerations in mind, we turned to constructing a model 10-mer library to study our hypotheses. For a total amino acid set of 16 amino acids a number of feasible monomer subsets exist. We chose the combination where amino acids Asp, Phe, His, Lys, Met, Pro, Trp are combined in subset 1 ( $ss_1$ ), and Ala, Glu, Gly, Gln, Ser, Thr, Val, Tyr constitute subset 2 ( $ss_2$ ), trying to separate monomers of similar functionality into different subsets where possible. Cys, Ile, Asn, Arg were excluded for various reasons outlined below. Ile was excluded from the library design due to the fact that it is difficult to differentiate it from isomeric Leu without additional experimental manipulations. Cys was not used to avoid any potential complications associated with on-resin oxidation of the displayed peptides. Asn was excluded due to the fact that it consistently underwent near quantitative deamidation to Asp during the PAM saponification step. Arg was not used because we found that peptides containing an internal Arg residue tend to yield collision-induced dissociation (CID) fragmentation spectra of poor quality.

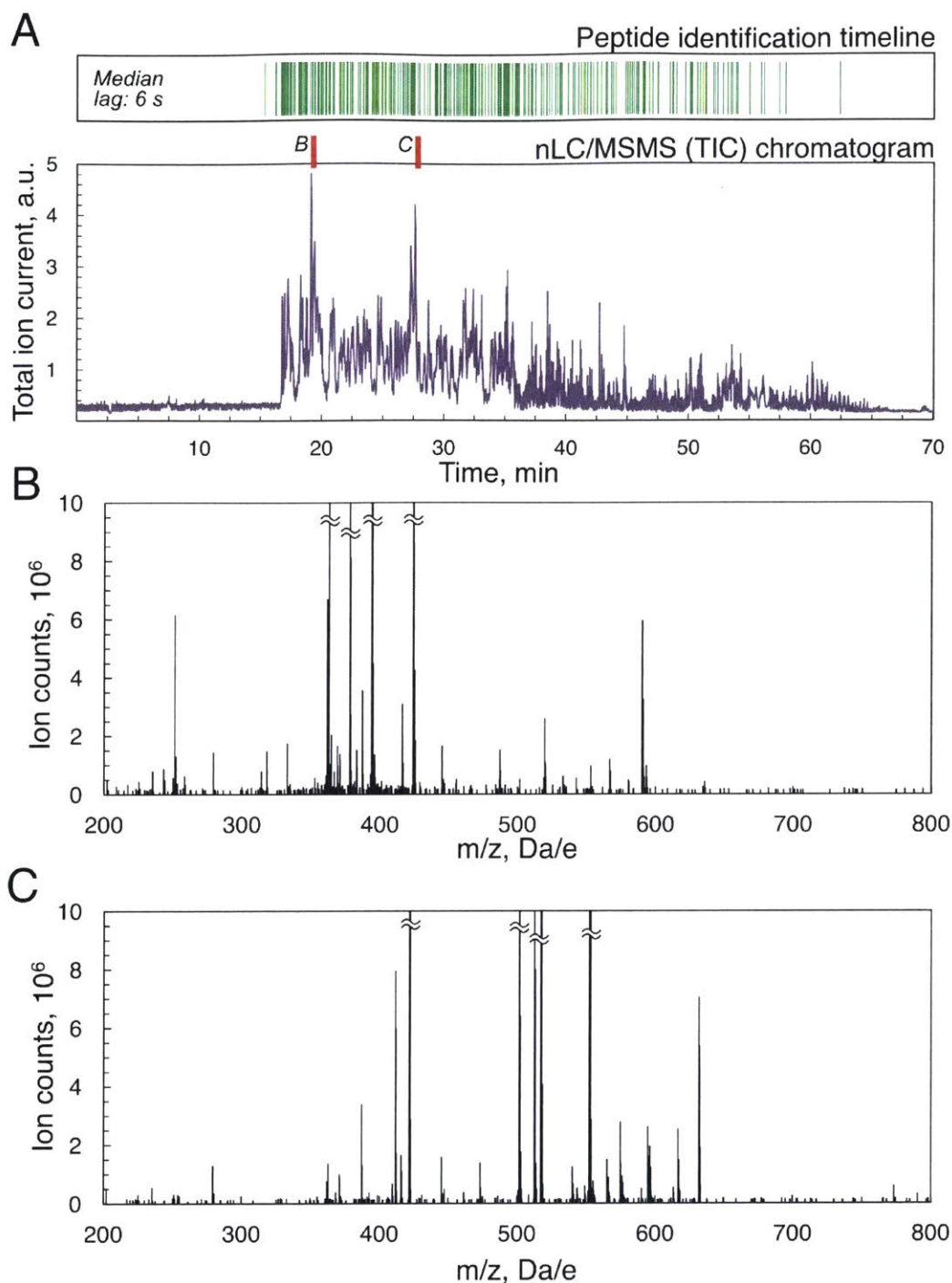
The library of the general design  $ss_1-ss_2-ss_1-ss_2-ss_1-ss_2-ss_1-ss_2-ss_1$ -Gly (**library 1**) was constructed using established Fmoc SPPS procedures<sup>32</sup> on Tentagel S-NH<sub>2</sub> 30  $\mu$ m resin (Rapp Polymere, 0.20 mmol/g amine loading) functionalized with PAM linker for TFA-orthogonal release of peptides from the solid support prior to nLC analysis. Next, we established a standard protocol for nLC/MSMS analysis of mixtures containing peptides from a few hundred beads, and found that nLC can successfully detect and separate peptides for the subsequent MSMS analysis running a 70 minute-long gradient even when only 2% of the total sample volume was submitted for analysis.

CID and higher-energy collisional dissociation (HCD) peptide fragmentation spectra were obtained for each precursor ion automatically in a data-dependent fashion on a Thermo Orbitrap Lumos mass spectrometer. *De novo* peptide sequencing of the acquired data was performed in PEAKS 7 or PEAKS 8 (BioInformatics Solutions Inc.)<sup>33,34</sup> in the following way. CID and HCD spectra were merged, pre-filtered to remove noise, and sequenced allowing Met-oxide as a variable post-translational modification. 15 candidate sequence assignments were created for each merged secondary scan.

Finally, we created automated Python-based routines for post-processing data analysis to eliminate noise, synthetic impurities, duplicates, resolve certain sequencing ambiguities, and to select the best candidate sequence assignment for each merged MS/MS scan. Briefly, in the first step of this process, *a priori* library design criteria are used to eliminate all sequence candidates of length other than 10, not bearing a C-terminal Gly residue or not having a correct monomer in each of the allocated positions. During this step peptide candidates with incorrect amino acid ordering and incorrect dipeptide assignments are

removed from further consideration as a result of the inherent properties of the AMS. For example, one sequence candidate for a peptide containing Asp-Ala subsequence can be interpreted by the sequencing software as Glu-Gly if no fragmentation is observed between these two residues, but since Glu-Gly is not a valid dipeptide in the AMS design, this sequence is rejected during the data analysis step (Fig. 5.3). Next, for each remaining spectrum a single candidate is kept, discarding all other peptides with lower sequencing scores (average local confidence (ALC) scores, from 0 to 99), and duplicate sequences are labeled as non-unique. Finally, the resulting unique sequence assignments are refined further by excluding prominent synthetic impurities that were not eliminated in the previous steps. Peptides containing oxidized Met residues; deamidation of Gln to Glu, which occasionally happens during saponification of PAM ester; sodium adducts; and a few less prominent side-reactions are identified and corresponding sequences are discarded. The list of remaining sequence assignments (or “peptides”) is considered to be the final result of the workflow, and can be utilized further to perform appropriate statistical analysis.

First, to illustrate the ability of nLC to separate mixtures containing hundreds of peptides, we analyzed a 306 bead sample of library 1 peptides as described above (sample statistics are also provided in section 5.3.5). Next, we overlaid the chromatogram with the peptide identification timeline (the list of retention times for parsed sequences identified from the analysis) as shown in Fig. 5.4A. We found that that the chromatogram maintains individual features, and that peptides are identified throughout the duration of the nLC gradient. Additionally, we manually analyzed primary mass spectra obtained for this and other samples to ensure that individual ion signals do not consistently overlap, which becomes an important consideration during the isolation of individual precursor ions for further CID and HCD fragmentation. Two characteristic primary spectra observed throughout the course of the study are shown in Fig. 5.4B and 5.4C. Based on these observations we concluded that the use of nLC for the analysis of complex mixtures of synthetic peptides is fully adequate.



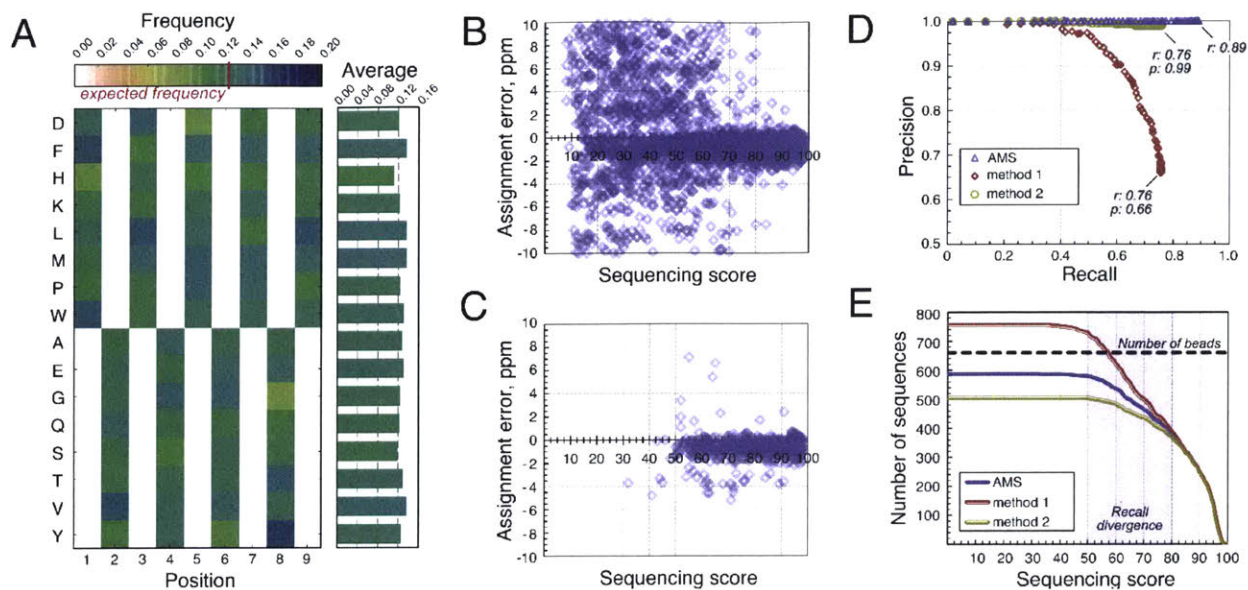
**Figure 5.4. Analysis of nLC efficiency**

A) Overlay of a nLC/MSMS (TIC) chromatogram for the 306 bead sample containing library 1 peptides with the peptide identification timeline (each vertical line represents an identified peptide). Peptides are identified throughout the gradient run time (median time between successive peptide identification events is 6 sec, mean value is 10 sec). B) Primary mass spectrum snapshot at 19.13 min of the LCMS run shown in A). Four most intense peaks are out of scale. The spectrum is clean with little to no m/z overlap between eluting peptides, which allows individual isolation of precursor ions for fragmentation. C) Analogously to B), primary mass spectrum snapshot at 27.61 min.

Having established the workflow for high-throughput decoding of peptide libraries, we sought to evaluate the performance of the method. To this end, we analyzed a naïve, i.e. not preselected in any way, aliquot containing 660 beads bearing library 1 peptides by nLC/MSMS and performed the downstream data analysis as described above. As summarized in Table 5.2, each of the filtration steps played a significant role in filtering out incorrect assignments, and the final sequence list consisted of 587 unique peptides (0.89 sequence identification rate). Manual analysis of the peptides rejected in each filtration step verified that the automated process leads to an identical outcome. As shown in Figure 5.5A, the resulting data set has positional amino acid frequency distribution close to uniform ( $\chi^2$ -test; P-value=0.42,  $\chi^2$  statistic table is provided in section 5.3.4), suggesting that the workflow did not bias the results based on amino acid composition, at least not in an immediately obvious way. Additionally, as Figures 5.5B and 5.5C demonstrate, post-*de novo* noise filtration process primarily removed peptides with low sequencing scores and high assignment mass errors while retaining high confidence and low mass error ones. More specifically, the unprocessed PEAKS output data set had average ALC of  $57 \pm 27$  (one standard deviation) and assignment mass error of  $-0.8 \pm 5.7$  ppm, while in the final peptide data set these parameters were  $82 \pm 14$  and  $-0.7 \pm 1.0$  ppm respectively, reinforcing the notion that low confidence assignments and noise are selected against during the filtration process.

**Table 5.2. Data analysis summary for a sample consisting of 660 beads bearing library 1 peptides**

Data analysis stage	Number of sequences
Unfiltered PEAKS output	5526
Matched	2396
Matched, unique	913
Matched, unique, refined	587



**Figure 5.5. Evaluation of the role of AMS in increasing the confidence in sequencing results**

All data are from the 660-bead sample of library 1 peptides (587 unique sequences identified). A) Color-coded positional amino acid frequency is shown on the left, and the mean amino acid frequencies are plotted on the right. Each non-zero cell in the matrix has the expected value of 0.125, and the observed values map close to it. B) Sequencing quality scatterplot for the unfiltered PEAKS output. C) Sequencing quality scatterplot for the final filtered data set. Most peptides with low sequencing score and/or large assignment errors are removed during the post-sequencing filtration. D) Precision-recall curves for different data filtration methods. Method 1 is inferior to AMS in both precision and recall. E) Total number of sequences recovered as a function of sequencing score for different data filtration methods. Results diverge in the region of medium (50-85) sequencing scores.

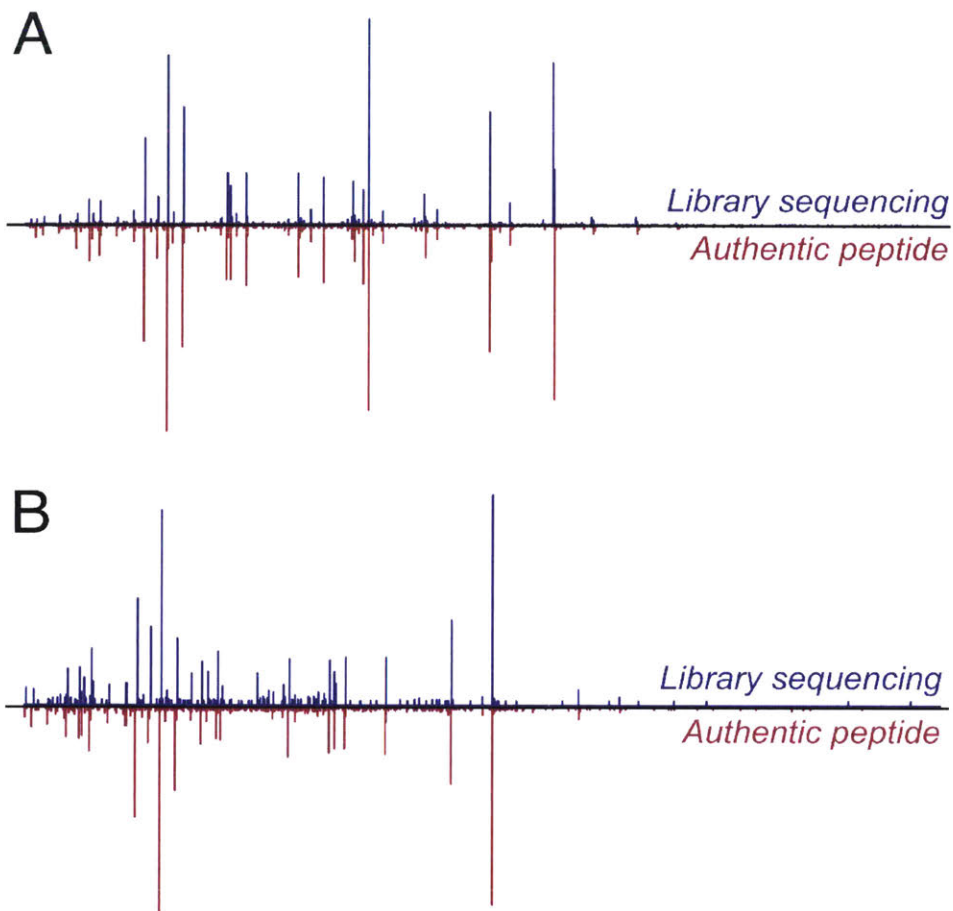


### 5.2.3. Critical Evaluation of the Alternating Monomer Set Design

Next, we turned to assessing the utility of the AMS design for gaining confidence in sequencing results. This was accomplished by parsing the raw PEAKS output for a single AMS library analysis in three different ways, and comparing the outcomes. Operating under the assumption that the AMS does not bias *de novo* sequencing, the data set was parsed assuming the AMS design (i.e., removing all assignments inconsistent with the AMS design), and the resulting data set was taken to be the **ground truth** one. In **method 1** the PEAKS output was parsed *without* imposing the AMS design rules, that is, assuming that the library monomer composition is uniform with respect to all 9 positions, and the monomer set consists of the 16 amino acids comprising the sum of two alternating subsets. The resulting dataset was expected to contain some erroneous assignments, inconsistent with the AMS library design. Analysis of sequences present in the method 1 output but absent from the ground truth data set can be used to estimate the frequency at which AMS eliminates *a priori* incorrect assignments. In **method 2**, we parsed the PEAKS output assuming the AMS design while considering single most confident peptide candidate per spectrum. Comparison of the resulting data set to the ground truth set allowed us to establish the lower bound for the frequency at which the AMS approach recovers sequences containing incorrect dipeptide assignments.

We applied method 1 and method 2 to the PEAKS analysis of a library 1 aliquot (660 bead sample from above; Figure 5.5D and 5.5E). Parsing the dataset enforcing the AMS design yielded a ground truth set of 587 sequences. We found that processing the dataset using method 1 leads to 756 unique sequences, of which 502 peptides were also found in the ground truth dataset, i.e. the overall sequence recall for this experiment was 0.76 (sequence recall is defined as a ratio of decoded sequences also found in the ground truth data set to the total number of peptides in a sample), and the overall precision was 0.66 (by precision here we mean the fraction of identified sequences found in the ground truth dataset). Method 2 yielded 505 unique peptides (recall: 0.76), 500 of which (precision: 0.99) were from the ground truth dataset. Both of these results compare unfavorably against the ground truth dataset: method 1 suffers from lower recall (0.76 vs. 0.89) and lower precision (0.66 vs. 1.00), while method 2 leads to lower recall (0.76 vs. 0.89).

A closer look at these data sets reveals that the data subsets consisting of *de novo* peptides with high sequencing confidence ( $ALC \geq 83$ ) overlapped almost exactly, whereas subsets of medium sequencing scores ( $50 \leq ALC \leq 83$ ) diverged significantly. This behavior is particularly evident from the analysis of the absolute sequence recall as a function of a sequencing score: recall between the data sets diverged in the ALC range of  $\sim 50$  to  $\sim 83$ ; almost no sequences were found at ALC values lower than 50; and no significant divergence was observed at high ALC values (Figure 5.5E).



**Figure 5.6. Evaluating the reliability of library decoding**

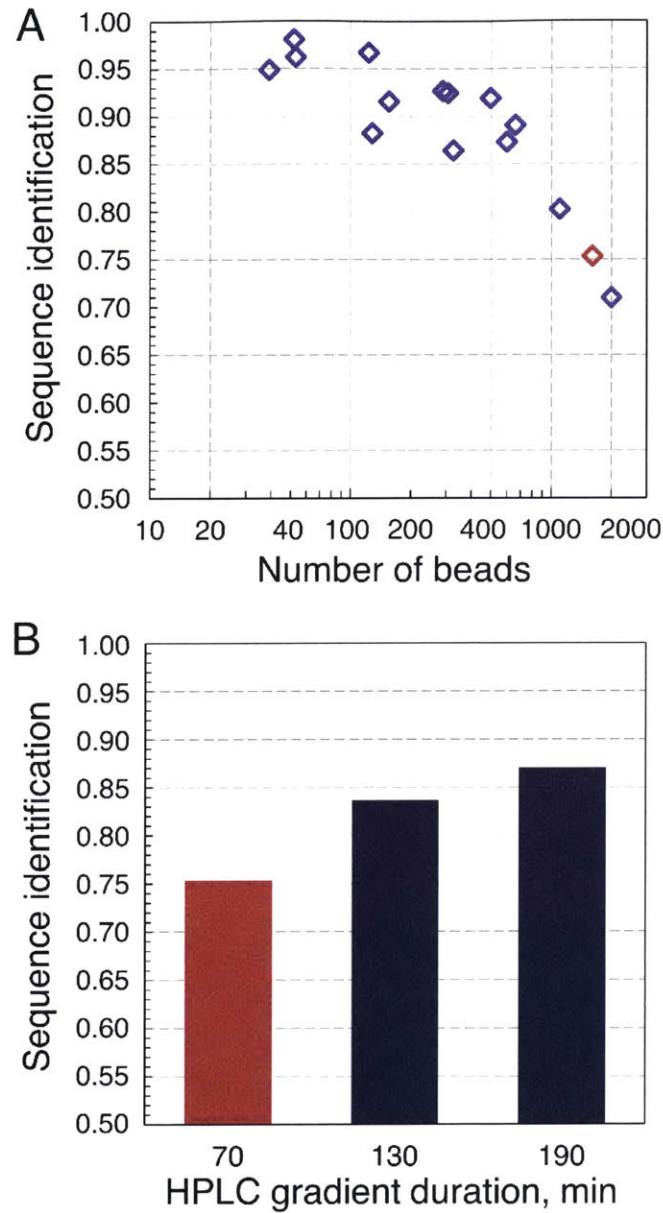
Sequences assigned to spectra from a library analysis were resynthesized and subjected to analogous analytical conditions. Original library-derived spectra are shown in blue; spectra of authentic peptides are displayed in red. A) Overlaid raw CID fragmentation spectra (collision energy = 20.6 eV, precursor ion: 710.82 Da/e) for GCβFLDEVEFPHG peptide ( $\beta = \beta$ -alanine). B) Overlaid raw CID fragmentation spectra (collision energy = 21.1 eV, precursor ion: 654.77 Da/e) for GCβFADASEFPHG peptide.

Based on these outcomes, we concluded that the use of the AMS in combination with post-sequencing filtration routines greatly improves sequencing accuracy by rejecting 254 incorrectly assigned peptides (38% of 660 analyzed beads) and helps recovering sequences that are not considered top candidates by the software (87 peptides recovered, 13% of analyzed beads). These effects are most prominent for peptides of medium sequencing scores. Additionally, our results suggest an alternative strategy for reliable MS/MS sequencing of peptide libraries — namely, discarding peptides with sequencing scores lower than ~85. Such a strategy obviates the need for AMS at the expense of a much lower sequence recall. For instance, parsing the data described above using method 1 and discarding sequences with ALC scores less than 85 yields 310 unique peptides (recall: 0.47, precision: 0.97). A fully analogous experiment performed for a separate 306 bead sample of library 1 peptides (data analysis is provided in section 5.3.5) corroborated conclusions listed here.

To further probe reliability of our approach, we resynthesized individual peptides identified as described above and subjected peptidyl resins to the library analysis conditions. We found that in all cases MSMS spectra of resynthesized peptides agreed well with those observed in a corresponding library sequencing experiment (Fig. 5.6, remaining spectra and characterization of authentic peptides are shown in section 5.3.6), reinforcing our belief in the accuracy of PEAKS assignments.

#### 5.2.4. Analysis of Method Throughput

Next, we studied the potential throughput of our approach using library 1 as a model AMS library. In particular, we were interested to investigate sequence identification rate as a function of sample complexity. To this end, we prepared bead aliquots and manually counted the exact number of beads for samples comprised of less than 1000 beads. For more complex analytes, the number of beads was estimated in three ways, and the average value was assumed. The targeted complexity of the prepared aliquots was 50, 150, 300, 600, or 1200 beads, but the actual bead counts deviated from these numbers. All aliquots were analyzed by nano-LC/MSMS (CID and HCD) running a linear 70 minutes long 2→48% acetonitrile in water gradient. All downstream data processing was performed as described above. As summarized in Fig. 5.7, we found that the simplest bead mixtures yielded highest sequence identification rates: all samples comprised of approximately 50 beads recovered more 95% of analyzed peptides. We also observed gradual decrease in sequence recall values that was generally inversely proportional to sample complexity for the samples comprised of between 50 and 660 beads, and a steep drop in sequence identification rate for more complex samples. We attributed these observations to the phenomenon of “peptide interference”: too many peptides eluting off a nLC column per unit time may cause mass spectrometer to omit analysis of some, and prevent the isolation of individual precursor ions for others. Indeed, when we reanalyzed one of the more complex samples comprised of approximately 1600 beads extending the LC gradient to two or three hours, we obser-



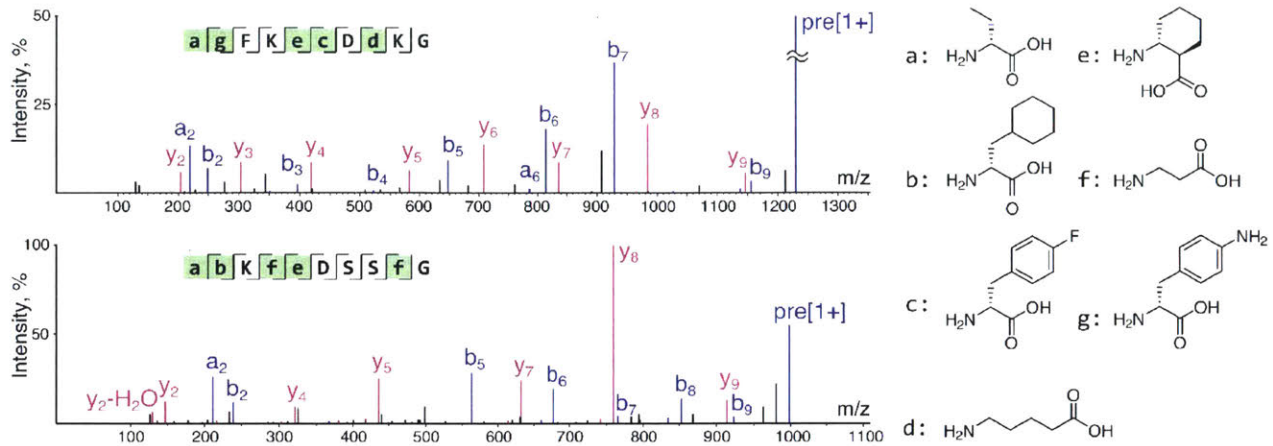
**Figure 5.7. Analysis of the method throughput**

A) Sequence identification rate as a function of sample complexity. Gradual reduction of recall values is observed as samples become more complex. B) Analysis of a 1600 bead library sample (marked red in panel A) under different nLC conditions. Extending the gradient time improves sequence recall.

ved uniform increase of the sequence identification rates from 0.75 to 0.87, which corresponded to extra 185 peptides identified when running a three-hour gradient. These results suggest— assuming that the overall sequence recovery of 85% is deemed satisfactory — that AMS library samples can be routinely analyzed at a rate of at least 600 peptides/hour, and that samples comprised of thousands of peptides can also be successfully decoded by extending LC gradients accordingly.

### 5.2.5. Application to Libraries Containing Non-proteogenic Amino Acid Residues

As mentioned earlier, the ability to sequence peptides comprised of structurally diverse unnatural amino acids is one of the major advantages of mass spectrometry over other decoding techniques. Accordingly, we sought to demonstrate the applicability of our approach towards sequencing libraries comprised of non-proteogenic amino acids. For this study, we prepared another model library (**library 2**), comprised of proteogenic and non-proteogenic  $\alpha$ -,  $\beta$ -, and  $\delta$ -amino acids (synthesis details are in section 5.3.2), and analyzed it using the standard protocol, encoding unnatural amino acids as fixed post-translational modifications on unused proteogenic amino acids during the *de novo* sequencing step in PEAKS. Analysis of three samples containing 183, 212, and 220 beads revealed that library 2 can be decoded nearly as efficiently as library 1, with sequence identification rates reaching 86%. Additionally, the overall quality of fragmentation spectra was comparable to those observed for libraries of proteogenic peptides, as evaluated by manual inspection. Two representative high-quality MSMS spectra and their respective sequence assignments featuring multiple unnatural amino acids are demonstrated in Fig. 5.8. Taken together, these observations indicate that our approach can be utilized to decode structurally diverse libraries of polypeptides and peptidomimetics.



**Figure 5.8. Decoding of the peptide libraries containing non-canonical amino acids**

A, B) Merged, post-processed CID and HCD spectra for a 615.84 Da/e precursor ion and a 499.78 Da/e precursor ions with corresponding assignments and decoded sequences. One-letter encoding of non-natural amino acids highlighted in green are shown on the right.

## 5.2.6. Conclusion

In conclusion, in this work we investigated the feasibility of designing peptide libraries to increase the reliability and throughput of peptide decoding. Accordingly, we developed new MSMS-friendly library designs and established experimental techniques as well as data analysis procedures to fully exploit these features. Our results indicate that the proposed library design principle improves sequencing accuracy and peptide recovery rate relative to the established techniques. We also found that such libraries can be rapidly analyzed by nLC/MSMS, and at least 600 peptides per hour can be decoded while keeping the peptide identification rate above 85%. We anticipate that these methods will be useful for the decoding of “hit” beads yielded from screens of huge libraries, i.e. greater than  $10^8$  members, enabled by monosized microparticulate resins.<sup>35,36</sup> This method is equally applicable to libraries comprised of proteogenic and non-proteogenic amino acids of various structure. Combined with the fact that our approach is experimentally straightforward — no extra chemical manipulations or specialty reagents are required at any stage during library synthesis or analysis — these results suggest that the approach described herein may be a feasible high-throughput alternative to the established library decoding strategies.

We believe that the described method is most valuable in analyzing sparse libraries of great diversity (both in terms of member size and variable region length). We consider libraries of the theoretical diversity on the order of  $10^7$  members consisting of 6 variable positions occupied by one of 16 possible amino acids to be the low diversity ones. MSMS sequencing of hexapeptides from such a library should be more robust than the sequencing of analogous decapeptides, and consecutively, the value of AMS can be expected to decrease for libraries of low diversity. Analyzing peptide libraries of high redundancy<sup>37</sup> using our approach may lead to a challenge of identifying the number of beads on which a given sequence was displayed; this challenge may become prominent during decoding of selection hits where some sort of sequence convergence may be natural. Although technically possible, such identification was not attempted in this work. Finally, we anticipate that the presented experimentation and the reasoning behind it are not inherently limited to OBOC libraries, and can be used, for instance, to decode solution phase peptide libraries, which is a topic of our ongoing investigations.

## 5.3. Experimental

### 5.3.1. General

Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) and proteogenic N<sup>α</sup>-Fmoc protected D-amino acids were purchased from Creosalus Inc. (KE). (1R,2R)-Fmoc-2-amino-1-cyclopentane carboxylic acid (Acp), (1R,2R)-Fmoc-2-amino-1-cyclohexane carboxylic acid (Ach), Fmoc- $\alpha$ -D-aminobutyric acid (Abu), Fmoc- $\beta$ -alanine ( $\beta$ ), Fmoc- $\delta$ -aminovaleric acid (Ava), Fmoc-D-cyclohexylalanine (Cya), Fmoc-D-4-fluorophenylalanine (Fph), Fmoc-D-4-aminophenylalanine (Aph), Fmoc-D-naphthylalanine (Nap), Fmoc-D-perfluorophenylalanine (Pfp), Fmoc-D-4-phenyl-phenylalanine (Phf), and Fmoc-D-3,4-dichlorophenylalanine (Cph) were from Chem-Impex International (Wood Dale, IL). N<sup>α</sup>-Boc-Glycyl ester of (4-hydroxymethyl)phenylacetic acid was purchased from PolyPeptide Group (Strasbourg, France). 30  $\mu$ m monosized Tentagel S NH<sub>2</sub> resin was purchased from Rapp Polymere (Tuebingen, Germany).

*N,N*-Dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile were from EMD Millipore. Triisopropyl silane (TIPS), and 1,2-ethanedithiol (EDT) were from Alfa Aesar. Solvents for HPLC-MS were purchased from Fluka. All other reagents were purchased from Sigma-Aldrich.

All reagents were used as received. Proteogenic amino acid residue encodings are reserved for corresponding D-amino acids throughout the chapter. In other words, libraries were synthesized from D-amino acids where applicable.

### 5.3.2. Peptide and Library Synthesis

Individual peptides and peptide libraries were synthesized using established batch Fmoc SPPS techniques. Briefly, Fmoc-protected amino acids were dissolved in appropriate volume of 0.38 M HATU in DMF, and activated with 3 eq. of DIEA. Couplings were allowed to proceed for 10 to 30 minutes at room temperature with periodical gentle rocking of the resin suspension. N<sup>α</sup>-Fmoc protecting group removal was achieved by two 3 min treatments with 20% (v/v) piperidine in DMF. Side chain protection was as follows: Asn(Trt), Asp(OtBu), Cys(Trt), Gln(Trt), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu).

Library 1 was synthesized by manual split-pool Fmoc SPPS. Fully assembled library 1 contained peptide ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>- $\beta$ CKQSDCLAGSVCG-PNGFCG-CONH-Tentagel at 0.02 mmol/g loading, and ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-ss<sub>2</sub>-ss<sub>1</sub>-Gly-PAM-CONH-Tentagel at 0.18 mmol/g. ss<sub>1</sub> was any amino acid of the following set: Asp, Phe, His, Lys, Met, Pro, Trp, Leu. Analogously, ss<sub>2</sub> was Ala, Glu, Gly, Gln, Ser, Thr, Val, Tyr.



Resin (6 g of Tentagel S-NH<sub>2</sub> 30 μm resin from Rapp Polymere, 0.20 mmol/g amine loading) was first coupled to 4.5 mmol of Boc-Gly-PAM-OH and 0.5 mmol of Fmoc-Gly-OH to achieve dual functionalization of the resin. Then Boc group was removed by treating with trifluoroacetic acid (TFA, twice, each with 40 mL for 2 minutes). The resin was washed with DMF (3 times, 50 mL each) and DCM (3 times, 50 mL each). 20 mL of 0.4 M Alloc-OSu in dry DCM and 3 mL of DIEA was added to the resin to functionalize the unmasked H<sub>2</sub>N-Gly-PAM with Alloc protection group. After removal of the Fmoc group (two 3 minute treatments with 20% (v/v) piperidine in DMF), the constant region (βCKQSDCLAGSVCGPNGFCG) was assembled following standard Fmoc protocols. Then, Alloc protection on Gly-PAM was removed with 400 mg of Pd(PPh<sub>3</sub>)<sub>4</sub> and 8 mL of phenylsilane in 32 mL of dry DCM for 20 minutes at room temperature, and split-pool Fmoc SPPS ensued. The fully assembled library was deprotected with 2.5% (v/v) EDT, 2.5% (v/v) H<sub>2</sub>O, and 1% (v/v) TIPS in TFA for 2 hours at room temperature. The deprotected library was washed with TFA, DCM, diethyl ether (five times each) and dried under reduced pressure.

Library 2 was synthesized on a spatially segregated Tentagel resin prepared using an unpublished functionalization protocol (*Z. Gates, A. Vinogradov, B. Pentelute, manuscript in preparation*). In this case, surface accessible area comprised 50% of the total bead volume. Interior of the resin bore H<sub>2</sub>N-(X)<sub>9</sub>-Gly-PAM-CONH-Tentagel peptide (X: D, K, F, Y, S, β, Ava, Abu, Acp, Ach, Cya, Fph, Aph, Nap, Pfp, Phf, Cph; expected frequency for D, K, F, Y, S and Abu was 10%, and for the other monomers it was 3.6%) at 0.20 mmol/g loading. Surface accessible part of the resin contained H<sub>2</sub>N-(X)<sub>9</sub>-CONH-Tentagel peptide at 0.02 mmol/g loading. Otherwise, synthesis of library 2 followed the procedure outlined above for library 1.

Peptides GCβWLSADEFPHG, GCβFLDEGYGPWG, GCβWLDEDTFMGG, GCβWLDTDPFPHG, GCβWLDTSEFPHG, GCβWLDPSLFPHG, GCβWADESAGPWG, GCβFLDEVEFPHG, and GCβFADASEFPHG were synthesized on Tentagel S-NH<sub>2</sub> 30 μm resin functionalized with Boc-Gly-PAM following procedures outlined above.

### 5.3.3. Library Cleavage and Nano-LC/MSMS Analysis

*PAM ester saponification.* A deprotected, dried peptidyl resin bearing library peptides was suspended in water and sonicated until homogeneity. 2 μl aliquots containing 50-2000 beads were taken for analysis. The number of beads in each sample was counted manually under a bright-field microscope for samples containing less than 1000 beads or estimated for more complex suspensions. PAM ester saponification was carried out by adding 4 μl of 1M NaOH in water to the peptidyl resin suspensions. The cleavage reaction was allowed to proceed for 45 minutes at room temperature, after which it was quenched with 40 μl of 1% (v/v) TFA in water. The beads were spun down (1 min at 14000 rpm), and the supernatant was carried forward for nano-LC analysis.

*Nano-LC/MSMS analysis of library bead mixtures.* Nano-LC/MSMS analysis was performed on Thermo Fisher Orbitrap Fusion Lumos Tribrid Mass Spectrometer coupled to Thermo Fisher EASY-nLC 1200 System equipped with Acclaim PepMap RSLC C18 column (250 mm x 75 µm ID, 2 µm 100Å silica). The standard nano-LC method was run at 40 °C and a flow rate of 300 nL/min with the following gradient: 1% of 80% acetonitrile in water with 0.1% formic acid added (solvent B') in water containing 2% methanol and 0.1% FA (solvent A') ramping linearly to 5% B' in A' over 2 minutes, followed by 5-61% B' in A' ramping linearly over 68 minutes, followed by 61-99% B' in A' ramping linearly over 5 minutes and finally 99% B' in A' for 5 minutes. MSMS acquisition over the course of the method duration was performed in a data-dependent style (Top N=15, z=2-10, intensity threshold = 10<sup>5</sup>) with a dynamic precursor exclusion for 20 seconds after each scan. CID and HCD fragmentation spectra were acquired for every selected precursor ion. Orbitrap was used as a detection method for both primary (resolution=120000) and secondary (resolution=30000) mass spectra.

*De novo peptide sequencing in PEAKS.* De novo peptide sequencing was performed in PEAKS 7 or PEAKS 8 from Bioinformatics Solutions Inc. (ON, Canada). Sequencing data obtained as described in the section above was refined as follows. HCD and CID scans were merged within a 0.2 minute and 10 ppm window, mass precursor correction was performed, and primary mass filtration was performed as appropriate. De novo sequencing was performed allowing 15 ppm assignment errors and 0.05 Da individual fragment mass errors. For library 1 methionine oxide was selected as a variable PTM. For library 2, 2-aminobutanoic acid was sequenced as a variable PTM on Gly, while Acp, Ach, Cha, Fph, Aph, Nap, Pfp, Phf and Cph were sequenced as fixed PTMs on unused Pro, Cys, Thr, Met, Leu, Asn, Glu, Trp and His, respectively. 15 candidate sequences were obtained for each preprocessed scan.

*Post-de novo data analysis.* After performing de novo sequencing as specified above, the list of all peptide candidates and a single top candidate peptide lists were exported as .csv files and further analyzed in Python 2.7.10. A number of open source libraries were used for parsing data and analyzing the outcomes: namely, Numpy,<sup>38</sup> Pandas,<sup>39</sup> Scipy,<sup>40</sup> and matplotlib.<sup>41</sup>

First, working with a list of all peptide candidates, sequences not matching library design rules (sequence length, C-terminal Gly, correct monomers in every position) were eliminated from further consideration. Next, for scans with multiple remaining sequence candidates, a single peptide with highest ALC score per scan was retained, while the rest were excluded. After this, remaining sequences were labeled as “unique” and “non-unique” based on a number of criteria. Comparing peptides to each other, a number of attributes, including retention time difference, measured mass difference, and sequence distance (Hamming distance) were considered, and if peptides were found to be similar to each other, the one with lower ALC score was labeled as “non-unique”. Finally, the list of unique sequences was refined further by a) identifying pairs of peptides with Hamming distance=1 and a Q/E mismatch, and eliminating sequences

bearing “E” in the mismatched position; b) identifying pairs of H<sup>+</sup>/Na<sup>+</sup> sequences and removing Na<sup>+</sup>-derived ones; c) removing Met-oxide containing sequences if unoxidized ones were also present; d) identifying pairs of peptides with Hamming distance=1 and a Q/A or E/A mismatch, and eliminating sequences bearing “A” in the mismatched position (unidentified, fairly prominent side reaction). The list of remaining sequences was considered to be the end result of library decoding.

#### 5.3.4. Amino Acid Frequency Distribution $\chi^2$ -Test for the Sample of 660 Library 1 Beads

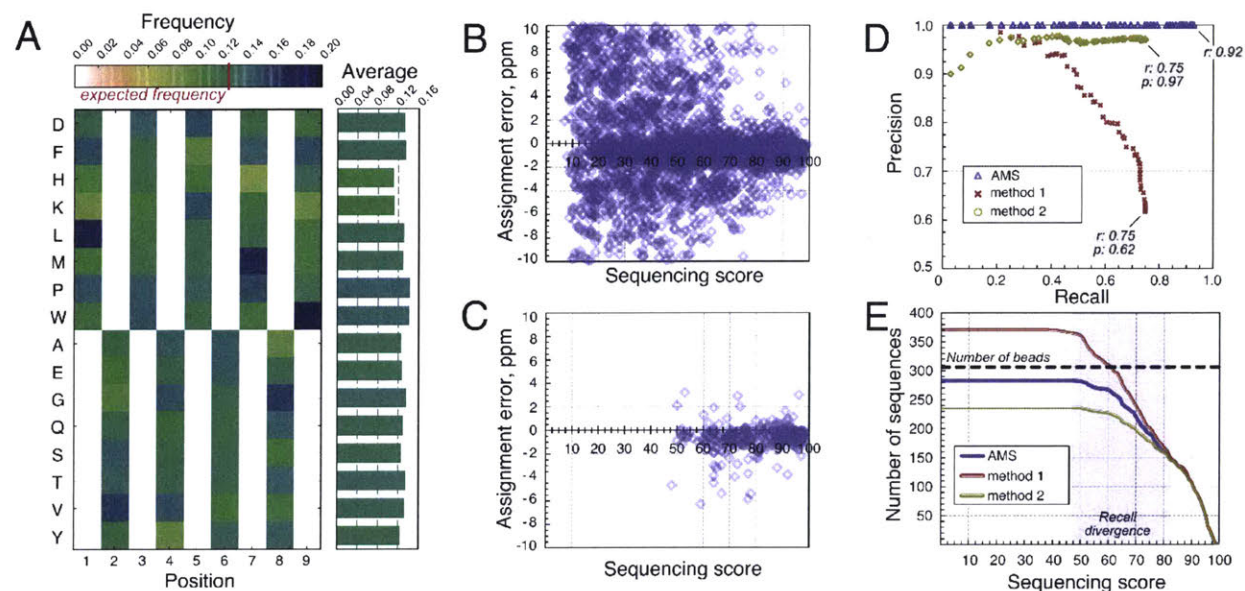
As Table 5.3 indicates, we found no evidence that there is statistically significant difference between assigned and designed amino acid frequencies (P-value=0.42) in the studied sample (n=587).

**Table 5.3.  $\chi^2$ -test statistic table for the amino acid distribution in the sample of 660 library 1 beads**

	pos1	pos2	pos3	pos4	pos5	pos6	pos7	pos8	pos9	Sum	P-value
D	0.18		1.54		5.12		0.74		0.39	7.97	0.09
F	2.92		1.76		0.60		0.60		0.60	6.47	0.17
H	4.60		0.00		2.09		1.20		0.39	8.28	0.08
K	0.26		0.74		0.00		0.18		0.55	1.74	0.78
L	0.04		2.53		0.18		0.96		2.53	6.23	0.18
M	0.16		0.79		1.26		0.60		0.60	3.41	0.49
P	0.74		0.55		1.26		0.60		1.47	4.62	0.33
W	2.53		0.55		0.09		0.04		0.01	3.22	0.52
A		0.08		0.43		0.00		0.03		0.54	0.91
E		0.04		0.08		0.43		0.29		0.84	0.84
G		0.03		1.01		1.01		6.82		8.88	0.03
Q		0.60		0.08		0.55		0.96		2.18	0.53
S		0.08		2.09		0.04		0.39		2.59	0.46
T		0.26		0.03		0.03		1.84		2.15	0.54
V		2.53		0.04		0.60		0.01		3.17	0.37
Y		1.76		0.09		2.82		5.80		10.47	0.01
Sum	11.42	5.37	8.48	3.84	10.60	5.48	4.90	16.13	6.54	72.76	0.42
P-value	0.12	0.62	0.29	0.80	0.16	0.60	0.67	0.02	0.48	0.42	

### 5.3.5. Analysis of a Library 1 Sample Containing 306 Beads

A sample containing 306 beads from library 1 was analyzed as described above. Additional data analysis was performed following the procedure for a 660 bead sample outlined in the discussion section of this Chapter and is summarized in Figure 5.9. Results of this analysis corroborate those obtained for the 660 bead sample.



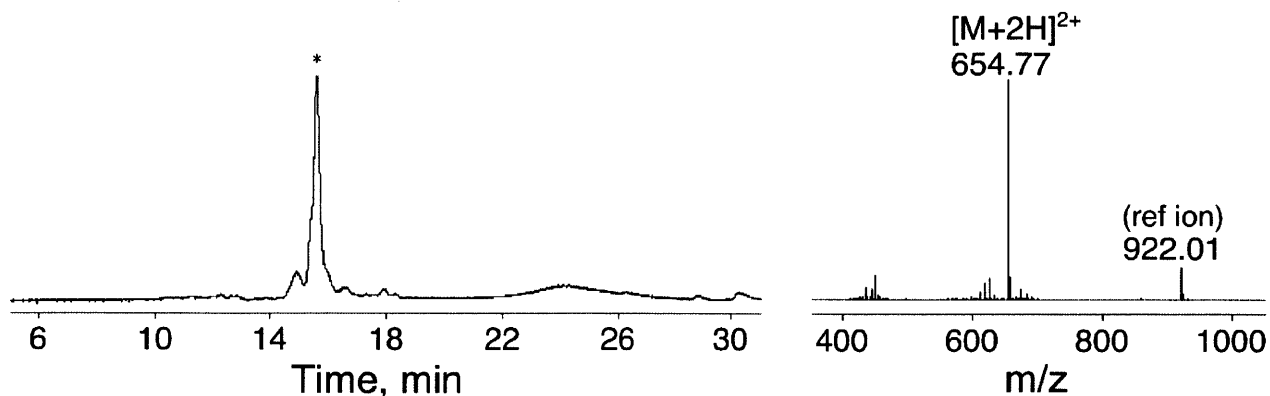
**Figure 5.9. Evaluation of the role of AMS in increasing the confidence in sequencing results**

All data are from the 306-bead sample (283 unique sequences identified). A) Color-coded positional amino acid frequency is shown on the left, and the mean amino acid frequencies are plotted on the right. Each non-zero cell in the matrix has the expected value of 0.125, and the observed values map close to it. B) Sequencing quality scatterplot for the unfiltered PEAKS output. C) Sequencing quality scatterplot for the final filtered data set. Most peptides with low sequencing score and/or large assignment errors are removed during post-processing. D) Precision-recall curves for different data filtration methods. Method 1 is inferior to AMS in both precision and recall. E) Total number of sequences recovered as a function of sequencing score for different data filtration methods. Results diverge in the region of medium (45-85) sequencing scores.

### 5.3.6. Matching Fragmentation Spectra from Library and Authentic Peptides

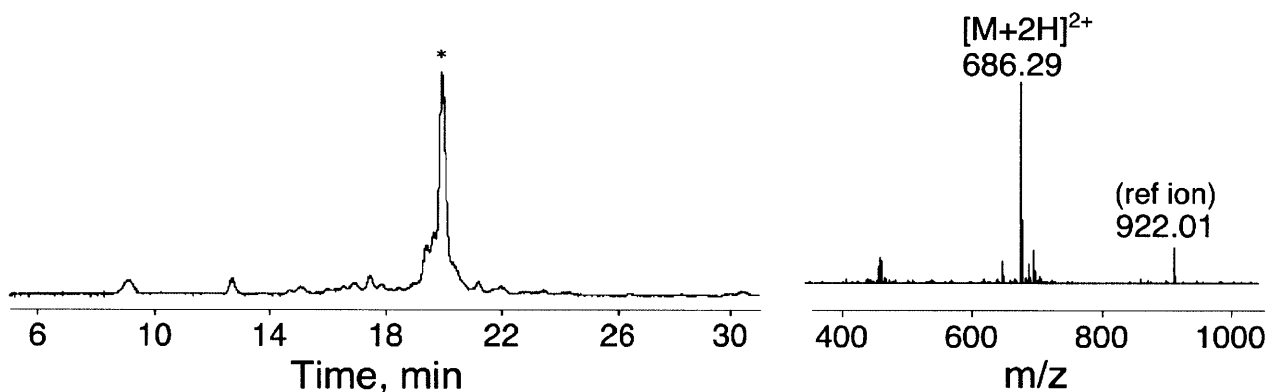
Analytical control over peptide quality was achieved by cleaving an aliquot of deprotected peptidyl resin with 10% (v/v) trifluoromethanesulfonic acid, 10% (v/v) *p*-cresol and 10% (v/v) thioanisole in TFA for 15 minutes on ice. Peptide carboxylates obtained in this way were analyzed on Agilent 6550 iFunnel Q-TOF LC-MS coupled to Agilent 1290 Infinity HPLC system equipped with Phenomenex Jupiter C4 column (150 x 1.0 mm ID, 5  $\mu$ m 300Å silica). The analysis method was as follows. At 40 °C and a flow rate of 0.1 mL/min, the following gradient was used: 1% acetonitrile with 0.1% formic acid added (FA, solvent B) in water with 0.1% FA (solvent A) for 5 min, 1-51% B in A ramping linearly over 25 min, and then 51-70% B in A ramping linearly over 5 minutes. 10–20 ng/ $\mu$ L peptide solutions were subject to analysis. All chromatograms below are plots of total ion current (TIC) versus time.

For the experiments described below, a corresponding library was analyzed as described above, except MSMS conditions differed in the following way. Agilent 6550 iFunnel Q-TOF LC-MS coupled to Agilent 1290 Infinity HPLC system equipped with Phenomenex Jupiter C4 column (150 x 1.0 mm ID, 5  $\mu$ m 300Å silica) was used to perform the analysis. The HPLC was performed at 40°C and a flow rate of 0.1 mL/min with the following gradient: 1% acetonitrile with 0.1% formic acid added (FA, solvent B) in water with 0.1% FA (solvent A) for 5 min, 1-41% B in A ramping linearly over 65 min, and then 70% B in A for 7 minutes. The entirety of the sample was analyzed. MSMS acquisition over the course of the method duration was performed in a data-dependent style (Top N=10,  $z=2-10$ , intensity threshold = 7500) with a dynamic precursor exclusion for 30 seconds after each scan. CID fragmentation spectra at two different collision energies were acquired for every selected precursor ion. Authentic peptides were analyzed in an analogous fashion. Appropriate raw CID spectra were manually overlaid and are demonstrated in Fig. 5.19-5.25.



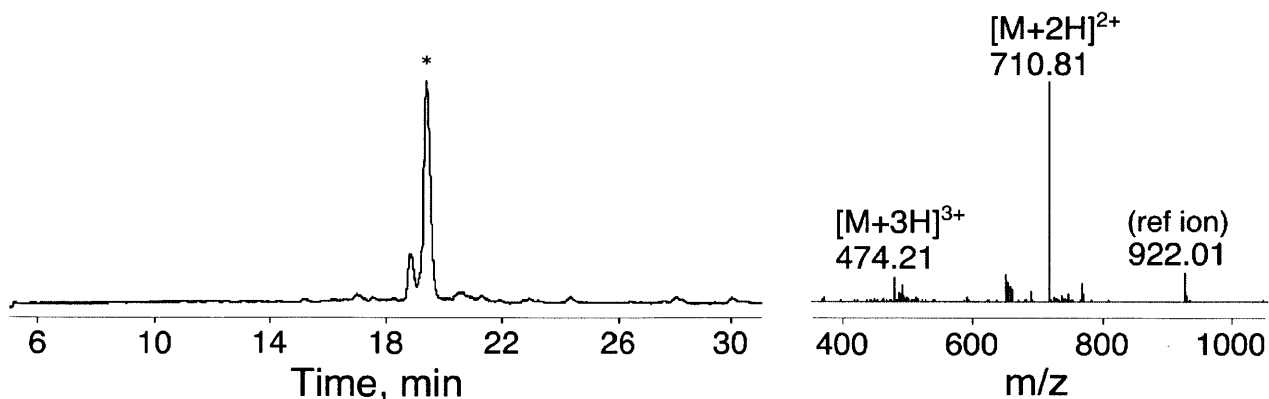
**Figure 5.10. HPLC-MS analysis of crude H<sub>2</sub>N- GCβFADASEFPHG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβFADASEFPHG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1307.51 Da, Observed mass = 1307.53 Da.



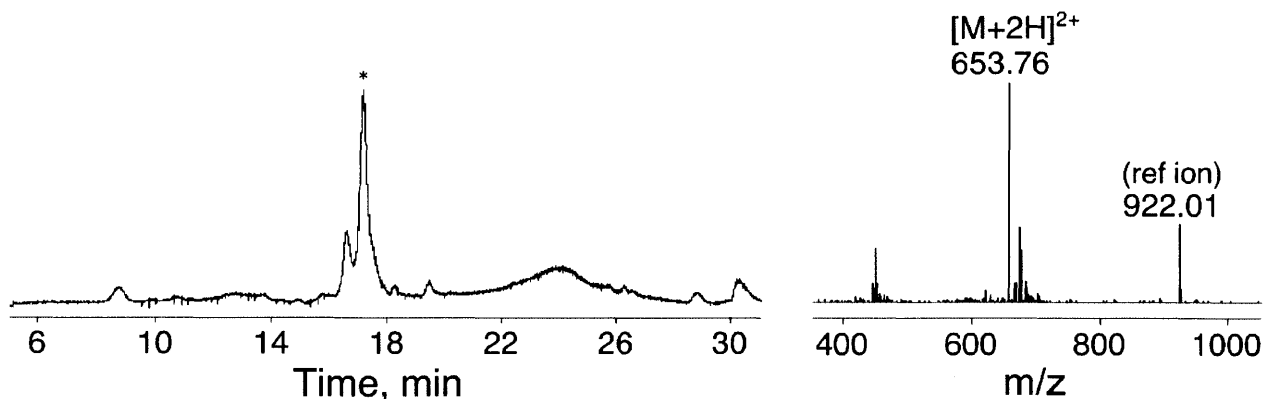
**Figure 5.11. HPLC-MS analysis of crude H<sub>2</sub>N- GCβFLDEGYGPWG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβFLDEGYGPWG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1370.55 Da, Observed mass = 1370.57Da.



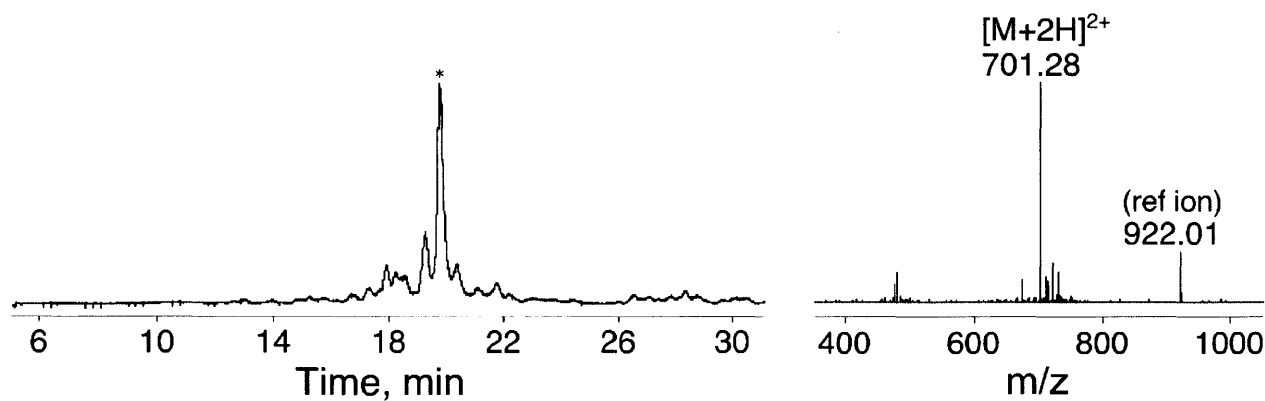
**Figure 5.12. HPLC-MS analysis of crude H<sub>2</sub>N- GCβFLDEVEFPHG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβFLDEVEFPHG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1419.60 Da, Observed mass = 1419.61 Da.



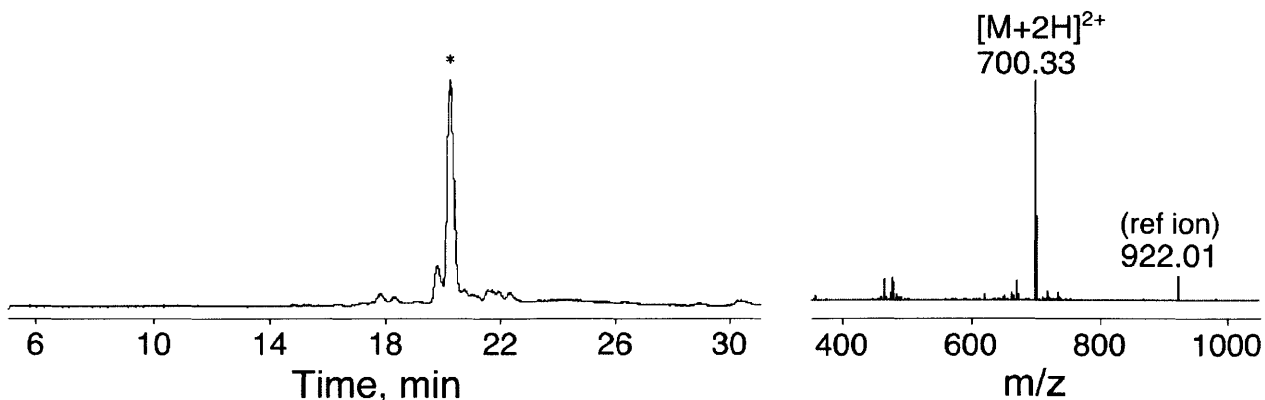
**Figure 5.13. HPLC-MS analysis of crude H<sub>2</sub>N- GCβWADESAGPWG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβWADESAGPWG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1305.50 Da, Observed mass = 1305.51 Da.



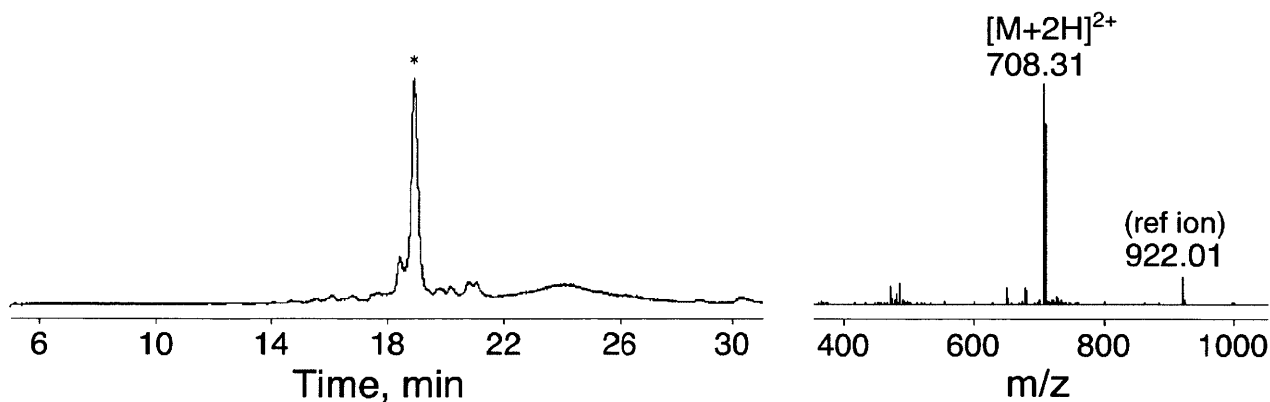
**Figure 5.14. HPLC-MS analysis of crude H<sub>2</sub>N- GCβWLDEDTFMGG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβWLDEDTFMGG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1400.53 Da, Observed mass = 1400.54 Da.



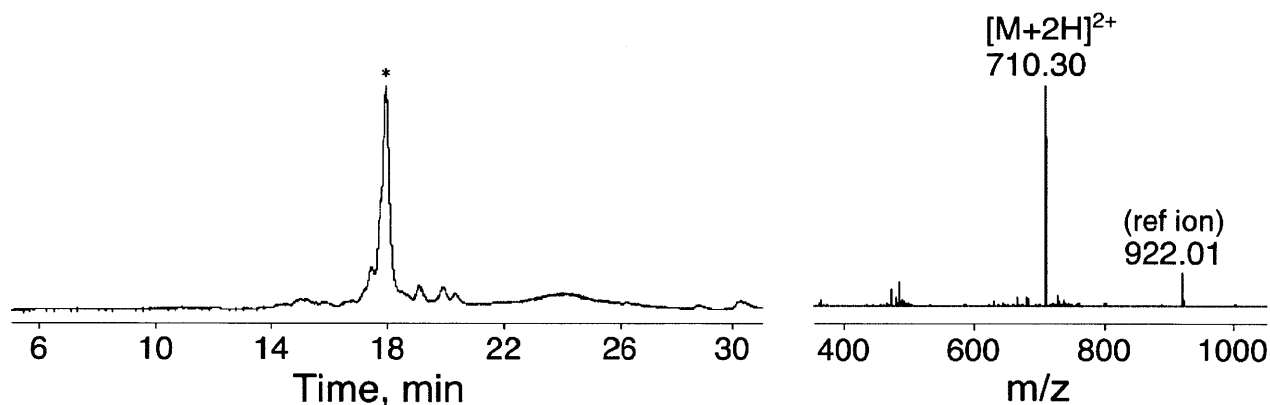
**Figure 5.15. HPLC-MS analysis of crude H<sub>2</sub>N- GCβWLDPSLFPHG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβWLDPSLFPHG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1398.63 Da, Observed mass = 1398.64 Da.



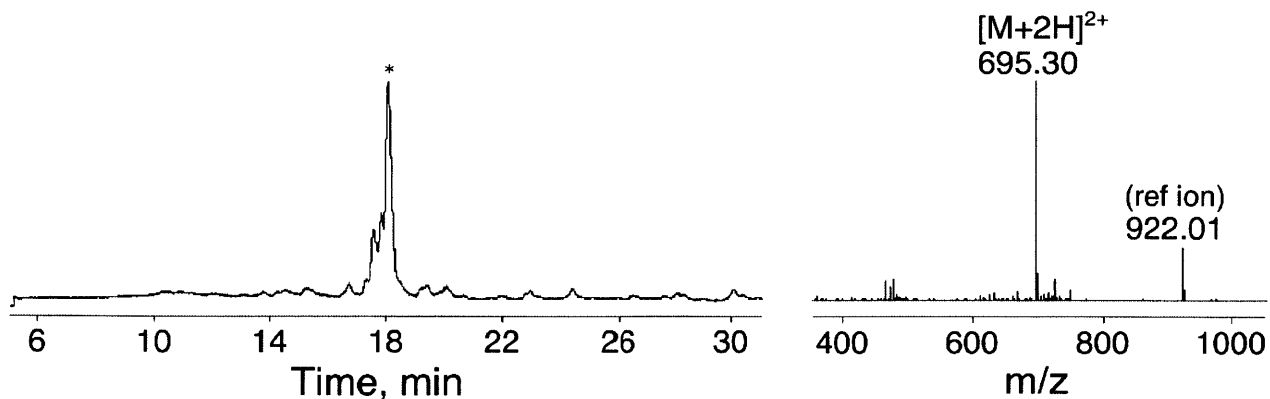
**Figure 5.16. HPLC-MS analysis of crude H<sub>2</sub>N- GCβWLDTPFPFHG-COOH**

HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβWLDTPFPFHG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1414.59 Da, Observed mass = 1414.61 Da.



**Figure 5.17. HPLC-MS analysis of crude H<sub>2</sub>N- GCβWLDTSEFPFHG -COOH**

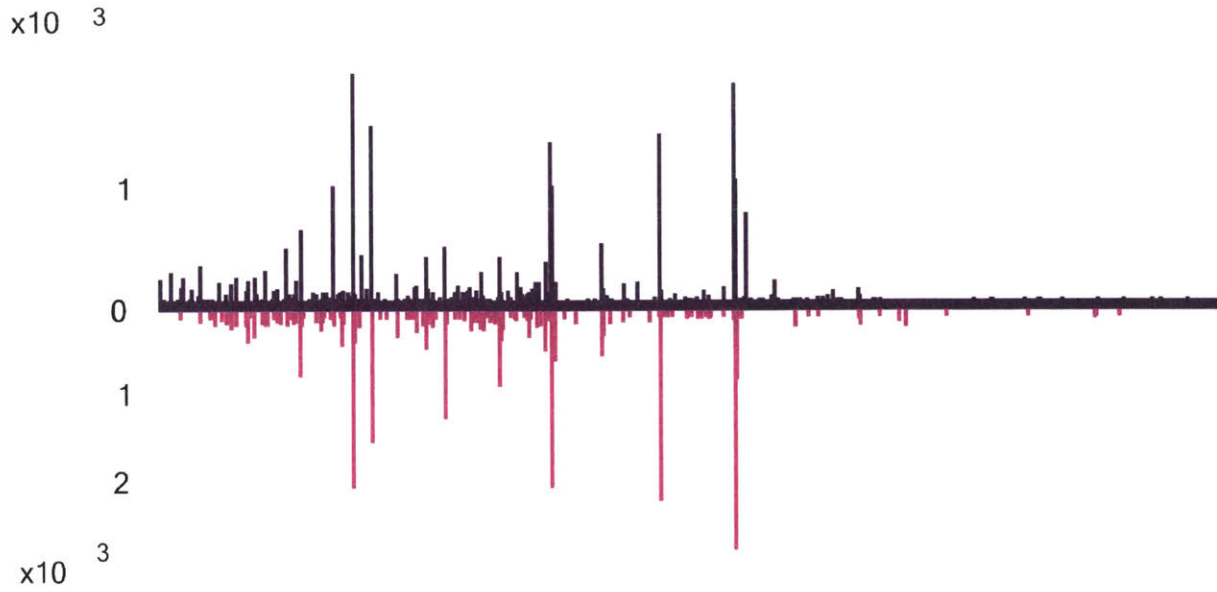
HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβWLDTSEFPFHG -COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1418.58 Da, Observed mass = 1418.59 Da.



**Figure 5.18. HPLC-MS analysis of crude H<sub>2</sub>N- GCβWLSADEFPHG-COOH**

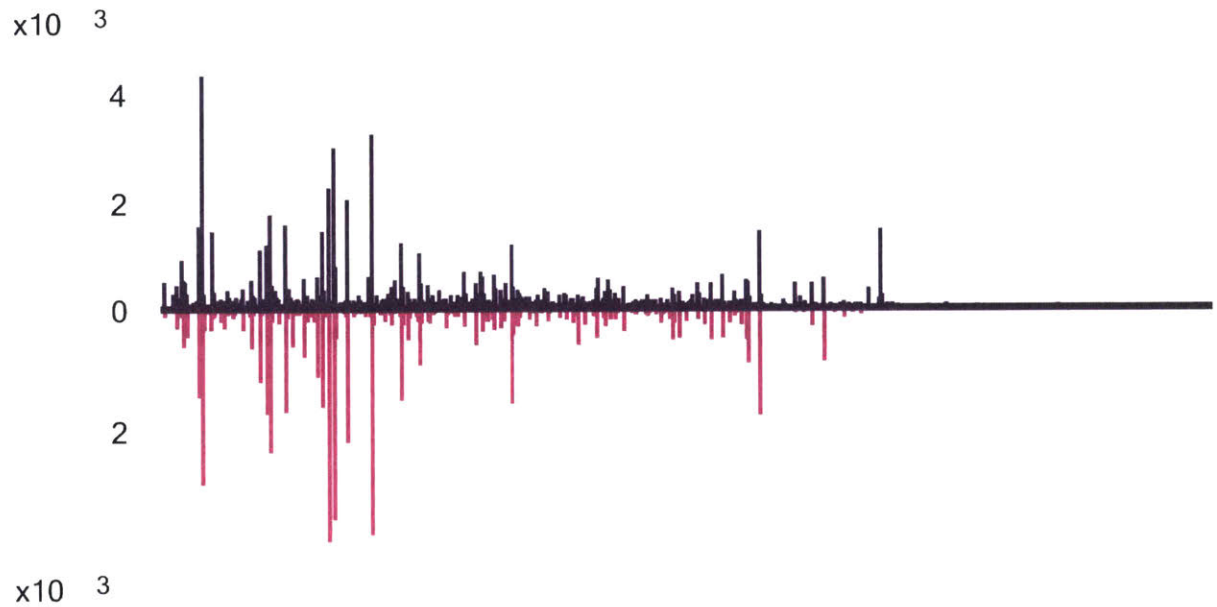
HPLC-MS (TIC) chromatogram for the crude H<sub>2</sub>N- GCβWLSADEFPHG-COOH on the left with a mass spectrum of major peak (labeled with an asterisk) on the right. Calc. monoisotopic mass = 1388.57 Da, Observed mass = 1388.59 Da.





**Figure 5.19. Overlaid raw CID fragmentation spectra for GCβWLSADEFPHG**

Overlaid raw CID fragmentation spectra (collision energy = 20.0 eV, precursor ion: 695.30 Da/e) for GCβWLSADEFPHG ( $\beta$  =  $\beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).



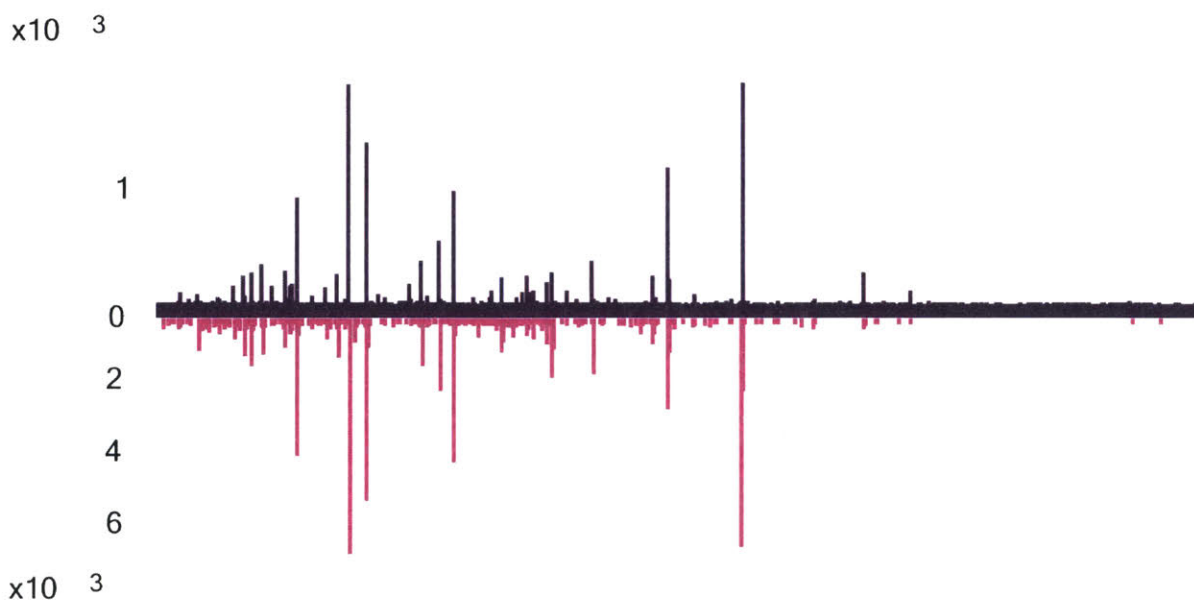
**Figure 5.20. Overlaid raw CID fragmentation spectra for GCβFLDEGYGPWG**

Overlaid raw CID fragmentation spectra (collision energy = 22.2 eV, precursor ion: 686.29 Da/e) for GCβFLDEGYGPWG ( $\beta$  =  $\beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).



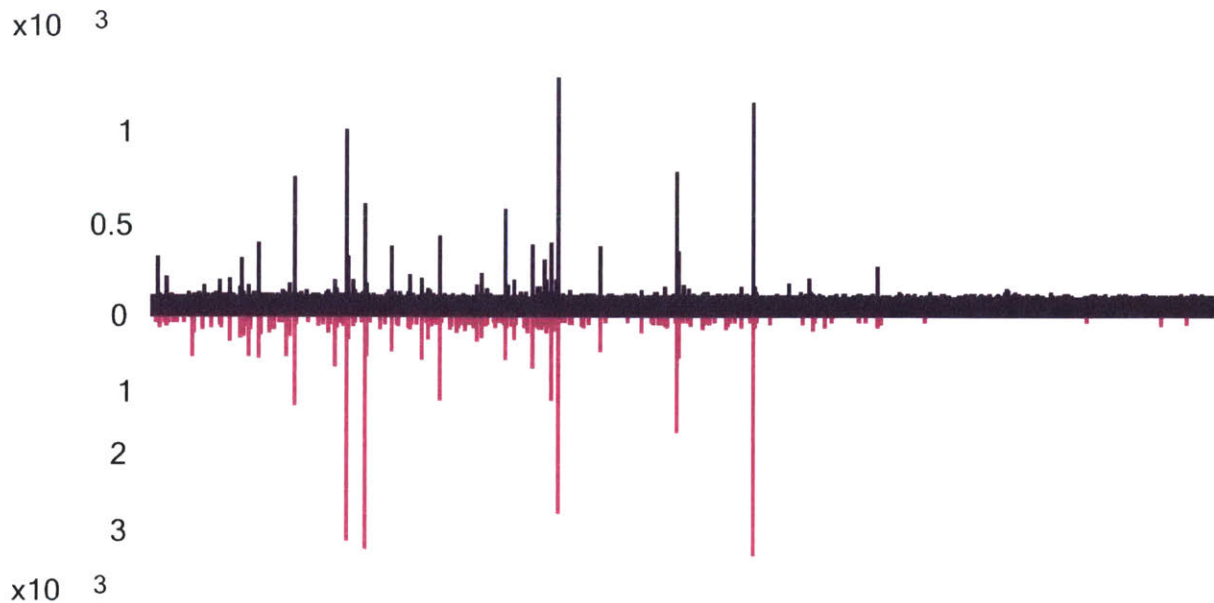
**Figure 5.21. Overlaid raw CID fragmentation spectra for GC $\beta$ WLDEDTFMGG**

Overlaid raw CID fragmentation spectra (collision energy = 22.8 eV, precursor ion: 701.28 Da/e) for GC $\beta$ WLDEDTFMGG ( $\beta$  =  $\beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).



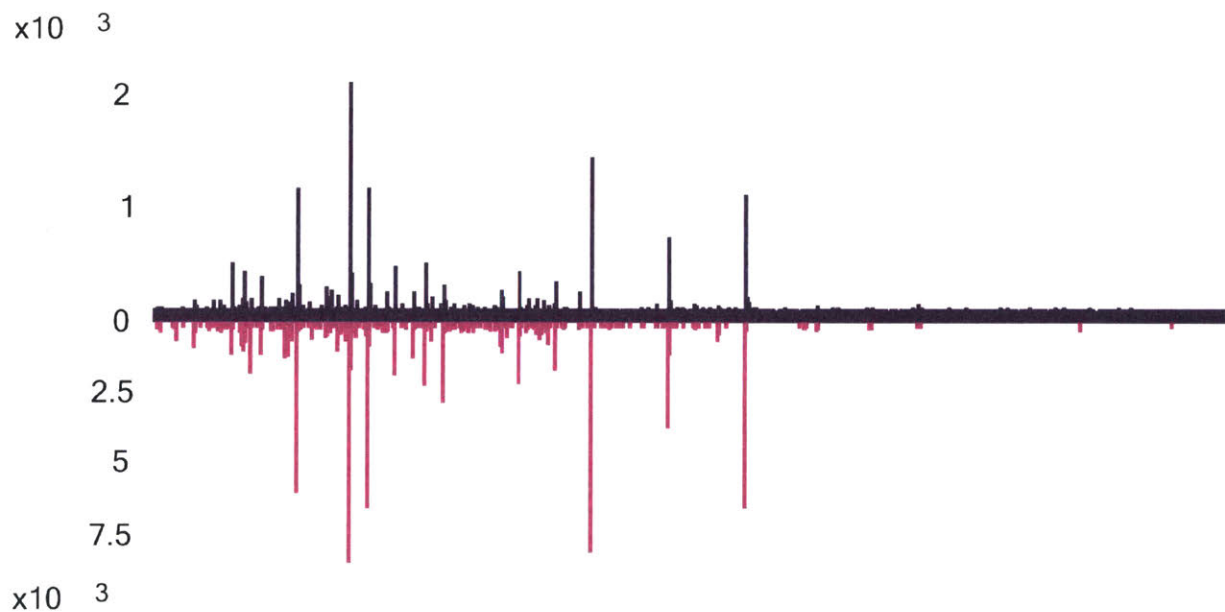
**Figure 5.22. Overlaid raw CID fragmentation spectra for GC $\beta$ WLDTDPFPHG**

Overlaid raw CID fragmentation spectra (collision energy = 23.0 eV, precursor ion: 708.31 Da/e) for GC $\beta$ WLDTDPFPHG ( $\beta$  =  $\beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).



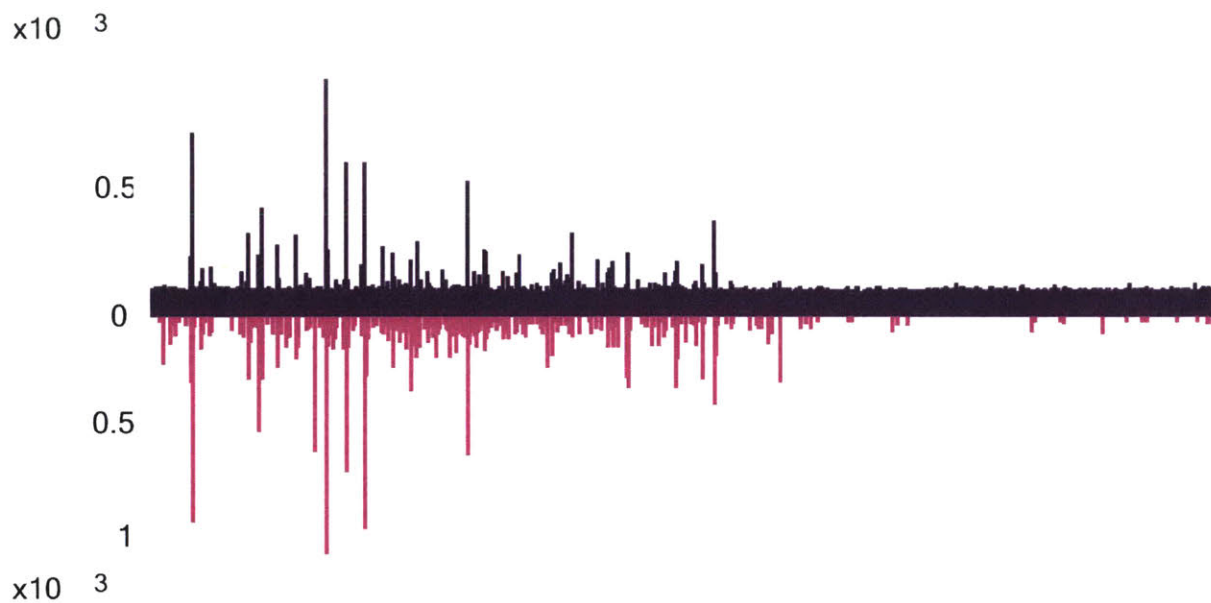
**Figure 5.23. Overlaid raw CID fragmentation spectra for GCβWLDTSEFPHG**

Overlaid raw CID fragmentation spectra (collision energy = 20.6 eV, precursor ion: 710.30 Da/e) for GCβWLDTSEFPHG ( $\beta = \beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).



**Figure 5.24. Overlaid raw CID fragmentation spectra for GCβWLDPSLFPHG**

Overlaid raw CID fragmentation spectra (collision energy = 22.7 eV, precursor ion: 700.33 Da/e) for GCβWLDPSLFPHG ( $\beta = \beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).



**Figure 5.25. Overlaid raw CID fragmentation spectra for GCβWADESAGPWG**

Overlaid raw CID fragmentation spectra (collision energy = 21.0 eV, precursor ion: 653.76 Da/e) for GCβWADESAGPWG ( $\beta$  =  $\beta$ -alanine) peptide identified from a library sequencing analysis (purple color, top spectrum) and its authentic, separately re-synthesized counterpart (pink color, bottom spectrum).

## 5.4. Acknowledgements

This work was supported by the DARPA (Award No. 023504-001, B.L.P.). C.Z. is a recipient of the Bristol-Myers Squibb Fellowship in Synthetic Organic Chemistry. We would like to thank Mr. Mark D. Simon, Mr. Ethan D. Evans, Mr. Surin K. Mong and Mr. Alexander J. Mijalis for insightful discussions and helpful suggestions.

## 5.5. References

- (1) Goodwin, S.; McPherson, J.; McCombie, W. *Nat. Rev. Genet.* **2016**, *17*, 333–351.
- (2) Smith, G. *Science.* **1985**, *228*, 1315–1317.
- (3) Smith, G.; Petrenko, V. *Chem. Rev.* **1997**, *97*, 391–410.
- (4) Boder, E.; Wittrup, D. *Nat. Biotechnol.* **1997**, *15*, 553–557.
- (5) Wittrup, D.; Boder, E. *Methods Enzymol.* **2000**, *328*, 430–444.
- (6) Mattheakis, L.; Bhatt, R.; Dower, W. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9022–9026.
- (7) Wilson, D.; Keefe, A.; Szostak, J. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3750–3755.
- (8) Zuckermann, R.; Kodadek, T. *Curr. Opin. Mol. Ther.* **2009**, *11*, 299–307.
- (9) Fisher, B.; Gellman, S. *J. Am. Chem. Soc.* **2016**, *138*, 10766–10769.
- (10) Cheng, R.; Gellman, S.; DeGrado, W. *Chem. Rev.* **2001**, *101*, 3219–3232.
- (11) Fowler, S.; Blackwell, H. *Org. Biomol. Chem.* **2009**, *7*, 1508–1524.
- (12) Schlippe, Y.; Hartman, M.; Josephson, K.; Szostak, J. *J. Am. Chem. Soc.* **2012**, *134*, 10469–10477.
- (13) Hipolito, C.; Suga, H. *Curr. Opin. Chem. Biol.* **2012**, *16*, 196–203.
- (14) Needels, M.; Jones, D.; Tate, E.; Heinkel, G.; Kochersperger, L.; Dower, W.; Barrett, R.; Gallop, M. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 10700–10704.
- (15) Lam, K.; Lehman, A.; Song, A.; Doan, N.; Enstrom, A.; Maxwell, J.; Liu, R. *Methods Enzymol.* **2003**, *369*, 298–322.
- (16) Boeijen, A.; Liskamp, R. *Tetrahedron Lett.* **1998**, *39*, 3589–3592.
- (17) Joo, S.; Xiao, Q.; Ling, Y.; Gopishetty, B.; Pei, D. *J. Am. Chem. Soc.* **2006**, *128*, 13000–13009.
- (18) Thakkar, A.; Cohen, A.; Connolly, M.; Zuckermann, R.; Pei, D. *J. Comb. Chem.* **2009**, *11*, 294–302.
- (19) Semmler, A.; Weber, R.; Przybylski, M. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 215–219.
- (20) Lee, S.; Lim, J.; Tan, S.; Cha, J.; Yeo, S.; Agnew, H.; Heath, J. *Anal. Chem.* **2010**, *82*, 672–679.
- (21) Cha, J.; Lim, J.; Zheng, Y.; Tan, S.; Ang, Y.; Oon, J.; Ang, M.; Ling, J.; Bode, M.; Lee, S. *J. Lab. Autom.* **2012**, *17*, 186–200.
- (22) Thakkar, A.; Wavreille, A.; Pei, D. *Anal. Chem.* **2006**, *78*, 5935–5939.
- (23) Grant, G.; Crankshaw, M.; Gorka, J. *Methods Enzymol.* **1997**, *289*, 395–419.
- (24) Bathany, K.; Owens, N.; Guichard, G.; Schmitter, J. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 458–462.
- (25) Schreiber, J.; Quadroni, M.; Seebach, D. *Chimia (Aarau).* **1999**, *53*, 621–626.
- (26) Heerma, W.; Versluis, C.; Kostert, C.; Kruijtzter, J.; Zigrovic, I.; Liskamp, R. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 459–464.
- (27) Zhang, Y.; Fonslow, B.; Shan, B.; Baek, M.; Yates, J. *Chem. Rev.* **2013**, *113*, 233–2394.
- (28) Bantscheff, M.; Lemeer, S.; Savitski, M.; Kuster, B. *Anal. Bioanal. Chem.* **2012**, *404*, 939–965.
- (29) Altelaar, A.; Munoz, J.; Heck, A. *Nat. Rev. Genet.* **2012**, *14*, 35–48.
- (30) Mitchell, A.; Kent, S.; Engelhard, M.; Merrifield, R. *J. Org. Chem.* **1978**, *43*, 2845–2852.
- (31) Pulley, S.; Hegedus, L. *J. Am. Chem. Soc.* **1993**, *115*, 9037–9047.
- (32) Coin, I.; Beyermann, M.; Bienert, M. *Nat. Protoc.* **2007**, *2*, 3247–3256.

- (33) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342.
- (34) Ma, B.; Zhang, K.; Liang, C. *J. Comput. Syst. Sci.* **2005**, *70*, 418–430.
- (35) Rapp, W.; Fritz, H.; Bayer, E. *Proc. 12th Am. Pept. Symp.* **1991**, 529–530.
- (36) Bayer, E. *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 113–129.
- (37) Doran, T.; Gao, Y.; Mendes, K.; Dean, S.; Simanski, S.; Kodadek, T. *ACS Comb. Sci.* **2014**, *16*, 259–270.
- (38) Van der Walt, S.; Colbert, C.; Varoquaux, G. *Comput. Sci. Eng.* **2011**, *13*, 22-30.
- (39) McKinney, W. *Proceedings of the 9th Python in Science Conference.* **2010**, 51-56.
- (40) <http://www.scipy.org/>
- (41) Hunter, J. *Comput. Sci. Eng.* **2007**, *9*, 90-95.

## **Chapter 6. Discovery of Reactive Sequence Tags for C-Terminal Peptide and Protein Conjugation via Reactivity Screening of Intact Peptide Libraries**

The work presented in this chapter has been submitted for publication:

Vinogradov, A.; Loftis, A.; Pentelute, B. "Discovery of Reactive Sequence Tags for C-Terminal Peptide and Protein Conjugation via Reactivity Screening of Intact Peptide Libraries"

## 6.1. Introduction

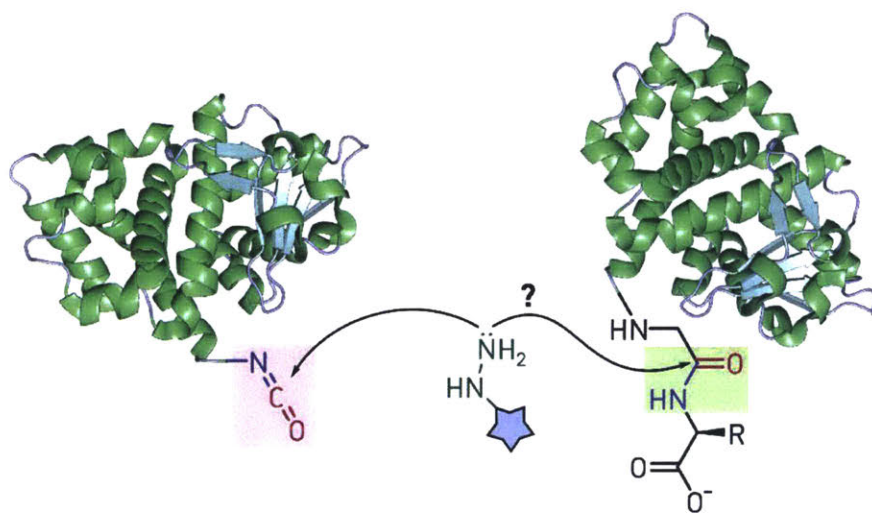
In this Chapter, we focus on the discovery of new methods for the C-terminal modification of peptides and proteins. The work presented here attempts to generalize the strategy established earlier in Chapter 3, where we developed an approach to the C-terminal conjugation of nucleophiles to peptide and protein isocyanates (Fig. 6.1). As powerful as isocyanate conjugation may be, the utility of the method is limited by the availability of precursor hydrazides, which is especially true for protein modification. Here, we expand the idea of conjugating various strong nucleophiles, such as hydrazine derivatives, to the C-terminus of fully proteogenic peptides and proteins, taking advantage of the inherently electrophilic nature of peptide bonds.

To discover peptide sequence tags which are able to undergo such a reaction, we turned to the methodology developed in Chapter 5, where we established a number of techniques for reliable high-throughput analysis of complex peptide mixtures using a combination of liquid chromatography and tandem mass spectrometry. In the work presented below, we further developed these techniques and demonstrated analysis of peptide libraries comprised of up to 13000 members. We utilized this methodology to perform reactivity screening of entire libraries, working towards the discovery of chemistry discussed above.

---

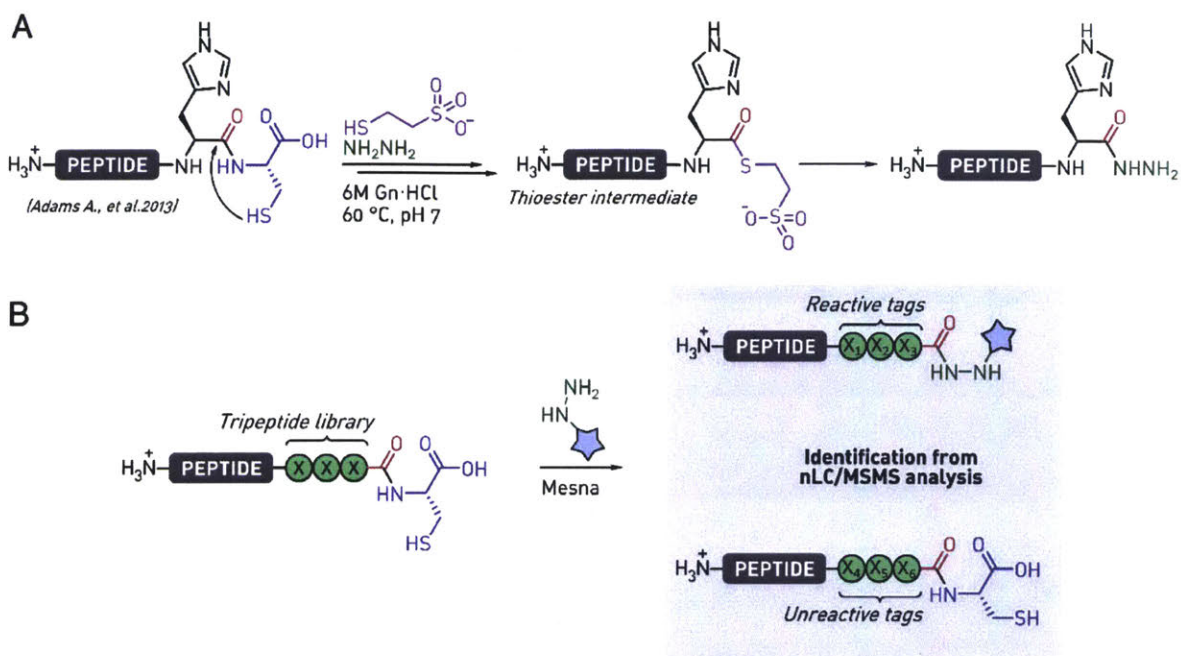
In 2013, Macmillan and coworkers described a method for the synthesis of C-terminal peptide and protein hydrazides from fully proteogenic precursors, which takes advantage of the unique reactivity of a cysteine residue.<sup>1</sup> The authors found that peptides and proteins containing C-terminal Gly-Cys and His-Cys dipeptide tags undergo selective, albeit fairly slow, hydrazinolysis in the presence of external thiol and hydrazine (Fig. 6.2A). Owing to the fact that described transformations are non-enzymatic and only require a proteogenic dipeptide tag, such a method has the potential to become a valuable tool for bio-conjugation<sup>2,3</sup> and protein sequence manipulation via a cascade of oxidation, thioesterification, and native chemical ligation reactions.<sup>4-6</sup> We hypothesized that since hydrazinolysis described above is at least to some extent sequence-dependent, a longer, more sophisticated peptide tag may provide further rate acceleration for the reaction.<sup>7,8</sup> Concomitantly, we thought that with the use of a more reactive sequence tag, the hydrazinolysis reaction may be generalized to a number of related less reactive nucleophiles, such as substituted hydrazines and hydrazides, allowing the direct conjugation of chemical probes to the C-terminus of fully proteogenic peptides and proteins via what can be called a transpeptidation reaction (Fig. 6.2B). The search of such a chemical tag is the focus of this study.





**Figure 6.1. The general idea of the study**

The protein render shown here is N-terminal domain of Anthrax toxin lethal factor (PDB: 1JKY).



**Figure 6.2. Site-selective hydrazinolysis of peptides and proteins**

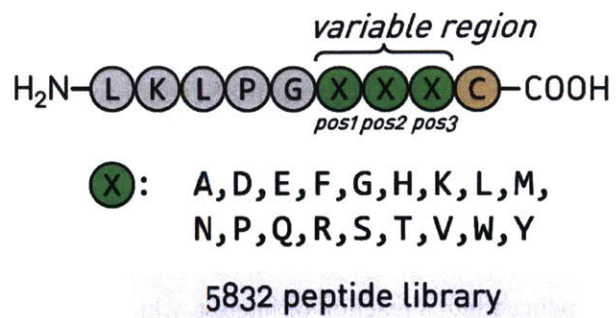
A) Known transformations of peptides bearing His-Cys and Gly-Cys motifs, and the proposed reaction pathway involving a Mesna thioester intermediate. B) Our strategy to discover peptide sequences reactive towards hydrazine derivatives.

## 6.2. Results and Discussion

### 6.2.1. Establishing Conditions for Library Reactivity Screening

To discover the optimal extended sequence, we turned to reactivity screening of intact libraries. Briefly, our idea can be summarized as follows. A small, redundant (3-4 amino acid residues-long, 5000-20000 members, redundancy > 10) one bead one compound peptide library is constructed and released into solution. The library is then introduced into a reaction of interest, which modifies peptides based on their reactivity. Without submitting the library to a selection step of any kind, the reaction mixture is then analyzed via nano-LC (nLC) coupled to tandem mass spectrometry (MS/MS), and the nature of reactive tags is recovered through *de novo* peptide sequencing. Such a strategy seemed appealing first of all because of its conceptual and experimental simplicity: the discovery process entails a single chemical reaction setup and a single nLC/MS/MS analysis, followed by data processing, which can be automated. Additionally, we reasoned that the discovery of reactions leading to peptide backbone cleavage may not be straightforward for other established screening approaches.<sup>9</sup> Intact library screening also removes any potential bias associated with a selection step. Finally, as we demonstrated in Chapter 5, nLC/MS/MS is capable of separating and analyzing mixtures containing thousands of synthetic library peptides, including those containing non-proteogenic functionalities. However, so long as nLC is utilized as the *only* method to discriminate between library members, the maximum library size is limited by the ability of nLC to successfully separate complex peptide mixtures. Our preliminary studies indicated that the upper bound for library size is around 25000 members, depending on the features of library design. This number may seem low in comparison to analogous procedures for the analysis of proteome peptides.<sup>10</sup> Our estimation was done for peptide libraries of a fixed length containing some sort of a constant region (fixed subsequence) in their design, which causes peptides to resemble each other in their chromatographical behavior. This fact decreases the chromatographical resolution of library members, and thus limits the throughput of the approach.

With these considerations in mind, we designed a first generation library, **library 1**, which incorporated a fixed C-terminal cysteine residue, three variable positions bearing 18 proteogenic amino acids each (5832 unique members; Ile and Cys were the only residues not allowed in the variable region), and an N-terminal pentapeptide (Leu-Lys-Leu-Pro-Gly) constant region to ensure desirable chromatographic behavior, appropriate solubility, and sequencing confidence for library members (Fig. 6.3). To ensure the synthesis of every library peptide and to increase the overall homogeneity of the library members, greater than 100-fold redundant library 1 was synthesized using established split-pool Fmoc SPPS techniques.<sup>11</sup> Next, we sought for optimal nLC/MS/MS conditions for the analysis of unreacted library. Running a 480 minute-long nLC gradient coupled to MS/MS (acquiring collision-induced dissociation (CID) and higher-energy collision-



**Figure 6.3. The design of library 1**

Three variable positions with 18 possible amino acids in each position account for 5832 theoretical library members.

nal dissociation (HCD) fragmentation spectra for every precursor ion) followed by *de novo* sequencing (PEAKS 8, BioInformatics Solutions Inc.)<sup>12,13</sup> and associated data analysis protocols resulted in the overall 84% sequence identification rate (4904 unique peptides). Detailed experimental procedures for the performed analysis are provided in section 6.3.4. In this experiment the library was analyzed at a rate of 730 peptides/hour, and we found that these results are consistent with the findings from Chapter 5, where we established that similar analysis for one-bead one-compound libraries can be performed at a rate of approximately 700 peptide/hour while maintaining ~85% sequence identification rate.

Tables 6.1 and 6.2 summarize the examination of the final peptide dataset, which revealed that peptides were identified throughout the gradient duration over a time span of 5 hours (from 66 to 373 min of the analysis time), indicating that the library was sufficiently separated for the MSMS analysis (on average, each member had ~3 s of allocated MSMS time). Additionally, the peptides generally had high sequencing scores (mean score:  $90 \pm 10$ ; one standard deviation; on a 0 to 99 scale) and low assignment errors ( $0.6 \pm 1.0$  ppm), suggesting that the *de novo* sequencing yielded high quality spectral assignments (Fig. 6.4). The overall mass distribution for identified peptides also had a Gaussian shape with high symmetry, indicating that the analysis did not bias the outcomes towards peptides of a certain mass or *m/z* values (Fig. 6.4). Taken together these results suggest that the established workflow for library analysis was adequate in its ability to decode the structure of library 1 peptides.

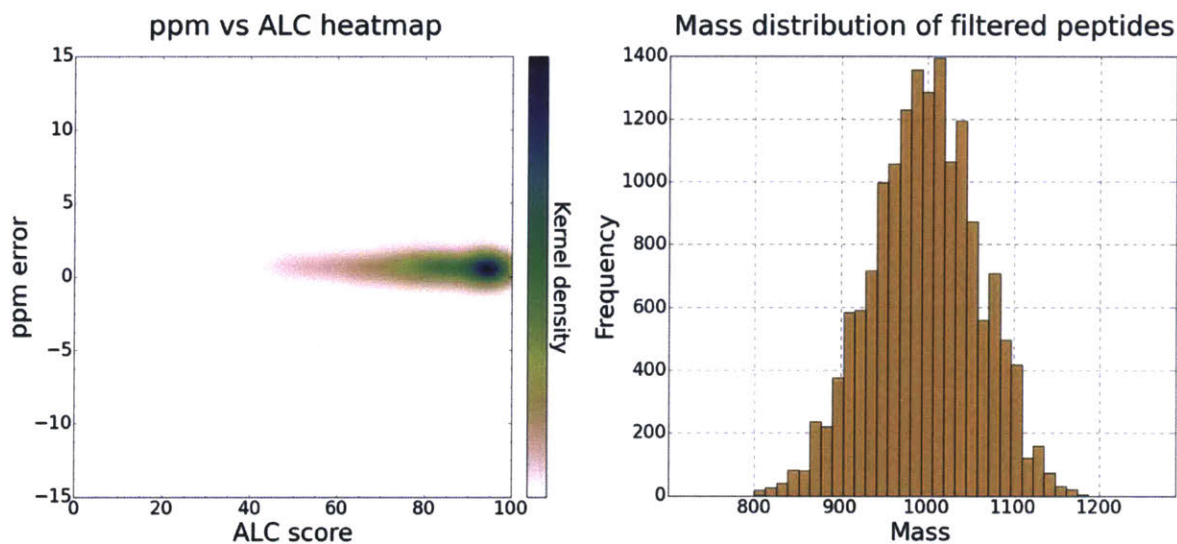
Finally, we established a nucleophile optimal from the decoding perspective, that is, a hydrazine derivative which, when reacted with a library peptide, does not impair the sequencability of the resulting product, and which has the lowest false discovery rate in the starting library background. To establish the false discovery rate (FDR) for several nucleophiles, a sample of naïve library 1 was analyzed as described above. PEAKS *de novo* sequencing was performed allowing either of the four C-terminal modifications: unsubstituted C-terminal hydrazide (+14.03 Da), N'-perfluoroarylhydrazide (+180.01 Da), N'-nicotinic hydrazide (+119.05 Da) or N'-adipic acid hydrazide (+156.10 Da). All sequences containing either of these modifications were postulated to be false positive. The FDR was estimated using two related metrics: a) the number of sequences containing the modification of interest in the raw PEAKS output b) number of sequences fully resembling reaction products identified from a data processing step performed as described above. The findings are summarized in Table 6.3. Most of the peptides assigned a false modification had poor sequencing scores and high assignment errors (Fig. 6.5), which allowed us to select the threshold values for the actual experiments accordingly. We found that, among tested hydrazine derivatives, perfluorophenylhydrazine (PFPh) yielded the lowest number of *a priori* erroneous assignments with PFPh-modified C-terminus (0.6% of the total dataset). In a series of independent preliminary experiments, we also found that PFPh-modified peptides could be sequenced as well as their unmodified analogues, and concluded that PFPh is a nucleophile choice for our purposes.

**Table 6.1. Summary of the raw PEAKS output for the sequencing analysis of naïve library 1**

	count	mean	std	min	25%	50%	75%	max	variance
ALC	27378	71.385	22.628	6	55	77	92	98	512.047
Length	27378	8.888	0.89	6	9	9	9	15	0.792
m/z	27378	452.13	85.59	237.47	379.21	469.28	516.27	750.49	7326.24
z	27378	2.271	0.486	2	2	2	3	4	0.236
RT	27378	200.22	72.41	14.51	141.14	196.03	254.69	406.71	5243.56
Area	25526	1.2E+07	1.7E+07	1.5E+00	8.7E+05	4.8E+06	1.8E+07	2.3E+08	2.9E+14
Mass	27378	990.66	102.23	700.39	927.56	998.50	1054.64	1498.95	10450.91
Ppm	27378	0.54	4.24	-15.1	0	0.6	1.1	15	17.972

**Table 6.2. Summary of the final (parsed) dataset for the sequencing analysis of naïve library 1**

	count	mean	std	min	25%	50%	75%	max	variance
ALC	4904	89.641	10.066	18	86	94	96	98	101.319
Area	4868	2.3E+07	2.0E+07	5.8E+00	9.7E+06	1.8E+07	3.1E+07	2.3E+08	4.2E+14
Mass	4904	990.42	55.302	800.421	954.478	989.544	1028.511	1187.6	3058.361
RT	4904	201.59	62.762	66.38	157.528	200.625	246.897	373.27	3939.093
m/z	4904	434.771	80.848	243.651	343.844	470.258	498.393	594.81	6536.4
ppm	4904	0.576	0.976	-7.9	0.2	0.6	0.9	7.8	0.953
z	4904	2.374	0.517	2	2	2	3	4	0.267
Length	4904	9	0	9	9	9	9	9	0

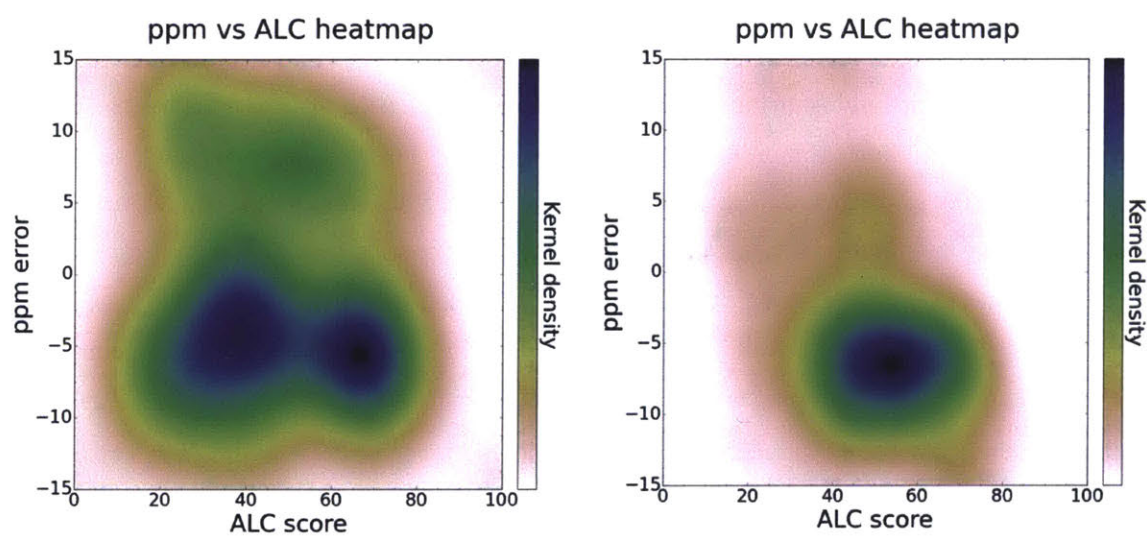


**Figure 6.4. Analysis of the naïve library 1 dataset**

On the left: Sequencing quality heat map for the final filtered dataset (library 1 analysis). On the right: Peptide molecular weight distribution for the final filtered dataset (library 1 analysis).

**Table 6.3. Summary of the FDR experiment**

Conjugation type	Total count	Fraction of the dataset	Filtered count	Fraction of the dataset
Hydrazine	1983	0.10	254	0.013
PFPh	495	0.02	113	0.006
Adipic dihydrazide	627	0.03	133	0.007
Isoniazid	1228	0.06	392	0.020



**Figure 6.5. Sequencing quality heat maps for the datasets containing erroneous assignments obtained from a naïve library analysis**

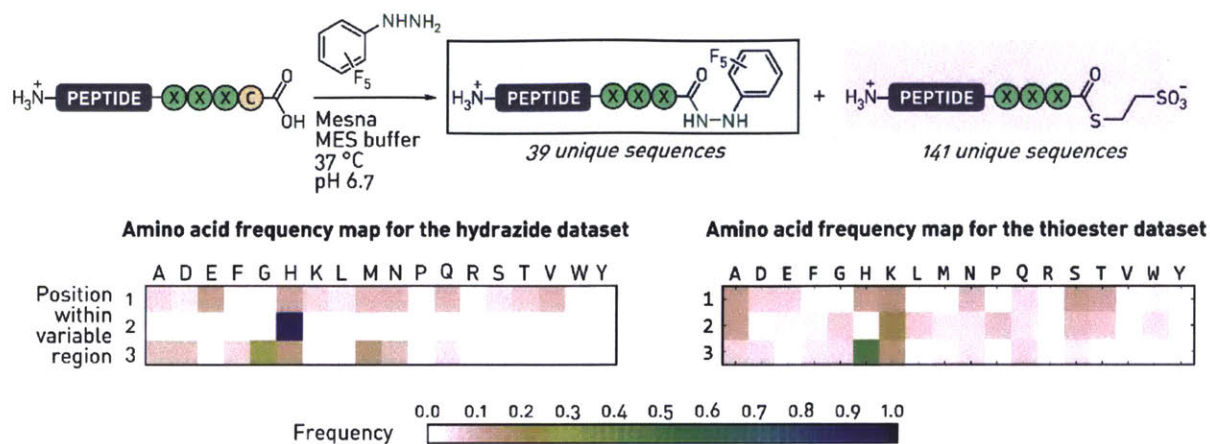
Left: PFPh data set, right: hydrazine data set. High assignments errors and low sequencing scores of these datasets indicate poor sequencing quality of these mock peptides.

## 6.2.2. Library 1 Screen

With the library and an appropriate probe in hand, we turned to the reaction discovery process. We sought to perform the reaction under stringent conditions where the most (>99%) library members do not yield any detectable product. As such, 2 mg/ml (~350 ng/ml for each peptide) library solution in 50 mM MES buffer at pH 6.7 was reacted with 5 mg/ml PFPh and 20 mg/ml sodium 2-mercaptoethanesulfonate (Mesna) at 37 °C for 24 hours, and then analyzed by nLC/MSMS as described above. *De novo* sequence assignment was performed allowing any of the three non-proteogenic modifications for every assignment (C-terminal PFPh conjugate, Mesna disulfide with Cys, and C-terminal Mesna thioester). The follow-up analysis was performed on those peptides assigned as either C-terminal PFPh conjugates (expected reaction product) or C-terminal Mesna thioesters (expected reaction intermediate), which had sequencing scores greater than 80, to minimize analysis of the data obtained from incorrect assignments. Integrated extracted ion currents corresponding to precursor ions from which reaction products were identified were directly used as a proxy for the reactivity estimation, under a tacit assumption that the ionization efficiency and the molar fraction within the library is more or less even for all library members. In other words, traditional reaction yields were not calculated. The decision to analyze peptide reactivities in this way was primarily driven by the fact that a starting material-product pair could not be identified for a sufficiently large portion of identified hits. In addition, direct quantification empirically proved to be more representative of observed reactivities. A list of most reactive library members obtained in this way was further manually compared to the original chromatogram to account for the errors associated with automated peak area integration.

Analyzed in this way, library 1 screen yielded 39 unique PFPh-conjugate sequences and 141 unique Mesna thioesters (full sequence list are provided in sections 6.3.8 and 6.3.9). Analysis of amino frequencies in the PFPh dataset revealed (Fig. 6.6) a strong consensus for one amino acid position: 90% sequences contained a His residue in the second variable region position. Position 3, i.e. amino acid residue adjacent to Cys, had a weaker consensus with the preference for Gly (38% peptides), which agrees with published observations.<sup>1</sup> Position 1, most distal from Cys, had essentially no amino acid convergence. The Mesna dataset differed significantly from the PFPh one, as no unambiguous pattern was observed, although positively charged residues (Lys and His) were overrepresented. Based on these results and analyses of relative reactivity as described above, we re-synthesized 13 peptides — 7 from the PFPh dataset and 6 from the Mesna one — to validate their reactivity in the batch format.





**Figure 6.6. Library 1 reaction screening**

The screen yielded 39 peptides identified as substituted C-terminal *N*'-(perfluorophenyl)hydrazides and 141 C-terminal thioesters. Color-coded positional amino acid frequency maps for these datasets are displayed on the bottom. Reaction conditions are specified in the text.

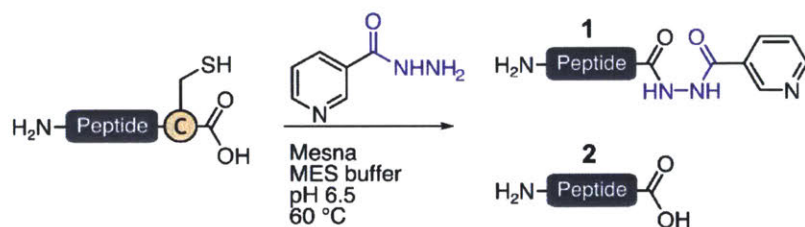
For validation, the peptides were introduced into the reaction with Mesna and nicotinic hydrazide. During our initial validation efforts, we found that nicotinic hydrazide is more representative of general peptide reactivities than PFPh, because reactions with the latter were, for many peptides, too slow to discriminate between their chemical reactivities. In most cases, final reaction mixtures consisted of the conjugation product, the product of the formal thioester hydrolysis and unreactive starting material. Table 6.4 summarizes HPLC yields of *N'*-substituted hydrazides and hydrolysis products for resynthesized peptides.

Interestingly, Mesna thioesters, which we assumed would be the major reaction products for the six sequences identified from the Mesna dataset (entries **1-6**),<sup>14</sup> were either not detected or comprised less than 1% of the crude reaction mixture in all cases, even when no nucleophile was added. At the same time, we found that these peptides yielded less than 30% of the expected *N'*-substituted C-terminal hydrazide, and less than 10% of the hydrolysis side-product, with the exception of peptide **6**, which reacted with nicotinic hydrazide giving 88% of the expected product. Comparing these outcomes against the reactivity of an arbitrary sequence containing a C-terminal cysteine residue (entry **20**), allowed us to conclude that thioester dataset-derived peptides **1-5** were false positive hits. Such low reactivity of Mesna dataset peptides is somewhat puzzling for two reasons. First, the precursor ions used to identify the Mesna thioester sequences during the screen for the most part were more than two to five-fold more intense than corresponding PFPh-derived ions, which led us to believe that these peptides ought to be reactive. Second, fragmentation spectra of most of these ions contained a prominent and characteristic  $[M-141.98]^+$  Da/e peak, suggesting that the corresponding peptides were indeed C-terminal Mesna thioesters, and were decoded correctly. These results could be explained if Mesna thioesters had ionization efficiency that is massively superior to that of unmodified peptides and C-terminal substituted hydrazides, but in this work we did not test this hypothesis.

In contrast to the results obtained for the thioesters, peptides from the PFPh dataset (entries **7-13**) were reactive towards nicotinic hydrazide. All peptides demonstrated at least 40% conversion, and hits **7** and **11** yielded more than 80% of the desired product. In general, peptides bearing a His residue in the second variable position were the most reactive ones, and additionally the peptide with a His-Gly-Cys tag (entry **11**) had higher conjugation to hydrolysis ratio than His-His-Cys peptides. Peptide **11** was both more reactive and selective (in terms of product to hydrolysis ratio) than an arbitrary sequence containing a previously reported His-Cys-COOH motif (entry **19**).

**Table 6.4. Reactivity of authentic hit peptides towards nicotinic hydrazide**

Peptide solutions (2 mg/mL) in 50 mM MES buffer at pH 6.7 were reacted with 50 mg/mL nicotinic hydrazide and 100 mg/ml Mesna at 60 °C for 48 hours and analyzed by HPLC/MS.



#	Sequence	ID from	HPLC yield, %	
			1	2
1 <sup>a</sup>	LKLPGKNHC	library 1	4 <sup>b</sup>	4
2 <sup>a</sup>	LKLPGHKQC		9	9
3 <sup>a</sup>	LKLPGKKSC		7	5
4 <sup>a</sup>	LKLPGSKKC		6	9
5 <sup>a</sup>	LKLPGKKGC		26	4
6 <sup>a</sup>	LKLPGKHHC		88	8
7	LKLPGHHHC		84	11
8	LKLPGEHDC		68 <sup>c</sup>	6
9	LKLPGWWGC		40	1
10	LKLPGLHFC		71	8
11	LKLPGHHGC		86	3
12	LKLPGHGHC		30	6
13	LKLPGNHHC		17 <sup>b</sup>	2
14	TLLGHTKHGC	library 2	90	3
15	TLLGKPTHGC		88	5
16	TLLGGLKHHC		84	7
17	TLLGHTGHHC		82	10
18	TLLGHGLHHC		83	8
19	TKELYSHC	P. C. <sup>d</sup>	15	12
20	TLLGPDGTKC	N. C. <sup>e</sup>	12	7

<sup>a</sup>) Identified, i.e. *de novo* sequenced, as Mesna thioesters from the screen; <sup>b</sup>) Main product was 1 Da heavier than expected, which was attributed to N→D deamidation under the reaction conditions; <sup>c</sup>) The product was observed as two chromatographically resolved peaks of equal intensity <sup>d</sup>) P. C. = positive control: arbitrary sequence with a known His-Cys motif <sup>e</sup>) N. C. = negative control: arbitrary sequence with a C-terminal Cys residue.

### 6.2.3. Library 2 Screen

Building on the results from the library 1 screen, we generated a larger second-generation library, **library 2**, which featured a fixed His one residue away from Cys, a four residue variable region accounting for 13310 theoretical members, and a different sequence (Thr-Leu-Leu-Gly) for the constant region (Fig. 6.8A). This new constant region was designed to account for the fixed His residue, which introduces an extra positive charge for every library member, and thus, affects the solubility, chromatographical behavior, and sequencability of library members. With this library we wanted to investigate the possibility of finding longer (up to 6 residues) reactive C-terminal tags, and the feasibility of analyzing peptide mixtures comprised of more than 10000 members at once using our approach. Sequencing of naïve library 2 resulted in 11070 unique peptides (83% identification rate), with secondary statistics resembling corresponding results for library 1 (Tables 6.5 and 6.6; Fig. 6.7). Thus, the identified peptides had an average sequencing score of  $82 \pm 10$ , assignments errors of  $-0.9 \pm 0.9$  ppm, symmetrical mass distribution, and were identified throughout a ~6.5-hour time window. We concluded that nLC/MSMS analysis of libraries comprised of more than 10000 members is practical, and turned to the reaction discovery using library 2.

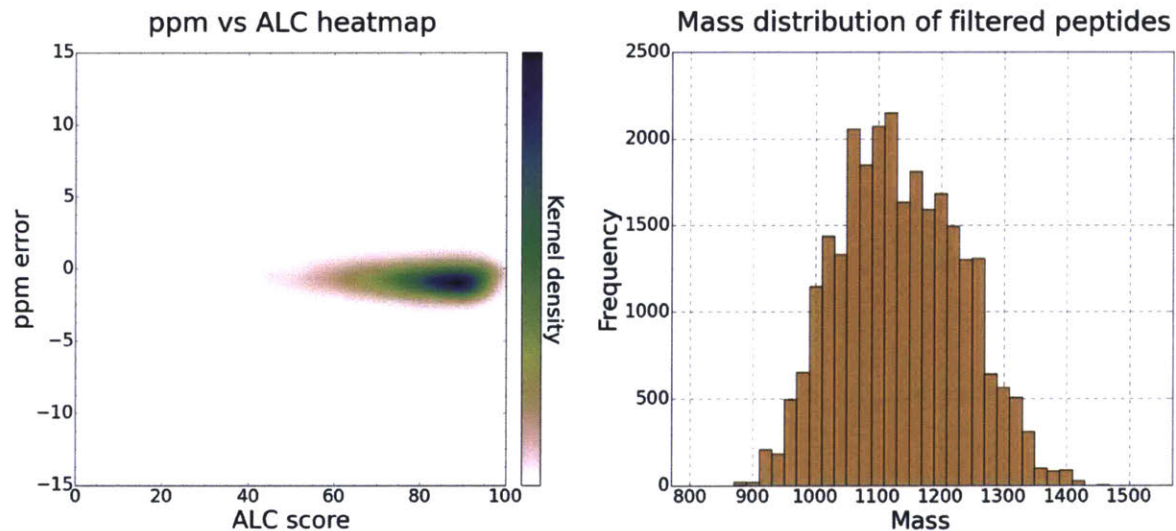
The reaction was performed similarly to the first screen, accounting for the fact that on average library 2 peptides were more reactive than library 1 ones due to the presence of the key His residue for all library 2 members. As before, our goal was to perform the reaction under stringent conditions, such that ~99% of the library members do not yield any detectable product. After some preliminary experimentation, the screen was performed under conditions similar to those described above for library 1 reaction, except the concentrations of individual peptides and Mesna were lower (150 ng/ml and 15 mg/ml, respectively), and the reaction time was reduced to 18 hours. nLC/MSMS conditions and the downstream analysis procedures were identical to the first screen, except Mesna thioesters were not analyzed (the thioester modification, however, was still allowed during the *de novo* assignment step). After the analysis, the final peptide list contained 203 unique sequences (full list is provided in section 6.3.10), which weakly converged on His/Gly in position 5, and had no consensus elsewhere, reinforcing the results obtained from library 1 screen (Fig. 6.8B). We re-synthesized five hit sequences with highest verified peak area (i.e., highest assumed PFPh-conjugation yields), and analyzed their reactivity in the reaction with nicotinic hydrazide identically to hits from library 1 (Table 6.4, entries **14-18**). All peptides reacted with the hydrazide in the presence of Mesna yielding at least 80% of the conjugation product, and His-Gly-Cys containing peptides (entries **14, 15**) demonstrated lower hydrolysis rates than His-His-Cys peptides (entries **16-19**).

**Table 6.5. Summary of the raw PEAKS output for the sequencing analysis of naïve library 2**

	count	mean	std	min	25%	50%	75%	max	variance
ALC	45688	71.5	21.2	4	61	79	87	99	448.4
Length	45688	9.8	1.0	4	9	10	10	15	0.9
m/z	45688	491.3	101.8	245.5	403.9	499.8	576.2	747.4	10362.5
z	45688	2.3	0.5	2	2	2	3	4	0.3
RT	45688	182.4	73.1	15.8	121.1	178.8	236.4	457.3	5339.3
Area	43644	9.0E+06	1.9E+07	1.0E+00	7.1E+05	2.5E+06	8.3E+06	3.8E+08	3.5E+14
Mass	45688	1098.1	129.4	700.4	1006.5	1109.5	1194.6	1492.8	16756.9
ppm	45688	-0.78	3.625	-15.2	-1.6	-1	-0.3	15	13.1

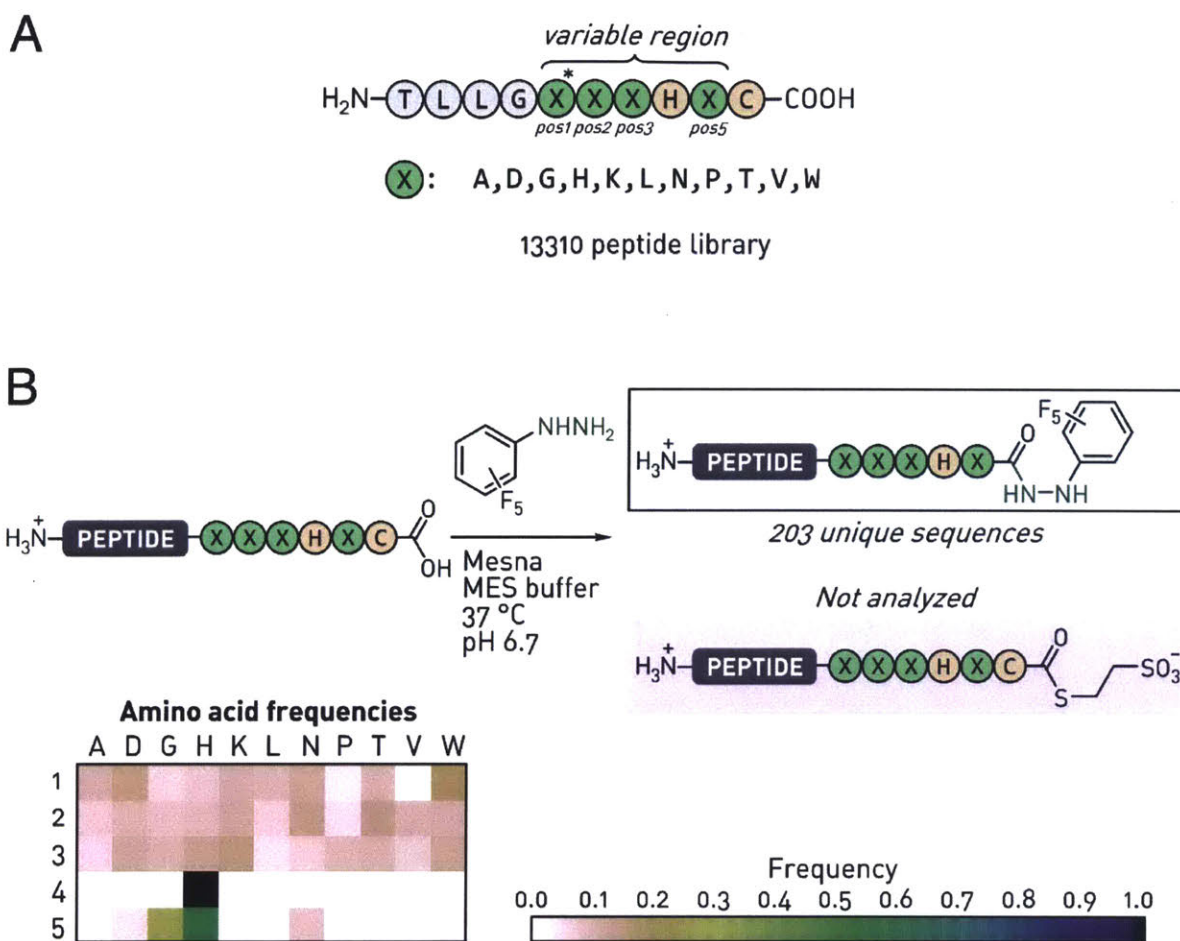
**Table 6.6. Summary of the final (parsed) dataset for the sequencing analysis of naïve library 2**

	count	mean	std	min	25%	50%	75%	max	variance
ALC	11070	82.4	10.5	20	78	85	90	99	109.8
Area	10963	1.5E+07	2.6E+07	1.2E+00	1.5E+06	5.4E+06	1.7E+07	3.8E+08	6.6E+14
Mass	11070	1117.2	87.0	870.4	1053.6	1108.5	1178.5	1428.6	7573.0
RT	11070	195.0	75.5	37.9	133.7	191.8	247.0	457.3	5698.7
m/z	11070	499.0	104.3	254.1	389.5	529.8	578.3	715.3	10878.8
ppm	11070	-0.9	0.9	-7.9	-1.4	-0.9	-0.4	7.9	0.8
z	11070	2.4	0.6	2	2	2	3	4	0.329
Length	11070	10	0	10	10	10	10	10	0



**Figure 6.7. Analysis of the naïve library 2 dataset**

On the left: Sequencing quality heat map for the final filtered dataset (library 2 analysis). On the right: Peptide molecular weight distribution for the final filtered dataset (library 2 analysis).



**Figure 6.8. Design and reactivity screening of library 2**

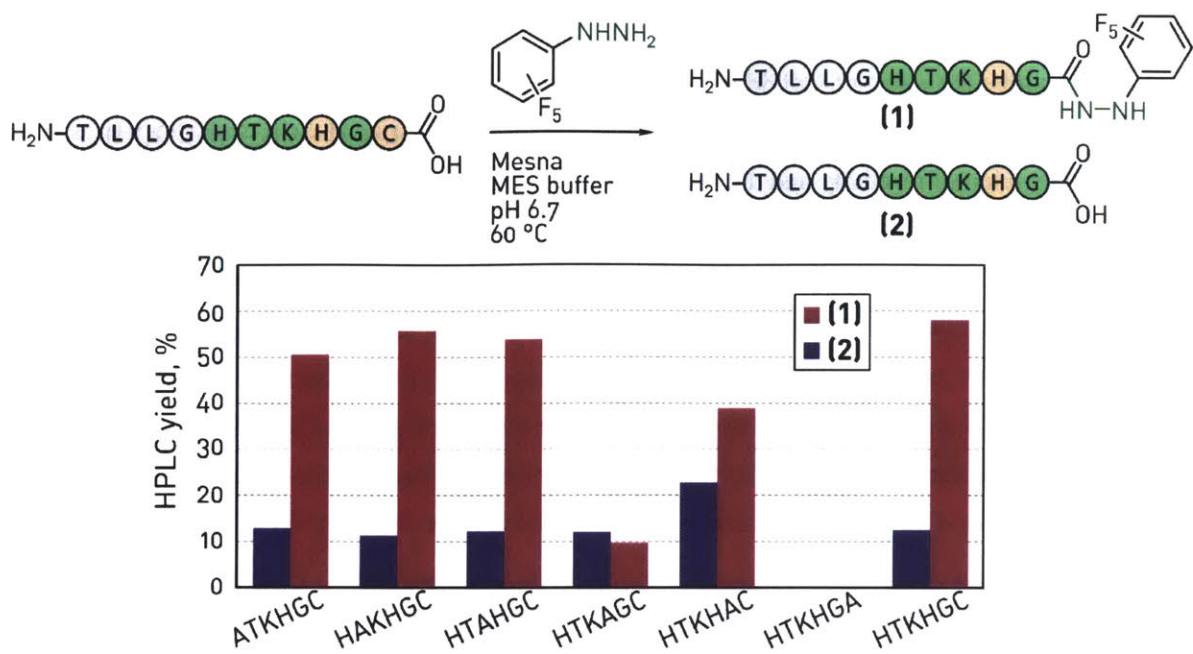
A) The design of library 2. Position 2, 3, and 5 each had 11 possible monomer amino acids, and position 1 had 10 (-Val relative to the rest), accounting for the total of 13310 peptides. B) Summary of the screen outcomes. The reaction on the entire library 2 yielded 203 peptides identified as substituted C-terminal *N*'-perfluorophenylhydrazides. Mesna thioesters were also identified, but not analyzed. Color-coded positional amino acid frequency maps for the *N*'-perfluorophenylhydrazide dataset is displayed on the bottom. Reaction conditions are specified in the text.

#### 6.2.4. Investigation of the Reaction Scope

To further probe the effect of individual amino acids on the overall reactivity of discovered tags, we performed alanine scan on peptide **14** (H<sub>2</sub>N-Thr-Leu-Leu-Gly-His-Thr-Lys-His-Gly-Cys-COOH), which demonstrated the highest reactivity and the lowest hydrolysis rate amongst the discovered sequences. Six alanine mutants and the parent peptide **14** (all at 2 mg/mL) were reacted with 7 mg/mL PFPh and 100 mg/mL Mesna in MES buffer at pH 6.7 and 60 °C for 48 hours. The reactions were analyzed by HPLC/MS, and compared to each other. As Fig. 6.9 illustrates, we found that amino acids in positions 1, 2 and 3 had subtle effects on the outcome of the conjugation, lowering product yields by no more than 10%, whereas His in position 4 and Cys were critical for the reactivity of the peptide. Mutating Gly in position 5 decreased the product yield, but more importantly doubled the amount of the hydrolysis product. Taken together, these results suggest His-Gly-Cys motif is the primary factor responsible for the reactivity of the peptide.

Next, we investigated the nucleophile scope of the reaction in more detail. To this end, we reacted 2 mg/ml solution of peptide **14** with different hydrazines and hydrazides (exact concentrations for different nucleophiles and detailed experimental procedures are in section 6.3.6) in the presence of 100 mg/ml Mesna in MES buffer at pH 6.7 and 60 °C for 48 hours, and analyzed the reaction outcomes by HPLC/MS. We found (Fig. 6.10) that various hydrazine derivatives and hydrazine itself conjugate to peptide **14** yielding more than 85% of the product in most cases. Reactions with hydrazine or succinic dihydrazide proceeded essentially quantitatively (98% product yield in both cases), whereas the conjugation with PFPh, which resulted in 61% product, was the only notable exception to the general reactivity displayed by the studied hydrazine derivatives. In general the conjugation proceeded with little hydrolysis (<5%) as long as the nucleophiles were soluble in water to required concentrations, which were on the order of ~100 mM depending on the probe. Interestingly, the reaction between peptide **14** and Mesna alone resulted in 34% hydrolysis product with no observable thioester, which suggests that the reaction mechanism is more complicated than a simple Mesna thioester formation followed by its reaction with a hydrazine derivative. At the same time, we observed that the reaction does not proceed in the absence of Mesna. The mechanism of uncovered transformations certainly needs thorough investigation.

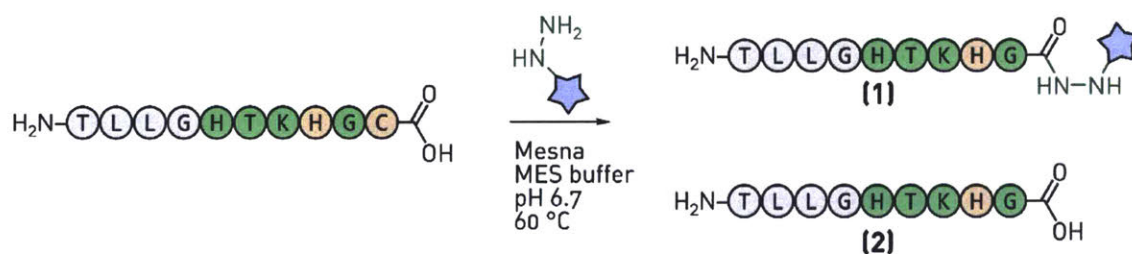
Finally, we explored the generality of the reactivity displayed by the discovered tag in different sequence contexts. Accordingly, we prepared a 32-kDa protein (LF<sub>N</sub>, N-terminal domain of Anthrax toxin lethal factor)<sup>15</sup> bearing His-Thr-Lys-His-Gly-Cys-COOH on the C-terminus, and reacted it with 50 mg/ml adipic acid dihydrazide in the presence of Mesna in MES buffer at pH 6.7 at 37 °C. The reaction proceeded significantly slower, but nevertheless after 96 hours the protein was smoothly converted (89% yield as estimated from the deconvolution results) to a corresponding product bearing a hydrazide moiety on a flexible linker (Fig. 6.11).

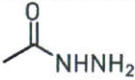
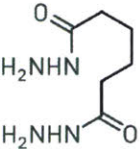
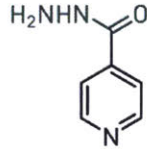
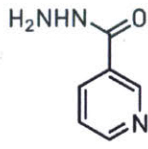
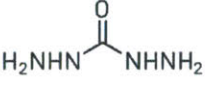
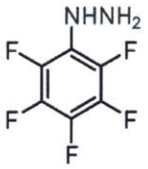
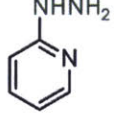
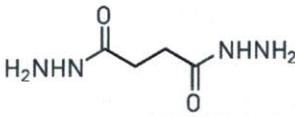


**Figure 6.9. Results of peptide 14 alanine scan experiments**

Reaction conditions are provided in the text. All alanine mutants also contained a TLLG constant region on the N-terminus.

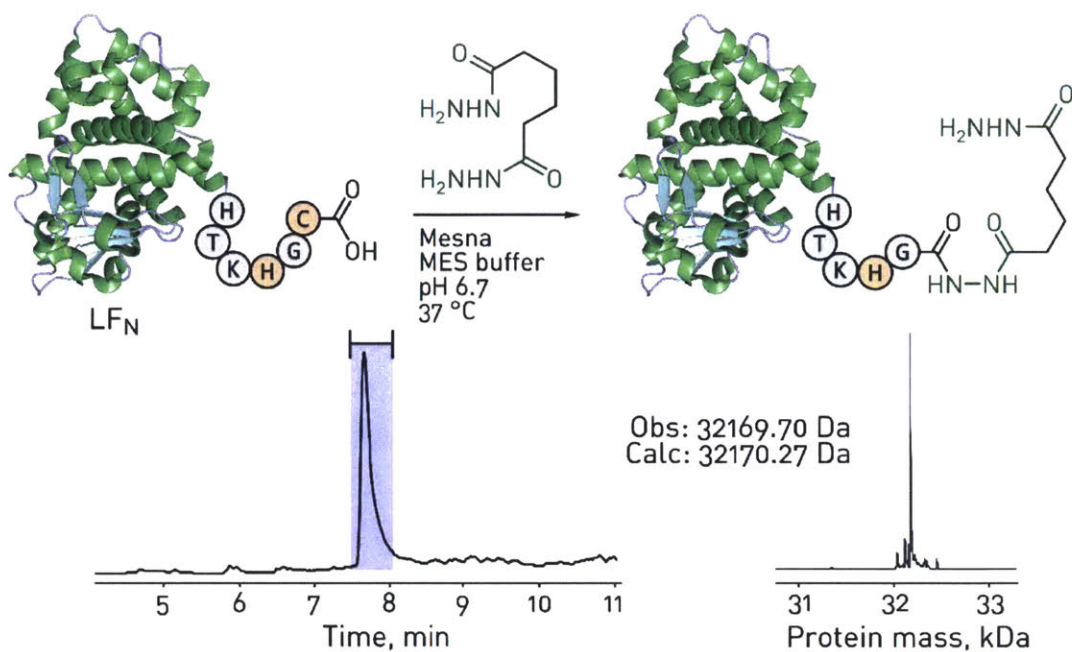




None	(1): 0% <sup>a</sup> (2): 34%	$\text{NH}_2\text{NH}_2$	(1): 98% (2): <1%		(1): 91% (2): 2%
	(1): 91% (2): 2%		(1): 88% (2): 3%		(1): 90% (2): 3%
	(1): 94% (2): 1%		(1): 61% (2): 11%		(1): 74% <sup>b</sup> (2): 1%
			(1): 98% <sup>c</sup> (2): 1%		

**Figure 6.10. Analysis of the reactivity exhibited by the discovered sequence tag**

Analytical yields of the conjugation and hydrolysis products in the reaction between peptide 14 and various hydrazine derivatives <sup>a)</sup> In this case, Mesna thioester was assumed to be product 1. <sup>b)</sup> An intramolecular cyclization product (-18 Da) was the major side product. <sup>c)</sup> The final reaction product was unsubstituted C-terminal peptide hydrazide.



**Figure 6.11. Synthesis of a C-terminal protein hydrazide using the discovered chemistry**  
 $\text{LF}_N\text{-HTKHGC-COOH}$  ( $3 \mu\text{M}$ ) was incubated with 100 mg/ml Mesna and 50 mg/ml adipic dihydrazide in 50 mM MES buffer at pH 6.7 for 96 hours at  $37^\circ\text{C}$ , and analyzed by HPLC/MS. TIC chromatogram and associated mass spectrum deconvolution (maximum entropy) result are shown on the bottom.

This experiment confirmed that the general reactivity demonstrated by C-terminal His-Gly-Cys tag is translatable from model 10-mer peptides to intact proteins.

### 6.2.5. Conclusion

In summary, in this study, we investigated the feasibility of selection-free peptide library screening for the discovery of sequence tags with unique reactivity. The approach relies on the combination of nLC and tandem mass spectrometry to directly analyze outcomes of chemical reactions performed on the entire libraries. We established that the analysis of reactions performed simultaneously on 13000 peptides is within the reach for the described method. We anticipate that constructing libraries bearing no or minimal constant region, and thereby increasing the diversity of peptides in terms of their chromatographical behavior, will allow routine analysis of libraries comprised of 20000 members and more.

Methods described here can be utilized to perform reaction discovery experiments of various kinds, although, in principle, the described nLC/MSMS techniques could also be extended to other applications. Reactions leading to the modification of amino acid side chains, backbone cleavage, different conjugations, and other types of transformations can be explored in this format by taking advantage of the unique ability of tandem mass spectrometry to decode the structures of peptides and peptidomimetics of a great variety. Conversely, the fact that fragmentation spectra are used as the only way to gain insight into the outcomes of reactions imposes a number of limitations to the types of modifications that can be introduced and analyzed using this approach. Foremost, reaction products of interest need to undergo facile fragmentation under CID and HCD conditions, and on top of that, these products need to have a reasonably low false discovery rate. Both of these considerations are intimately tied to the nature of a probe with which a peptide is reacted. In our experience, the types of peptide modifications that can easily be identified in a high-throughput format are those that are approximately 40-200 Da in size, have low basicity in gas phase and low propensity for off-pathway fragmentation (by which we understand any fragmentation that does not lead to the generation of *b* and *y* ions or derivatives of thereof) under CID/HCD conditions. In the work presented here, the use of perfluorophenylhydrazine as a reactive probe satisfied all of these criteria. C-terminal peptide *N'*-perfluorophenylhydrazides had low false discovery rate in the native library background, and could be successfully *de novo* sequenced.

To examine the feasibility of these considerations, we performed reactions to discover peptide sequence tags reactive towards substituted hydrazines and hydrazides in the presence of external thiol for two separate libraries. We found a number of reactive peptides pointing towards His-Gly-Cys as a reactive C-terminal tripeptide. We confirmed the reactivity of this tag in a traditional batch format, verified the importance of all three amino acids by preparing and studying single alanine mutants, and established that various hydrazine derivatives can be conjugated to peptides and proteins bearing this tripeptide in high

yields (>85%). This reaction represents a logical extension of the C-terminal isocyanate conjugation chemistry established in Chapter 3 to the case of fully proteogenic sequences, although mechanistically these reactions are quite different. Admittedly, site-selective hydrazinolysis of the discovered sequence tag is still fairly slow, and requires high concentrations of external thiol and hydrazine or hydrazide nucleophile. Further elaboration of the tripeptide motif and the discovery of a mutually selective peptide/hydrazide reactive pair may allow non-incremental improvement of the reaction rate and selectivity, increasing the practical utility of the uncovered transformations.

## 6.3. Experimental

### 6.3.1. General

2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-(7-aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU), and 2-chlorotrityl chloride resin (200-400mesh, 1.2 mmol/g) were purchased from Chem-Impex International. N<sup>α</sup>-Fmoc protected L-amino acids were obtained through Advanced ChemTech, and Novabiochem. Boc-Gly-OH and Boc-Ile-OH were from Midwest Bio-Tech.

*N,N*-Dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile were from EMD Millipore. Triisopropyl silane (TIPS), acetic hydrazide, and 1,2-ethanedithiol (EDT) were from Alfa Aesar. Solvents for HPLC-MS were purchased from Fluka. All other reagents were purchased from Sigma-Aldrich.

### 6.3.2. Library and Peptide Synthesis

*Library 1 synthesis.* 2-Chlorotrityl chloride resin (400 mg, 200-400 mesh, 1.2 mmol/g) was swollen in dry DCM for 10 min. Then, 2 mmol of Fmoc-Cys(Trt)-OH and 600  $\mu$ L DIEA dissolved in 5 mL of dry DCM was added to the resin. The suspension was swirled and left for 20 min to install the first amino acid, after which the solution was drained. The resin was washed 3X with DCM and methanol, and incubated with 5 mL of methanol for 10 min. After washing the resin 3X with DCM and DMF, regular split-pool Fmoc peptide synthesis ensued. N<sup>α</sup>-Fmoc protecting group removal was achieved by two 3 min treatments with 5 mL of 20% (v/v) piperidine in DMF. For the amino acid coupling step, 0.5 mmol Fmoc-protected amino acids (any L-proteogenic amino acid except Cys and Ile) were dissolved in 1.25 mL volume of 0.38 M HATU in DMF, and activated with 150  $\mu$ L of DIEA. Couplings were allowed to proceed for 10 to 30 min at room temperature with periodical gentle rocking of the resin suspension. Elongation of the constant region was performed much in the same way, except 2 mmol Fmoc-protected amino acids were dissolved in 5 mL volume of 0.38 M HATU in DMF, and activated with 600  $\mu$ L of DIEA was used for couplings. Side chain protection was as follows: Asn(Trt), Asp(OtBu), Cys(Trt), Gln(Trt), Glu(OtBu), His(Boc), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu). Fully assembled library was cleaved with 2.5% (v/v) EDT, 2.5% (v/v) H<sub>2</sub>O, and 1% (v/v) TIPS in TFA for 2 hours at room temperature, after which peptides were precipitated with cold (-80°C) diethyl ether. The resulting mixtures were washed three times with diethyl ether, dissolved in 25% water / 75% acetonitrile (v/v, 0.1% TFA) and lyophilized. Prior to an nLC/MSMS analysis a 20-30 mg library aliquot was additionally desalted on an SPE cartridge (C18, Alltech) and lyophilized.

*Library 2 synthesis.* Library 2 was synthesized analogously to library 1, except different amino acids were used for split pool. The library design was Thr-Leu-Leu-Gly-Xaa<sub>1</sub>-Xaa<sub>2</sub>-Xaa<sub>2</sub>-His-Xaa<sub>2</sub>-Cys-COOH where Xaa<sub>1</sub> is Ala, Asp, Gly, His, Lys, Leu, Asn, Pro, Thr, or Trp, and Xaa<sub>2</sub> is Ala, Asp, Gly, His, Lys, Leu, Asn, Pro, Thr, Val, or Trp (+Val w.r.t. Xaa<sub>1</sub>).

*Peptide synthesis.* Individual peptides were synthesized using batch Fmoc SPPS techniques on 100-200 mg of 2-chlorotrityl chloride resin (200-400 mesh, 1.2 mmol/g). Peptide synthesis and cleavage conditions were as described above. Peptides were purified on a Waters 600 HPLC system with a Waters 484 or 486 UV detector using water with 0.1% TFA added (solvent A) and acetonitrile with 0.1% TFA added (solvent B) as solvents. An Agilent Zorbax 300SB semi-preparative C3 column (300 Å, 5 µm, 9.4 x 250 mm) or an analogous C18 column were used at 5 mL/min linear flow rate for the purification of 30-35 mg of peptide crude. Unless noted, the following gradient was used for purification: 1% B in A for 5 minutes, then 1%-31% B ramping linearly over 90 minutes. Fraction containing pure material (as confirmed by HPLC/MS) were lyophilized. Alternatively, for crude peptides of high quality, solid phase extraction of 15-20 mg of material was performed on an SPE cartridge (C18, Alltech) and the resulting peptide was lyophilized.

### 6.3.3. Preparation of LF<sub>N</sub>-HKTHGC-COOH

*Protein expression and purification.* SUMO-LF<sub>N</sub>-LPSTGG-H<sub>6</sub> was expressed in BL21(DE3) *E. coli* cells. The cells were grown at 37 °C to an OD<sub>600</sub> ~0.6-1.0. The proteins were induced with 0.4 mM IPTG overnight at 30 °C. After induction, the cells were pelleted and resuspended in 20 mM Tris pH 7.5 and 150 mM NaCl with protease inhibitor cocktail (Roche), DNaseI (Roche), and lysozyme. The cells were sonicated five times for 20 seconds on ice then centrifuged for 30 minutes at 35,000 g and 4 °C. The lysate was purified using a HisTrap FF NiNTA column (GE Healthcare) pre-equilibrated with 20 mM Tris pH 8.5 and 150 mM NaCl. The protein was loaded onto the column then washed with the equilibration buffer, followed by 20 mM Tris pH 8.5, 500 mM NaCl, and 40 mM imidazole, then eluted in buffer containing 20 mM Tris pH 8.5, 500 mM NaCl, and 500 mM imidazole. The protein elution was further purified by SEC using a HiLoad 26/600 Superdex column (GE Healthcare) into 20 mM Tris pH 7.5 and 150 mM NaCl.

*Sortase-mediated ligation purification.* *Staphylococcus aureus*<sup>59-206</sup>SrtA (P94S/D160N/K196T; SrtA\*) was used to ligate SUMO-LF<sub>N</sub>-LPSTGG-H<sub>6</sub> to GGGTLLGHTKHGC. In order to obtain a native N-terminus, the small ubiquitin-like modified (SUMO) was cleaved off LF<sub>N</sub>-LPSTGG-H<sub>6</sub>. SUMO cleavage was achieved using 1 µg SUMO protease per mg of protein substrate at RT for 2 hours. Sortase-mediated ligations were done in the presence of Ni-NTA beads to sequester His<sub>6</sub> containing peptides and proteins. Ni-NTA beads were pre-equilibrated with water. In one pot, 50 µM LF<sub>N</sub>-LPSTGG-H<sub>6</sub>, 5 µM SrtA\*, and 1.6 mM GGGTLLGHTKHGC peptide were mixed in SrtA buffer (10 mM CaCl<sub>2</sub>, 50 mM Tris-HCl, 150

mM NaCl, pH 7.5) and reacted for 30 min at RT. Then, Ni-NTA beads were added to the mixture and incubated an additional 30 min at RT while rotating. After incubation, the beads were spun down and the supernatant was collected. The beads were washed twice with 20 mM Tris-HCl pH 7.5, 150 mM NaCl (to remove any non-specifically bound LF<sub>N</sub>). The supernatant and all washes were combined, concentrated, and buffer exchanged three times into 20 mM Tris-HCl, 150 mM NaCl, pH 7.5 to remove the excess GGGTLLGHTKHGC peptide. The purity of the desired ligated product was analyzed by HPLC/MS.

#### 6.3.4. Library Analysis

*Reaction conditions.* In the case of library 1 screening, solution of 20 mg/ml Mesna (15 mg/ml in the case of library 2) and 5 mg/ml perfluorophenylhydrazine (PFPh) in MES buffer (50 mM MES and 50 mM NaCl in water) was prepared, and the pH of resulting solution was adjusted to 6.7. 2 mg of library 1 or 2 was dissolved in this solution, and thermostated at 37 °C for 24 hours (library 1) or 18 hours (library 2). After, the reaction mixture was quenched with water containing 1% TFA (v/v) to a final library concentration corresponding to 25 pg/μl for each unreactive library member, filtered and subjected to nLC/MSMS analysis (1 μl of the quenched solution was injected, i.e. theoretical 25 pg/peptide was analyzed in both cases).

*Nano-LC/MSMS library analysis.* Nano-LC/MSMS analysis was performed on Thermo Fisher Orbitrap Fusion Lumos Tribrid Mass Spectrometer coupled to Thermo Fisher EASY-nLC 1200 System equipped with Acclaim PepMap RSLC C18 column (750 mm x 75 μm ID, 2 μm 100Å silica). Analysis was performed at 40 °C and a flow rate of 300 nL/min with the following gradient. 2% of 80% acetonitrile in water with 0.1% formic acid added (solvent B') in water containing 2% methanol and 0.1% FA (solvent A') ramping linearly to 52% B' in A' over 480 minutes, followed by 52-80% B' in A' ramping linearly over 10 minutes, followed by 80% B' in A' for 30 minutes. MSMS acquisition over the course of the method duration was performed in a data-dependent style (Top N=15, z=2-10, intensity threshold =  $3 \cdot 10^5$ ) with a dynamic precursor exclusion for 20 seconds after each scan. CID and HCD fragmentation spectra were acquired for every selected precursor ion. Orbitrap was used as a detection method for both primary (resolution=120000) and secondary (resolution=30000) mass spectra.

*De novo sequencing and data analysis.* *De novo* peptide sequencing was performed in PEAKS 8 from Bioinformatics Solutions Inc. (ON, Canada). Sequencing data obtained as described in the section above was refined as follows. HCD and CID scans were merged within a 0.2 minute and 10 ppm window, mass precursor correction was performed, and primary mass filtration was performed as appropriate. *De novo* sequencing was performed allowing 15 ppm assignment errors and 0.05 Da individual fragment mass errors. C-terminal Mesna thioesters, C-terminal N'-perfluorophenylhydrazide, or Mesna disulfide with Cys were allowed as post-translation modifications. 15 candidate sequences were obtained for each preprocessed scan.

After performing *de novo* sequencing as specified above, the list of all peptide candidates and a single top candidate peptide lists were exported as .csv files and further analyzed in Python 2.7.10. A number of open source libraries were used for parsing data and analyzing the outcomes: namely, Numpy,<sup>16</sup> Pandas,<sup>17</sup> Scipy,<sup>18</sup> and matplotlib.<sup>19</sup>

First, working with a list of all peptide candidates, sequences not matching library design rules (sequence length, constant region subsequence, correct monomers in every position) were eliminated from further consideration. Next, for scans with multiple remaining sequence candidates, a single peptide with highest ALC score per scan was retained, while the rest were excluded. After this, for sequences found in the dataset multiple times, the unique/non-unique assignment was performed. Assignments with higher ALC scores were assigned as unique, and all non-unique sequences were discarded. For sequences identified in two forms (reduced Cys and Mesna disulfide), one of the sequences (the one with lower ALC) was discarded. Finally, sequences with ALC scores lower than 80 were removed, and peptide lists obtained at this stage were considered the final product of the parsing. The data were visualized as needed using matplotlib.

### 6.3.5. HPLC-MS Analysis

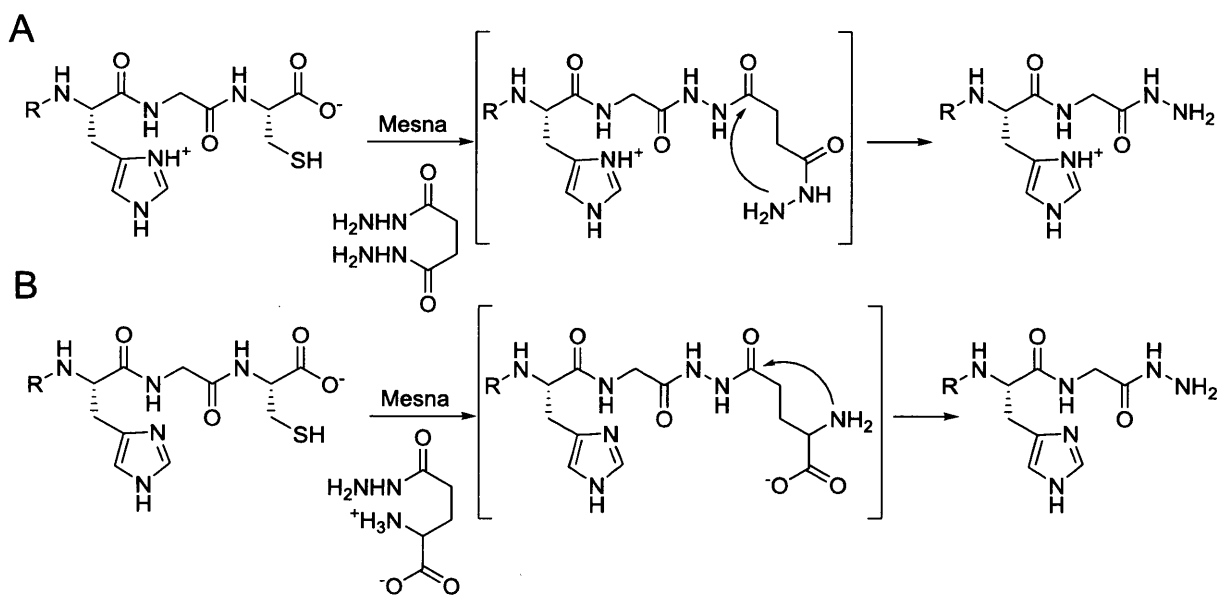
Batch HPLC/MS analysis (as opposed to the library analysis described above) was performed on an Agilent 6520 Accurate Mass Q-TOF LC-MS using an Agilent Zorbax 300SB C3 column (300 Å, 5 µm, 2.1 x 150 mm, used for the reaction yielding unsubstituted C-terminal hydrazides and for protein analysis) or an analogous C18 column for other peptides. At 40 °C and a flow rate of 0.8 mL/min, the following gradient was used: 1% acetonitrile with 0.1% formic acid added (FA, solvent B) in water with 0.1% FA (solvent A) for 2 min, 1-61% B in A' ramping linearly over 9 min, 61-95% B in A ramping linearly over 1 minute. Typically, 150 ng/µl solutions of peptides were subject to analysis. Unless noted, all chromatograms shown in this work are plots of total ion current (TIC) versus time. HPLC yields of reactions were calculated in the following way. For each product observed in the reaction mixture, all ions corresponding to that product were identified, and an extracted ion current chromatogram was generated for each ion. Extracted ion current chromatograms were integrated, and the area corresponding to individual reaction products were assumed to be sums of the corresponding extracted ion areas. Yields for each product were calculated as an area fraction of the total integrated area. Performed in this way, such analysis agreed well with manual integration of UV<sub>214</sub> traces (<5% relative difference). The data were analyzed using Agilent MassHunter Qualitative analysis software. MS deconvolution spectra were obtained using the maximum entropy algorithm.



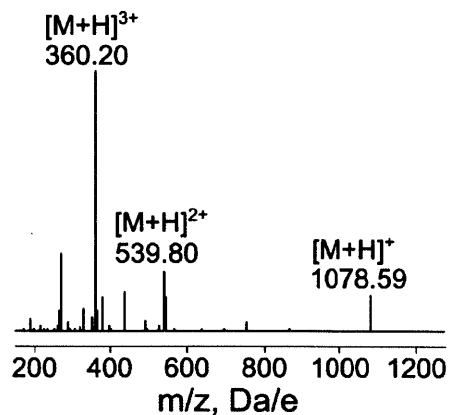
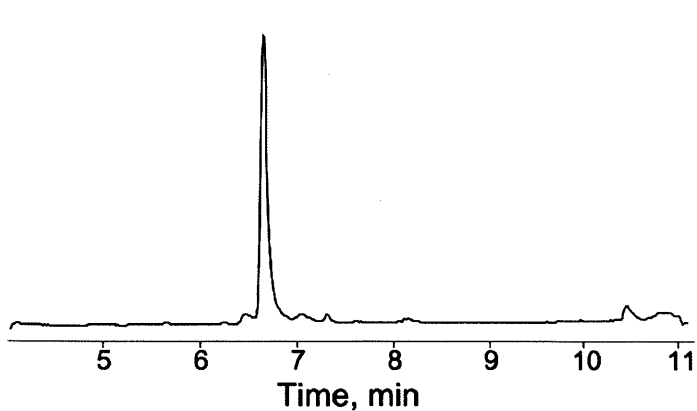
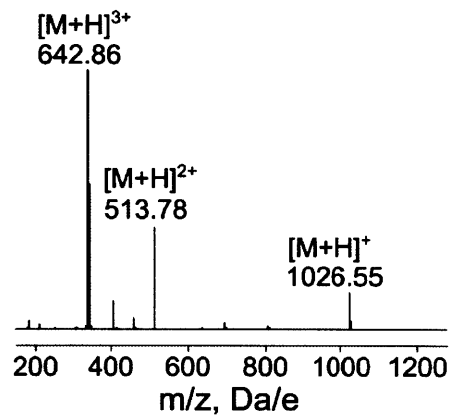
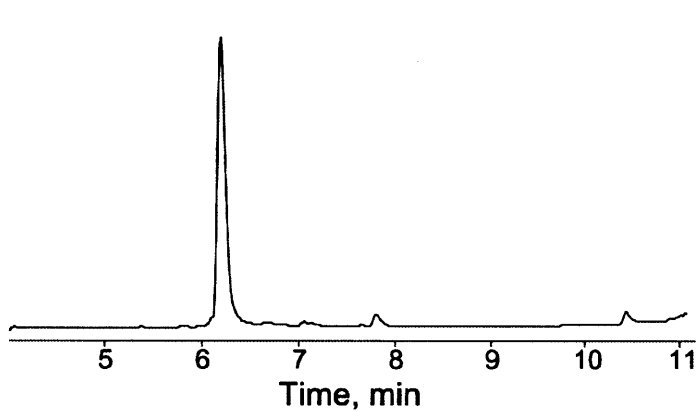
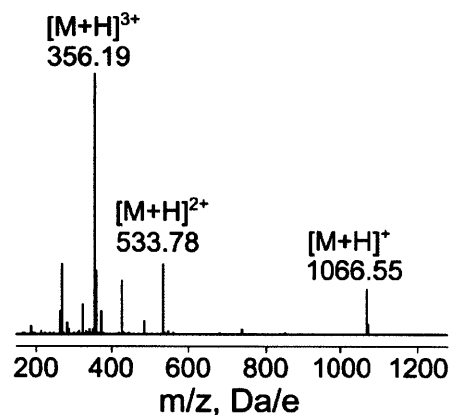
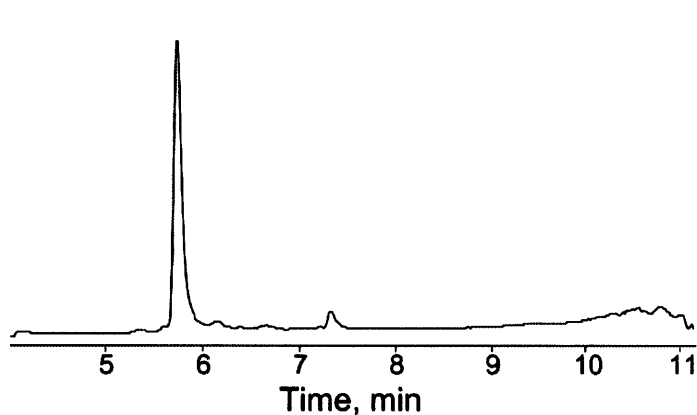
### 6.3.6. Conjugation to Different Nucleophiles

All batch conjugation reactions except for LF<sub>N</sub> labeling were performed in the following way. Solution of a hydrazine derivative was dissolved in MES buffer (50 mM MES, 50 mM NaCl, 100 mg/ml Mesna in water) and the pH of the resulting solution was adjusted to ~6.7. Peptides were then dissolved in this solution to 2 mg/ml concentration, and incubated in a thermostated water bath at 60 °C. Nucleophiles concentration were 7 mg/ml for perfluorophenylhydrazine, 50 mg/ml for isoniazid, nicotinic hydrazide, carbohydrazide, adipic dihydrazide, and 2-hydrazinopyridine, and 10 mg/ml for succinic dihydrazide and hydrazine. After 48 hours, the reaction mixtures were quenched with water containing 1% (v/v) TFA and analyzed by HPLC/MS as described above.

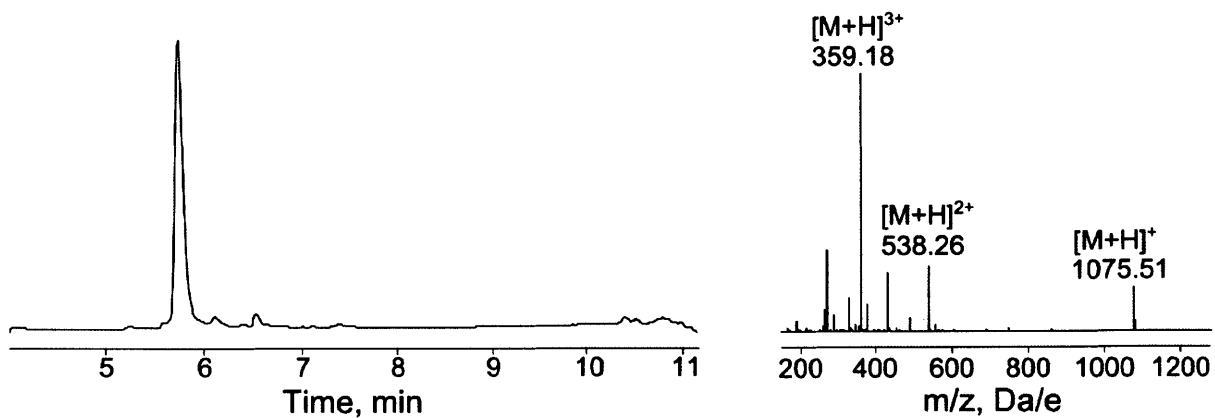
Reactions with succinic dihydrazide or glutamic acid  $\gamma$ -hydrazide (20 mg/ml) both yielded unsubstituted hydrazides as major reaction products for all tested peptides. Because no N'-substituted hydrazides were observed, and because only these two nucleophiles yielded such unexpected products we hypothesized that one potential route to these peptides can be an intramolecular cyclization of immediate conjugation products to liberate unsubstituted C-terminal peptide hydrazides as illustrated below.



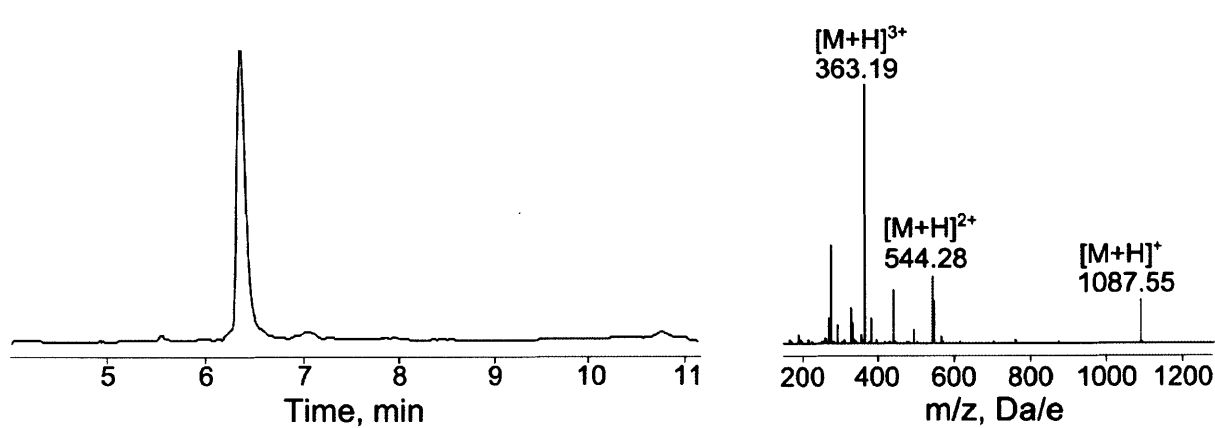
### 6.3.7. Peptide Synthesis: Analytical Data



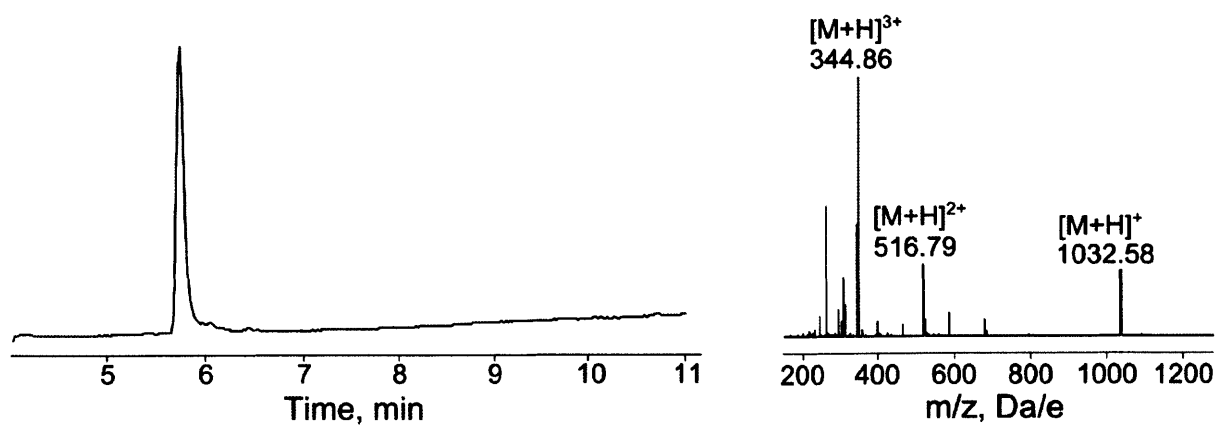
H<sub>2</sub>N-TLLGHTGHHC-COOH



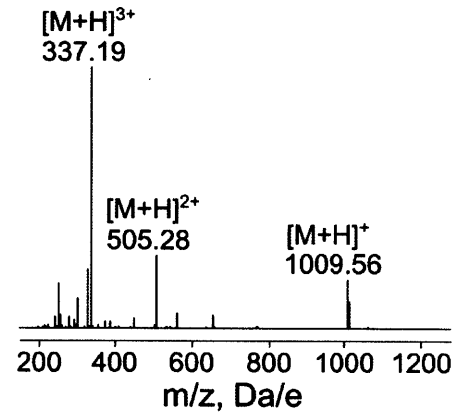
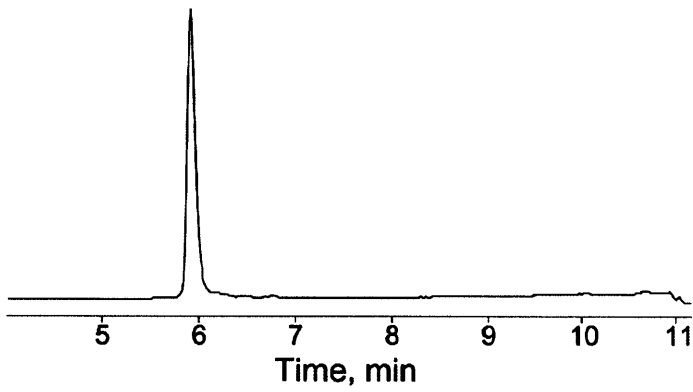
H<sub>2</sub>N-TLLGHGLHHC-COOH



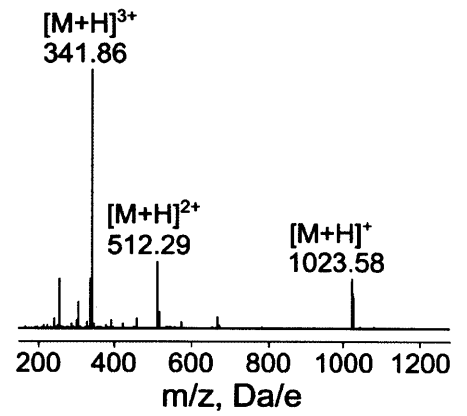
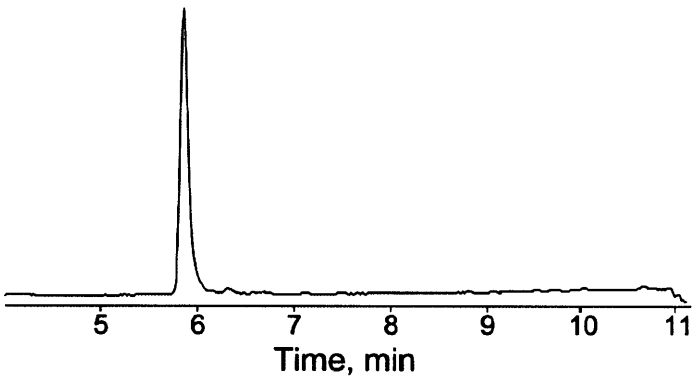
H<sub>2</sub>N-LKLPGKHHC-COOH



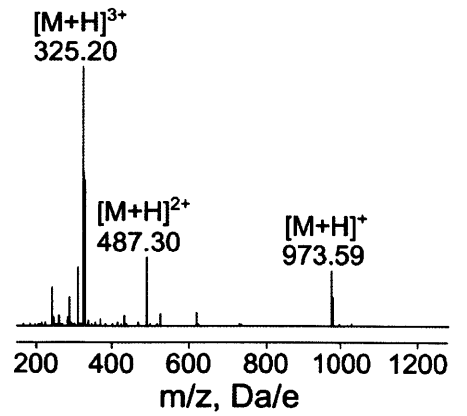
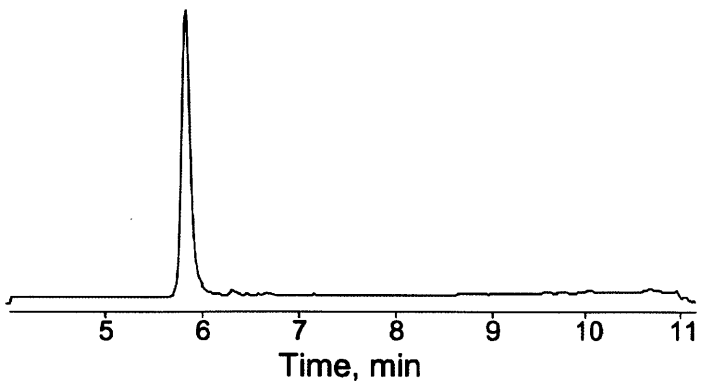
H<sub>2</sub>N-LKLPGKNHC-COOH



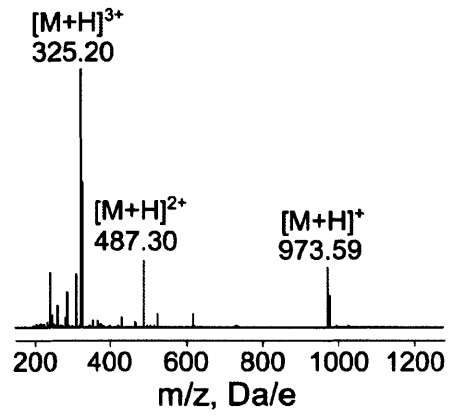
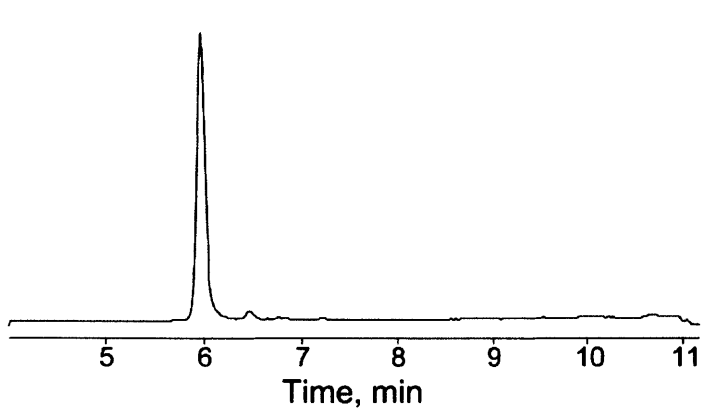
H<sub>2</sub>N-LKLPGHKQC-COOH



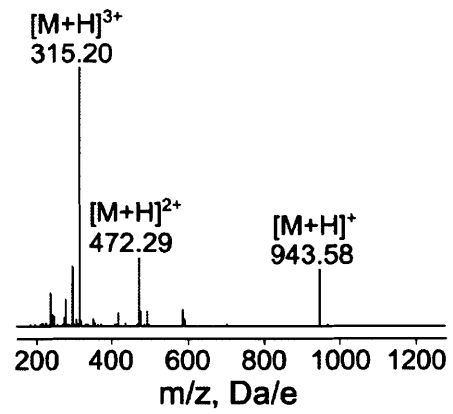
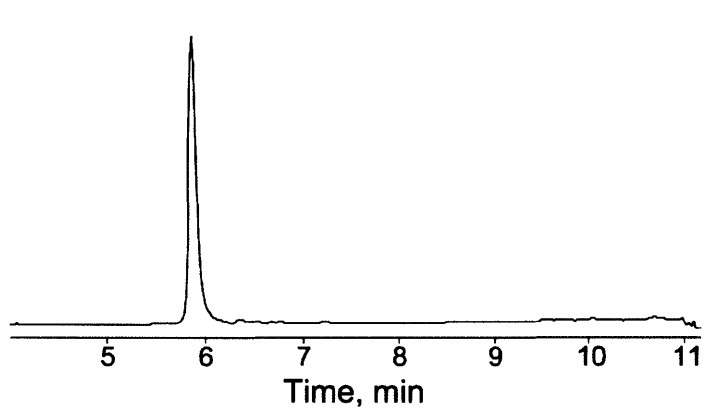
H<sub>2</sub>N-LKLPGKKSC-COOH



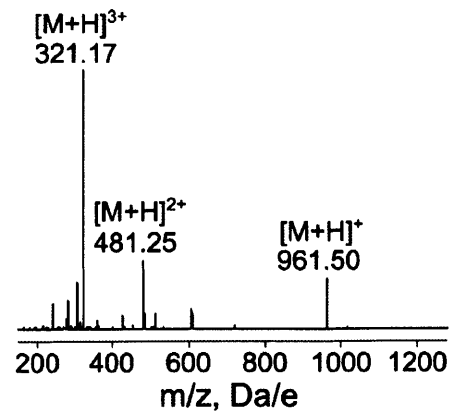
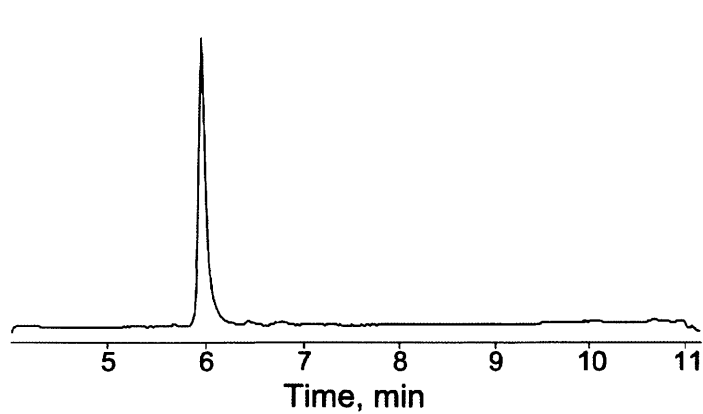
H<sub>2</sub>N-LKLPGSKKC-COOH



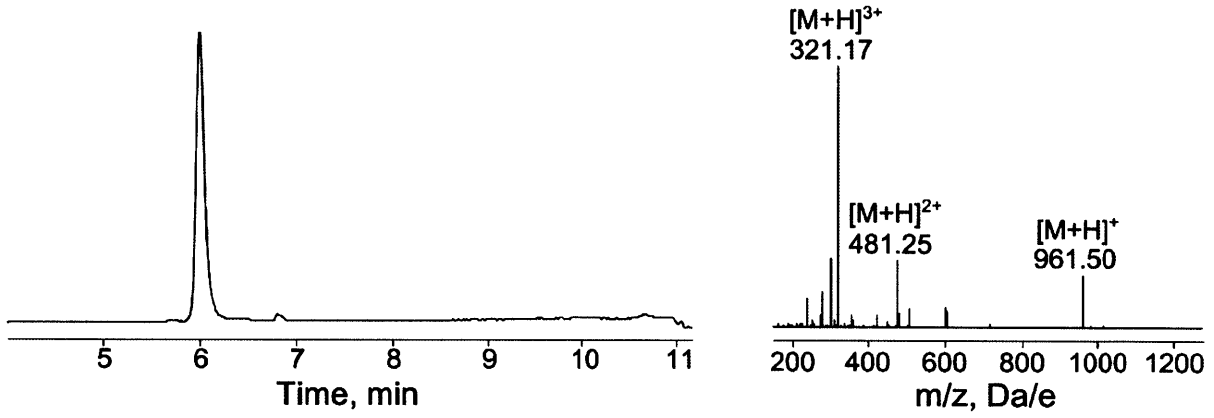
H<sub>2</sub>N-LKLPGKKGC-COOH



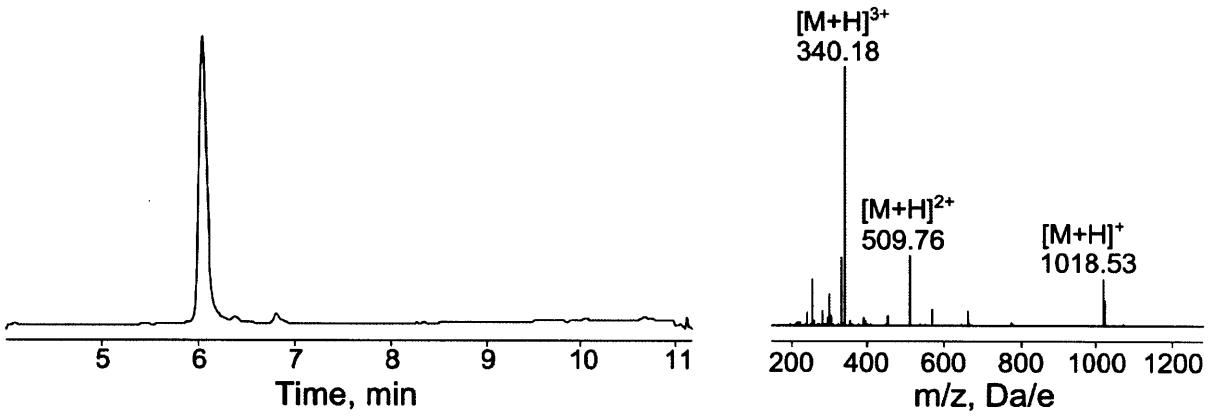
H<sub>2</sub>N-LKLPGHHGC-COOH



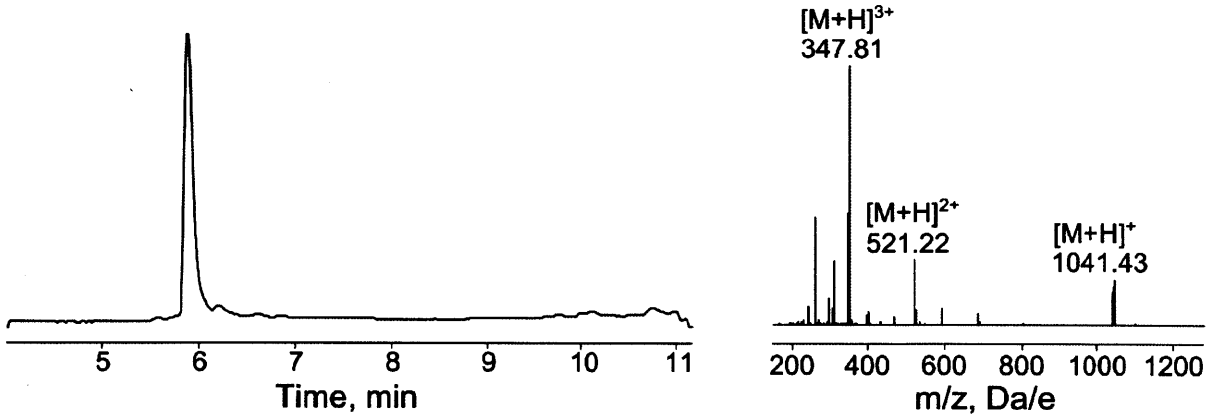
H<sub>2</sub>N-LKLPGHGHC-COOH



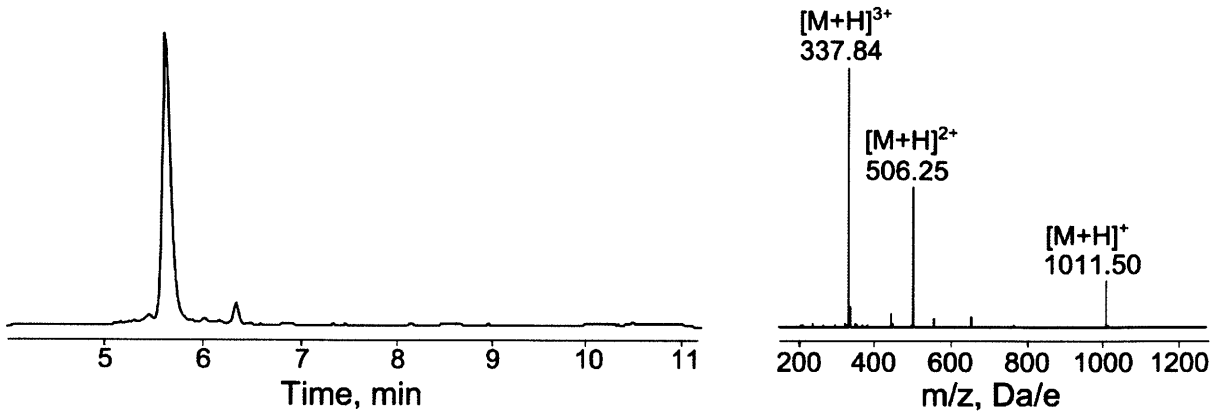
H<sub>2</sub>N-LKLPGNHHC-COOH



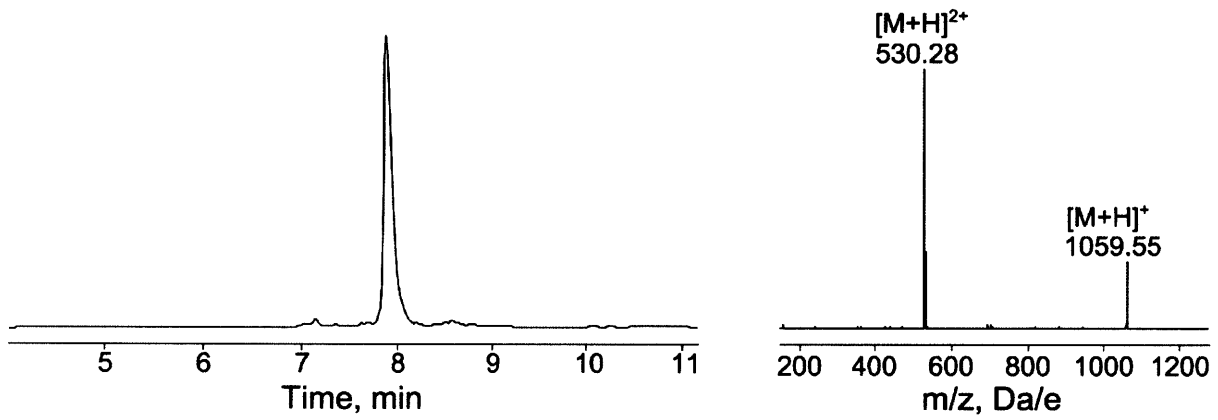
H<sub>2</sub>N-LKLPGHHHC-COOH



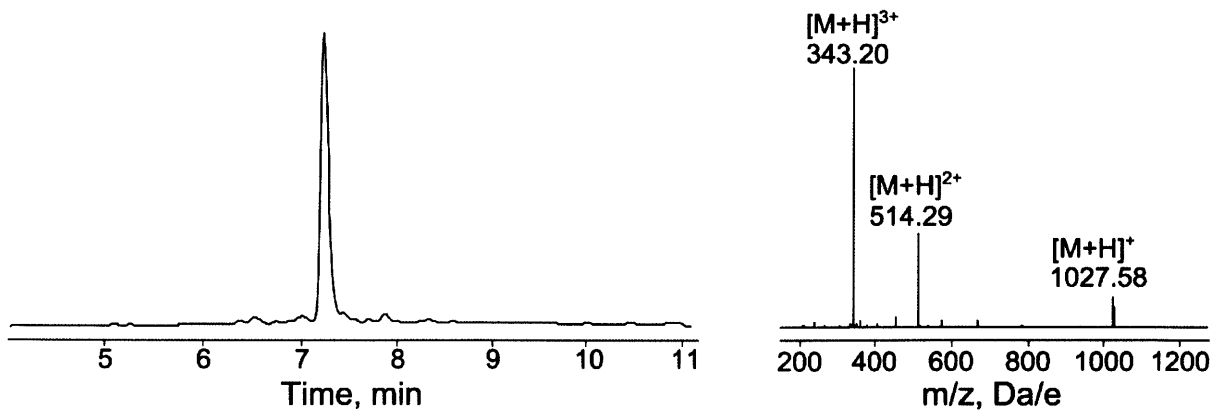
H<sub>2</sub>N-LKLPGEHDC-COOH



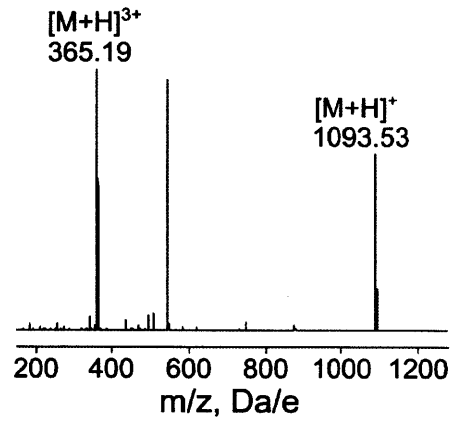
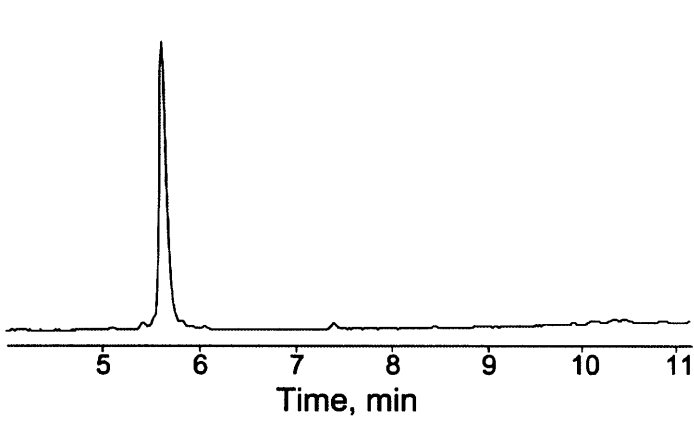
H<sub>2</sub>N-LKLPGWWGC-COOH



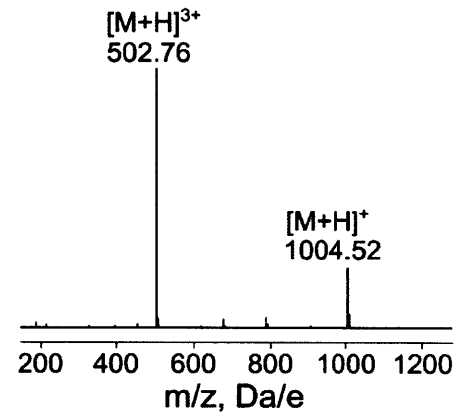
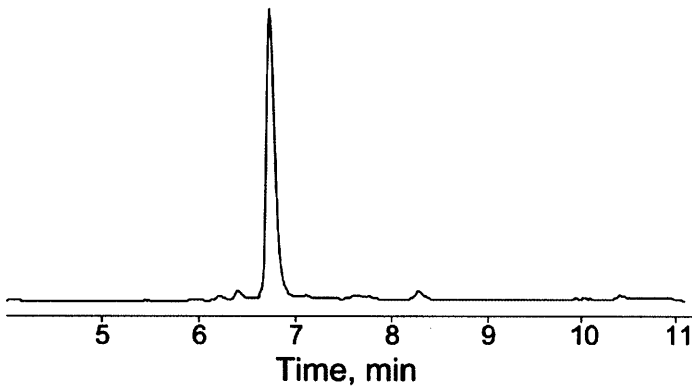
H<sub>2</sub>N-LKLPLGHFC-COOH



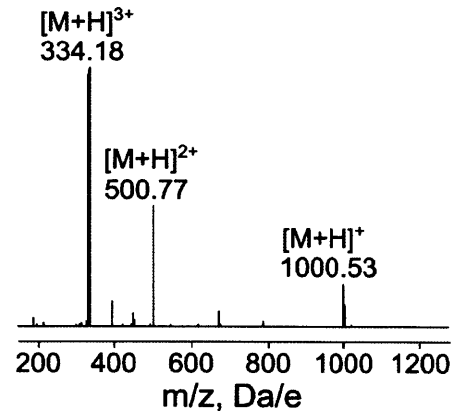
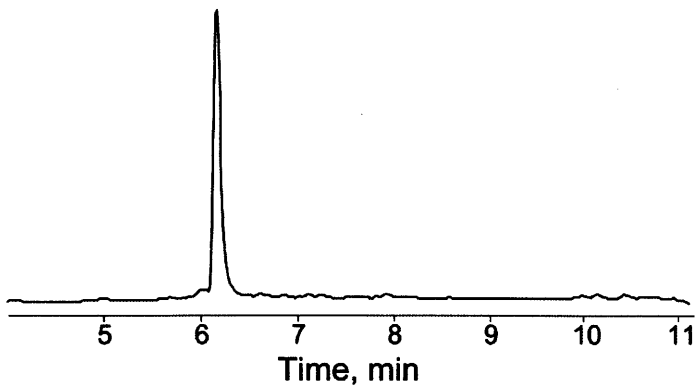
H<sub>2</sub>N-TLKELYSHC-COOH



H<sub>2</sub>N-TLLGPDGTKC-COOH

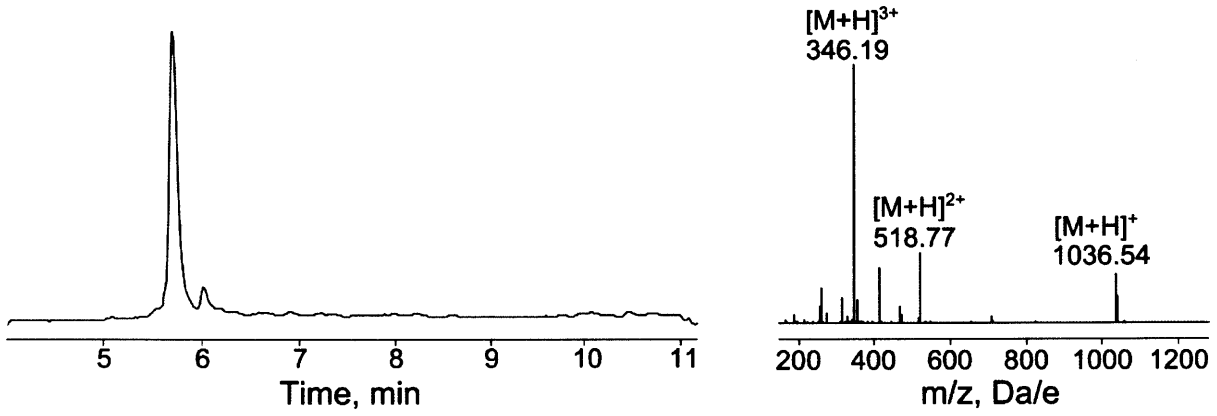


H<sub>2</sub>N-TLLGATKHGC-COOH

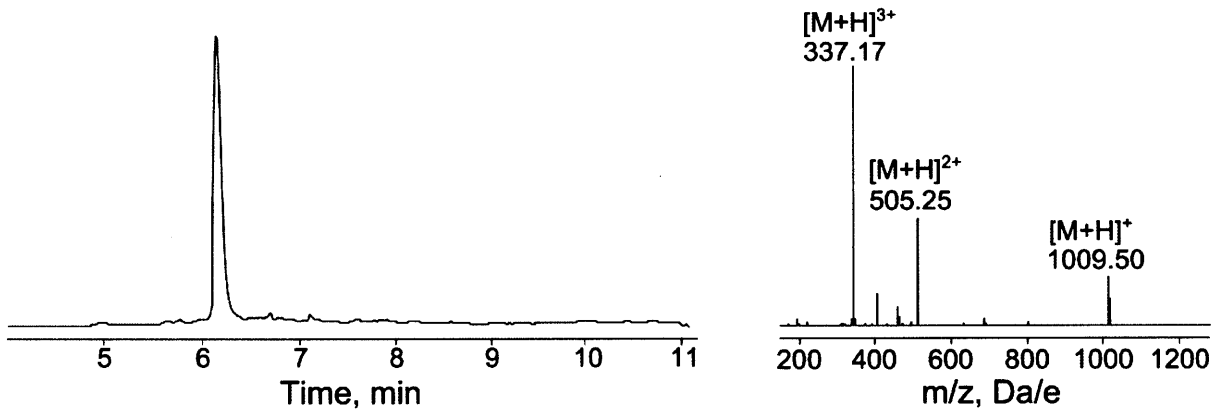




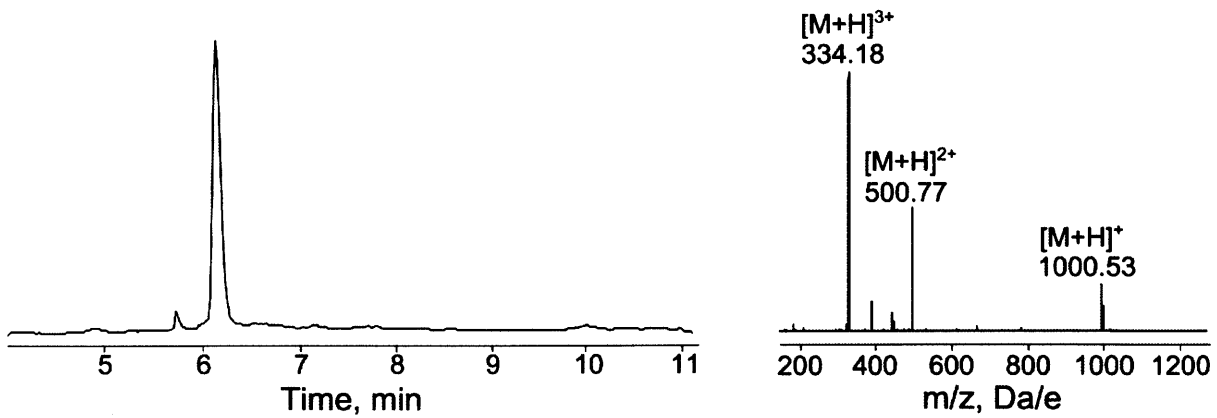
H<sub>2</sub>N-TLLGHAKHGC-COOH



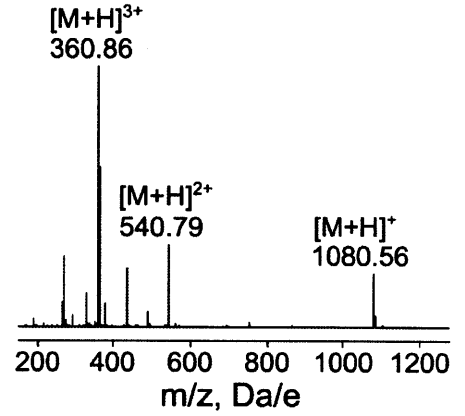
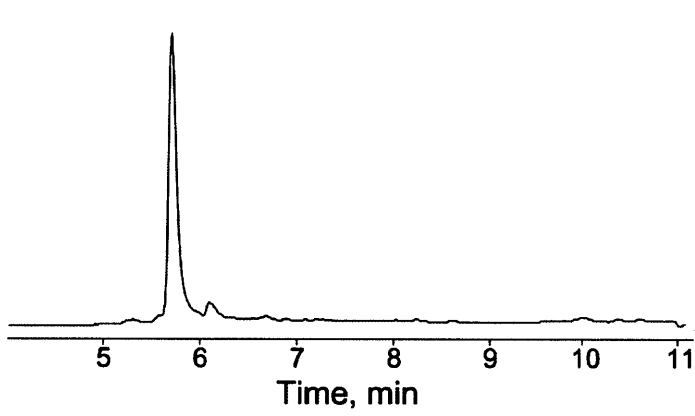
H<sub>2</sub>N-TLLGHTAHGC-COOH



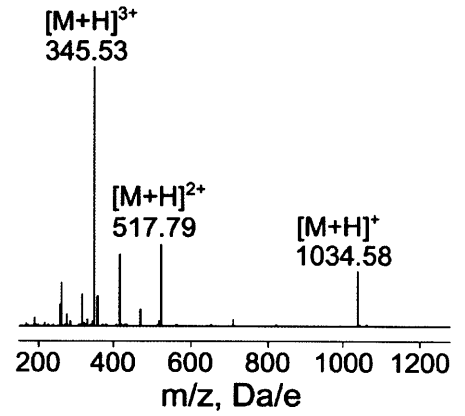
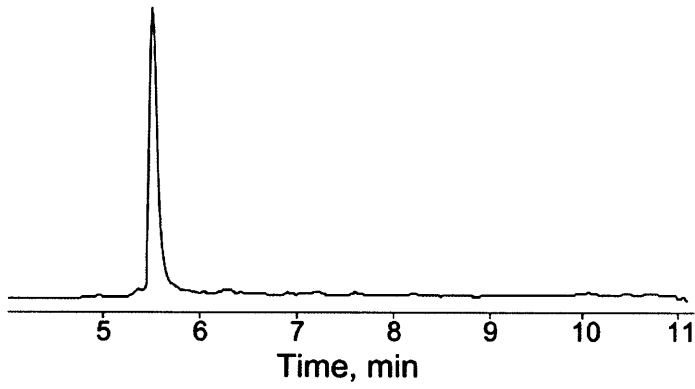
H<sub>2</sub>N-TLLGHTKAGC-COOH



H<sub>2</sub>N-TLLGHTKHAC-COOH



H<sub>2</sub>N-TLLGHTKHGA-COOH



### 6.3.8. Library 1 PFPh Dataset

ALC	Area	Mass	Peptide	RT	Scan	m/z	ppm	z
92	1.09E+06	1081.5244	LKLPGTHHp	58.84	11373	361.5144	0.2	3
96	1.07E+06	1037.4983	LKLPGHHGp	56.8	10715	346.8392	-2.2	3
81	1.07E+06	1037.4983	LKLPGHGp	57.43	10909	346.8397	-1	3
91	9.06E+05	1094.5198	LKLPGNHHp	55.05	10178	365.8466	-1.5	3
90	9.05E+05	1028.5342	LKLPKGHGp	54.85	10121	343.8515	1.6	3
89	8.38E+05	1103.5599	LKLPGLHFp	145.45	42175	368.8605	-1.9	3
97	8.15E+05	1117.5249	LKLPGHHHp	58.07	11164	373.5152	-2.9	3
91	7.52E+05	997.4921	LKLPGPHGp	88.87	21638	333.5028	-2.4	3
82	7.51E+05	1102.5532	LKLPKGHMP	71.7	15390	368.5229	-2.8	3
89	6.68E+05	1087.4769	LKLPGEHDp	88.45	21243	363.5002	0.1	3
95	6.44E+05	1047.5078	LKLPGFHGp	118.39	34118	350.1747	-2.2	3
92	6.14E+05	1095.5037	LKLPGDHHp	61.81	12295	366.1745	1.2	3
89	6.00E+05	1014.4822	LKLPGNHGp	72.92	15785	339.1672	0.6	3
91	5.90E+05	1001.487	LKLPGTHGp	80.14	18304	334.834	-4	3
93	5.40E+05	999.5077	LKLPGVHGp	100	26128	334.1746	-2.8	3
89	5.25E+05	1028.4978	LKLPGQHGp	73.51	15992	343.8381	-2.3	3
89	5.14E+05	1094.5198	LKLPGHHNp	54.2	9919	365.8468	1.9	3
88	4.85E+05	1013.5234	LKLPGLHGp	112.25	31462	338.8459	-4.3	3
96	4.84E+05	1089.5547	LKLPGVHFp	137.07	41335	364.1901	-2.7	3
89	4.68E+05	1042.5134	LKLPGQHAp	76.14	16892	348.5109	0.6	3
92	4.45E+05	971.4764	LKLPGAHGp	79.62	18121	324.8307	-3.2	3
91	4.18E+05	1043.4976	LKLPGEAHP	83.64	19632	348.8378	-2.7	3
92	4.03E+05	987.4713	LKLPGSHGp	74.24	16244	330.1623	-3.4	3
95	3.88E+05	1031.4797	LKLPGMHGp	103.39	27575	344.832	-2.3	3
90	3.86E+05	1029.4819	LKLPGEHGp	80.55	18445	344.1659	-2.8	3
86	3.86E+05	1086.5398	LKLPKGHDp	60.65	11942	363.1864	0.8	3
95	3.63E+05	1091.5339	LKLPGTHFp	123.29	36187	364.8498	-2.8	3
81	3.62E+05	1119.5288	LKLPGEHFp	119.59	34622	374.1813	-3.1	3
94	3.45E+05	1135.5279	LKLPGWWGp	150.8	44127	379.5162	-2.8	3
92	3.37E+05	1087.5425	LKLPGLHMP	132.7	39796	363.5194	-2.6	3
81	3.35E+05	1075.5061	LKLPGTHMP	105.72	28558	359.5069	-3.7	3
93	3.17E+05	1001.487	LKLPGSHAp	77.43	17322	334.8339	-4.1	3
92	3.13E+05	1088.5012	LKLPGNHMP	96.24	24594	363.8394	-1.4	3
90	2.56E+05	1085.5193	LKLPGQHNP	69.83	14742	362.8456	-1.1	3
90	2.52E+05	1015.4662	LKLPGDHGp	82.64	19272	339.4945	-1.5	3
85	2.15E+05	1095.5037	LKLPGHHDp	60.55	11908	366.1749	2.4	3
88	1.90E+05	1088.5012	LKLPGMHNp	97.46	25085	363.8392	-2.1	3
89	1.46E+05	1045.4954	LKLPGMHAp	105.06	28268	349.5035	-3.3	3
81	1.30E+05	1066.4692	LKLPGNMDp	130.86	39125	534.2394	-1.7	2

## 6.3.9. Library 1 Mesna Dataset

ALC	Area	Mass	Peptide	RT	Scan	m/z	ppm	z
93	4.28E+06	945.4412	LKLPGASHm	47.48	7918	473.7273	1.8	2
80	4.13E+06	982.4979	LKLPGKFGm	87.8	21220	492.254	-1.5	2
90	3.73E+06	1043.5256	LKLPGHKQm	35.14	4891	348.8487	1.8	3
86	3.55E+06	963.5245	LKLPGKKGm	32.55	4910	322.1816	2.2	2
90	3.55E+06	993.4987	LKLPGSKQm	45.08	7257	497.7559	1.6	2
92	3.48E+06	994.4827	LKLPGESKm	49.66	8554	498.2482	2.2	2
89	3.24E+06	993.4987	LKLPGNTKm	46.1	7536	497.7564	2.6	2
89	3.23E+06	993.5351	LKLPGKKS m	31.69	4617	332.1851	1.4	3
91	3.14E+06	1021.4936	LKLPGDQKm	48.74	8293	511.7532	1.2	2
86	3.14E+06	1021.4936	LKLPGDKQm	49.26	8449	511.7533	1.4	2
88	3.12E+06	963.4881	LKLPGQGKm	44.96	7223	482.7508	1.8	2
82	3.12E+06	963.4881	LKLPGAKNm	45.47	7370	482.7508	1.9	2
84	2.90E+06	1012.5197	LKLPGKPHm	40.02	5932	338.5135	1.9	3
91	2.83E+06	972.4521	LKLPGNAHm	46.53	7643	487.2328	2	2
87	2.83E+06	972.4521	LKLPGQGHm	46.02	7517	487.2325	1.3	2
93	2.78E+06	1017.4623	LKLPGTEHm	54.26	9944	509.7383	2.8	2
92	2.78E+06	959.4568	LKLPGATHm	50.49	8814	480.7352	2	2
92	2.78E+06	959.4568	LKLPGTAHm	51.01	8975	480.735	1.5	2
84	2.78E+06	1017.4623	LKLPGETHm	53.22	9618	509.7381	2.3	2
89	2.73E+06	1003.4467	LKLPGESHm	50.33	8761	502.7301	1.9	2
91	2.64E+06	966.4878	LKLPGTKSm	49.61	8542	484.2507	2	2
92	2.60E+06	989.5038	LKLPGNPKm	54.24	9933	495.7586	1.9	2
91	2.50E+06	936.4772	LKLPGSKAm	51.55	9139	469.2454	2	2
85	2.50E+06	1031.4966	LKLPGLMHm	98.28	25411	516.7523	-3.4	2
80	2.50E+06	936.4772	LKLPGTKGm	52.07	9287	469.2454	2	2
81	2.49E+06	949.4725	LKLPGNGKm	42.91	6624	475.743	1.9	2
91	2.44E+06	988.447	LKLPGNSHm	43.13	6679	495.2308	3.2	2
92	2.40E+06	1002.4626	LKLPGNTHm	47.08	7803	502.2379	1.6	2
90	2.37E+06	1014.5354	LKLPGVKHm	48.72	8285	339.185	0.9	3
93	2.36E+06	1047.5244	LKLPGFLHm	115.37	32823	524.7664	-3	2
86	2.36E+06	1029.5099	LKLPGNKHm	34.42	4748	344.1769	2.1	3
89	2.33E+06	1029.5099	LKLPGKNHm	33.41	4536	344.1769	2	3
90	2.31E+06	987.4517	LKLPGEAHm	53.16	9599	494.7327	2	2
84	2.29E+06	1020.5096	LKLPGNKQm	43.76	6845	511.2617	2.4	2
82	2.21E+06	986.4677	LKLPGQAHm	48.27	8153	494.2406	2	2
80	2.21E+06	986.4677	KLLPGAQHm	48.78	8304	494.2408	2.2	2
91	2.11E+06	1052.5259	LKLPGKHHm	29.79	4113	351.8491	2.5	3
92	2.10E+06	936.4772	LKLPGASKm	47.98	8066	469.2452	1.5	2
89	2.10E+06	1030.4939	LKLPGDKHm	37.47	5386	344.5045	0.7	3
88	2.08E+06	977.5038	LKLPGAQm	49.39	8484	489.7585	1.7	2

93	2.06E+06	936.4772	LKLPGSAK <sub>m</sub>	46.96	7760	469.2456	2.4	2
93	2.06E+06	989.4674	LKLPGTTH <sub>m</sub>	51.97	9254	495.7404	1.7	2
88	2.01E+06	986.5041	LKLPGAKH <sub>m</sub>	36.92	5278	329.8414	1.4	3
80	1.98E+06	932.4823	LKLPGKGP <sub>m</sub>	55.87	10434	467.2473	0.6	2
87	1.96E+06	1016.5146	LKLPGKTH <sub>m</sub>	34.96	4853	339.8449	1.4	3
84	1.94E+06	963.4881	LKLPGANK <sub>m</sub>	48.15	8124	482.7509	2.2	2
82	1.94E+06	1012.4833	LKLPGQPH <sub>m</sub>	57.06	10799	507.2485	2.1	2
86	1.93E+06	1001.5038	LKLPGTLH <sub>m</sub>	74	16167	501.7562	-2.9	2
94	1.87E+06	1060.4834	LKLPGSWH <sub>m</sub>	88.71	21587	531.2468	-1.2	2
90	1.87E+06	1051.4829	LKLPGTYH <sub>m</sub>	68.3	14241	526.7468	-0.7	2
81	1.85E+06	1002.4626	LKLPGHTN <sub>m</sub>	43.94	6899	502.2386	2.9	2
86	1.75E+06	1002.4626	LKLPGTNH <sub>m</sub>	49.74	8578	502.2382	2.2	2
94	1.73E+06	1017.481	LKLPGVMH <sub>m</sub>	84.33	19884	509.7444	-3.6	2
85	1.71E+06	1028.551	LKLPGKLN <sub>m</sub>	50.53	8826	343.857	1.1	3
91	1.70E+06	950.4929	LKLPGTAK <sub>m</sub>	50.23	8724	476.2532	1.9	2
82	1.70E+06	950.4929	LKLPGATK <sub>m</sub>	50.75	8893	476.2531	1.8	2
88	1.68E+06	993.4987	LKLPGKSQ <sub>m</sub>	41.87	6365	497.7565	2.8	2
92	1.67E+06	906.4667	LKLPGAGK <sub>m</sub>	48.13	8120	454.2401	1.9	2
92	1.67E+06	987.4881	LKLPGSLH <sub>m</sub>	74	16165	494.7491	-1.5	2
85	1.67E+06	906.4666	LKLPGKAG <sub>m</sub>	48.64	8268	454.24	1.6	2
83	1.63E+06	1005.4446	LKLPGHSM <sub>m</sub>	69.55	14658	503.7273	-1.6	2
82	1.62E+06	1020.5347	LKLPGELK <sub>m</sub>	79.69	18145	511.2726	-1	2
88	1.56E+06	936.4772	LKLPGAKS <sub>m</sub>	49.5	8515	469.2451	1.3	2
82	1.55E+06	1120.5198	LKLPGHWF <sub>m</sub>	133.97	40238	561.265	-0.8	2
93	1.53E+06	952.4721	LKLPGSKS <sub>m</sub>	43.84	6866	477.243	2.3	2
92	1.53E+06	931.4255	LKLPGSGH <sub>m</sub>	45.25	7313	466.7195	1.9	2
84	1.53E+06	931.4255	LKLPGGSH <sub>m</sub>	44.22	6998	466.7196	2.1	2
80	1.53E+06	931.4255	LKLPGHSG <sub>m</sub>	44.73	7159	466.7195	1.9	2
92	1.50E+06	920.4823	LKLPGKAAM	51.24	9045	461.2476	1.2	2
90	1.50E+06	1000.4833	LKLPGNVH <sub>m</sub>	62.05	12370	501.2481	1.3	2
83	1.50E+06	1034.5251	LKLPGQKQ <sub>m</sub>	45.76	7453	518.269	1.4	2
91	1.49E+06	1016.4783	LKLPGTQH <sub>m</sub>	50.05	8670	509.2459	2	2
91	1.48E+06	945.4412	LKLPGTGH <sub>m</sub>	50.13	8691	473.7271	1.5	2
80	1.48E+06	1089.5464	LKLPGRWP <sub>m</sub>	105.1	28284	545.7774	-2.6	2
82	1.47E+06	1063.4829	LKLPGEFH <sub>m</sub>	87.05	20949	532.7465	-1.3	2
93	1.45E+06	975.4517	LKLPGTSH <sub>m</sub>	49.45	8499	488.7327	2.1	2
87	1.44E+06	989.5038	LKLPGKPN <sub>m</sub>	52.32	9364	495.7583	1.2	2
89	1.42E+06	1052.4895	LKLPGHQH <sub>m</sub>	34.7	4800	351.8368	2.2	3
89	1.41E+06	1086.5354	LKLPGLWH <sub>m</sub>	117.08	33558	544.2728	-1.1	2
93	1.39E+06	961.4361	LKLPGSSH <sub>m</sub>	44.45	7065	481.7249	2.1	2
89	1.39E+06	972.4521	LKLPGHAN <sub>m</sub>	44.5	7079	487.2328	2	2
81	1.38E+06	1029.4987	LKLPGELH <sub>m</sub>	78.33	17652	515.7554	0.7	2
85	1.37E+06	1037.515	LKLPGHLH <sub>m</sub>	51.92	9237	346.845	1.2	3

89	1.31E+06	1015.4579	LKLPGNNHm	43.26	6711	508.7358	2.3	2
89	1.31E+06	1046.4712	LKLPGHQm	69.56	14661	524.2408	-1	2
85	1.30E+06	1005.4445	LKLPGSMHm	64.98	13236	503.7286	1.2	2
86	1.25E+06	1029.4735	LKLPGQNHm	45.16	7284	515.744	3	2
91	1.24E+06	1043.562	LKLPGHKKm	31.25	4138	348.861	2.2	3
92	1.23E+06	1072.512	LKLPGYKmm	98.2	25380	537.2603	-2.6	2
89	1.23E+06	1019.4602	LKLPGMTHm	71.51	15323	510.7363	0.9	2
89	1.20E+06	980.5035	LKLPGTTKmm	51.52	9129	491.2582	1.4	2
88	1.19E+06	950.4929	LKLPGTKAm	58.93	11405	476.2529	1.2	2
92	1.18E+06	971.4568	LKLPGSPHm	56.4	10597	486.7354	2.3	2
93	1.17E+06	1072.5449	LKLPGKFFm	126.64	37502	537.2781	-0.1	2
88	1.15E+06	950.4929	LKLPGKTAmm	51.26	9051	476.253	1.5	2
90	1.14E+06	998.4677	LKLPGNPHm	55.1	10191	500.2408	2.4	2
89	1.12E+06	989.431	LKLPGSDHm	50.47	8806	495.722	1.5	2
91	1.11E+06	1030.4939	LKLPGKHDm	38.48	5596	344.5049	2.1	3
92	1.10E+06	999.4517	LKLPGDPHm	61.66	12246	500.7323	1.3	2
90	1.10E+06	922.4252	LKLPGSANm	66.74	13770	462.2174	-2.3	2
91	1.08E+06	1101.5464	LKLPGHwKmm	66.29	13628	368.1873	-2.9	3
80	1.08E+06	948.4885	LKLPGRAAm	56	10474	475.2508	1.6	2
91	1.06E+06	1034.5615	LKLPGQKKm	34.5	4761	345.8607	1.8	3
95	1.05E+06	1101.5464	LKLPGKWHm	64.08	12968	368.1873	-2.9	3
86	1.03E+06	1013.4673	LKLPGEPHm	64.45	13079	507.7393	-0.3	2
81	1.01E+06	1049.5149	LKLPGRWGm	97.7	25180	525.7623	-1.7	2
82	1.00E+06	922.4615	LKLPGSKGm	46.25	7576	462.2368	0.4	2
88	9.87E+05	920.4823	LKLPGAKAm	58.97	11419	461.2476	1.2	2
89	9.85E+05	964.4721	LKLPGADKmm	53.37	9661	483.2441	4.6	2
90	9.57E+05	1024.4756	LKLPGDMKmm	71.24	15226	513.2429	-1.2	2
92	9.40E+05	962.4929	LKLPGSPKmm	55.69	10372	482.2529	1.4	2
81	9.00E+05	986.4677	LKLPGHAQmm	47.25	7857	494.2409	2.5	2
90	8.86E+05	986.5041	LKLPGHKAm	38.65	5635	329.8416	1.9	3
83	8.81E+05	1023.4915	LKLPGMKNm	66.78	13785	512.7504	-2.2	2
87	8.39E+05	1086.5354	LKLPGWLHm	120.93	35180	544.2726	-1.4	2
90	8.21E+05	964.4721	LKLPGDAKmm	51.39	9089	483.2455	7.6	2
80	8.08E+05	993.5351	LKLPGSKKmm	34.04	4665	332.1853	1.9	3
90	7.97E+05	1031.4415	LKLPGDEHm	54.28	9951	516.7277	2.3	2
88	7.74E+05	1016.4419	LKLPGNDHm	54.56	10020	509.2279	2.3	2
91	7.67E+05	929.4462	LKLPGA AHm	50.9	8940	465.7296	1.3	2
80	7.52E+05	997.4976	LKLPGSFLm	88.87	21638	333.5028	-8	3
89	7.36E+05	989.431	LKLPGDShm	49.36	8473	495.7224	2.2	2
84	7.36E+05	929.4462	LKLPGHA Am	54.18	9914	465.7301	2.4	2
93	7.04E+05	1041.5061	LKLPGMLFm	174.02	55041	521.757	-3.4	2
88	7.02E+05	1014.5354	LKLPGKVHm	42.17	6434	339.1853	1.8	3
91	6.89E+05	915.4306	LKLPGAGHm	48.39	8183	458.7221	2	2

91	6.81E+05	1040.4823	LKLPGWGFm	167.81	52786	521.2471	0.4	2
91	6.66E+05	1063.4829	LKLPGFEHm	91.64	22782	532.7471	-0.1	2
86	6.21E+05	1087.4941	LKLPGHYHm	47.78	8003	363.5049	1.9	3
89	6.06E+05	973.4725	LKLPGSVHm	62.06	12372	487.7426	1.2	2
83	6.06E+05	879.4194	KLLPGAASm	74.16	16221	440.7147	-2.1	2
85	6.03E+05	1014.499	LKLPGVQHm	65.12	13277	508.2552	-0.1	2
84	5.77E+05	1053.4736	LKLPGHEHm	36.89	5273	352.1648	1.9	3
83	5.25E+05	1080.5137	KLLPGWPFm	178.2	56501	541.2622	-0.5	2
82	5.09E+05	1064.4856	LKLPGWPMm	156	48371	533.2483	-0.3	2
86	5.08E+05	980.467	LKLPGSKDm	55.49	10314	491.2404	2.3	2
84	4.89E+05	994.4827	LKLPGTKDm	61.07	12070	498.2476	0.8	2
90	4.77E+05	1039.4579	LKLPGHHDm	39.57	5834	347.4928	1.7	3
88	4.35E+05	964.4721	LKLPGAKDm	59.32	11536	483.2422	0.7	2
80	2.69E+05	1024.4756	LKLPGKMDm	75.48	16676	513.2425	-2	2
85	6.99E+04	964.4721	LKLPGKADm	55.18	10218	483.2493	14.4	2

## 6.3.10. Library 2 PFPh Dataset

ALC	Area	Mass	Peptide	RT	Scan	m/z	ppm	z
80	4.94E+06	1142.5408	TLLGHTKHGp	107.98	27360	381.8534	-2.2	3
84	3.84E+06	1155.5361	TLLGKNGHhp	98.44	22893	386.1853	-1.7	3
84	3.61E+06	1102.5347	TLLGKPTHGp	167.27	54010	368.5179	-2.5	3
85	3.04E+06	1075.4985	TLLGKNGHGp	147.38	44530	359.5061	-1.9	3
88	2.77E+06	1222.5784	TLLGKTHHHp	82.83	16121	408.5329	-1.2	3
88	2.77E+06	1222.5784	TLLGHTKHhp	83.43	16346	408.5328	-1.4	3
82	2.58E+06	1195.5674	TLLGKNPHhp	107.43	27097	399.5292	-1.3	3
85	2.42E+06	1186.4944	TLLGNTDHHp	178.16	59371	396.5046	-1.9	3
89	2.06E+06	1151.5049	TLLGHTGHHp	109.74	28209	384.8416	-1.8	3
82	2.00E+06	1178.552	TLLGKGHHhp	78.8	14507	393.8574	-1.5	3
88	1.94E+06	1154.5771	TLLGGLKHhp	173.51	57111	385.8652	-3	3
88	1.89E+06	1213.6143	TLLGKTKHHp	81.2	15458	405.5447	-1.5	3
83	1.89E+06	1146.572	TLLGNKKHGp	95.98	21739	383.1972	-2	3
80	1.89E+06	1204.5564	TLLGNKWHGp	224.1	80793	402.5255	-1.5	3
83	1.86E+06	1163.5413	TLLGHGLHhp	152.81	47066	388.8538	-1.4	3
92	1.83E+06	1168.52	TLLGNTPHhp	169.55	55151	390.5132	-2	3
82	1.80E+06	1147.5098	TLLGHGPHhp	115.51	30865	383.5102	-0.9	3
90	1.73E+06	1227.5725	TLLGHKWHGp	182.07	61296	410.1973	-2.1	3
89	1.72E+06	1074.5398	TLLGLGKHGp	213.92	76105	359.1867	-1.3	3
89	1.71E+06	1032.4928	TLLGGAKHGp	158.24	49664	345.171	-1.6	3
84	1.71E+06	1145.6133	TLLGKVAHKp	159.87	50428	382.8777	-1.7	3
94	1.70E+06	1170.5356	TLLGDAKHhp	159.06	50050	391.1854	-1.1	3
84	1.70E+06	1151.5049	TLLGTHGHHp	112.88	29636	384.8414	-2.3	3
82	1.53E+06	1142.5408	TLLGKHTHGp	112.66	29535	381.8539	-0.8	3
89	1.52E+06	1213.5205	TLLGHNWHGp	224.63	81045	405.5134	-1.7	3
85	1.52E+06	1226.6096	TLLGNKKHHp	75.38	13166	409.8768	-0.9	3
84	1.43E+06	1142.468	TLLGDNHHGp	181.61	61063	381.829	-2.4	3
89	1.39E+06	1138.5459	TLLGKGPHhp	114.26	30276	380.5221	-1.3	3
93	1.38E+06	1227.5935	TLLGDKKHhp	92.01	19971	410.2047	-1.1	3
86	1.36E+06	1200.5251	TLLGWHTHGp	226.61	81963	401.1816	-1.7	3
85	1.35E+06	1214.5044	TLLGWGDHHp	243.91	89907	405.8415	-1.5	3
92	1.34E+06	1123.5349	TLLGLGPHhp	217.28	77647	375.5183	-1.7	3
86	1.34E+06	1227.5725	TLLGKWGHhp	183.89	62152	410.197	-2.7	3
86	1.32E+06	1222.5056	TLLGDNHHhp	131.08	38084	408.5087	-1.2	3
89	1.30E+06	1147.5349	TLLGWGKHGp	226.5	81913	383.5184	-1.3	3
82	1.29E+06	1120.4878	TLLGWTGHGp	292.74	107772	561.2502	-1.8	2
85	1.28E+06	1213.5205	TLLGHWGHnp	217.1	77586	405.5145	0.9	3
89	1.27E+06	1240.5564	TLLGWPTHhp	235.33	86062	414.525	-2.6	3
86	1.24E+06	1226.6096	TLLGKNKHhp	73.91	12600	409.8766	-1.2	3



91	1.23E+06	1200.5251	TLLGTGWHP	220.66	79218	401.1815	-2.1	3
88	1.22E+06	1160.624	TLLGKKKHGp	70.48	11114	387.8813	-1.8	3
88	1.21E+06	1200.4736	TLLGDNDHHP	190.67	65399	401.1644	-1.9	3
84	1.21E+06	1181.5154	TLLGNPNHHP	155.92	48543	394.8448	-2.3	3
87	1.19E+06	1128.4888	TLLGTGNHHP	156.08	48622	377.1693	-2.4	3
83	1.18E+06	1168.52	TLLGTNPHP	163.27	52087	390.5131	-2.1	3
92	1.17E+06	1240.5564	TLLGWTPHP	231.68	84352	414.525	-2.6	3
84	1.16E+06	1154.5771	TLLGAVKHHP	167.96	54348	385.8659	-1	3
83	1.13E+06	1160.5876	TLLGKNKHAp	105.58	26225	387.8692	-1.6	3
91	1.11E+06	1258.5308	TLLGWDHHP	248.75	92012	420.5171	-1	3
80	1.11E+06	1033.4768	TLLGTVGHGp	260.31	96742	517.7456	-0.1	2
82	1.10E+06	1133.5405	TLLGATKHHP	160.01	50500	378.8535	-1.5	3
90	1.08E+06	1213.5205	TLLGWNHHP	219.46	78666	405.5137	-1.1	3
89	1.07E+06	1155.4998	TLLGNNAHHP	149.1	45326	386.1733	-1.5	3
84	1.07E+06	1151.5049	TLLGTHHHGp	113.65	30001	384.8417	-1.4	3
83	1.07E+06	1114.571	TLLGLPKHGp	224.47	80966	372.5306	-0.9	3
88	1.00E+06	1159.5198	TLLGTTTHHP	168.37	54556	387.5128	-2.6	3
91	9.93E+05	1200.4736	TLLGNDDHHP	189.89	65019	401.1645	-1.7	3
93	9.92E+05	1142.5044	TLLGNTAHP	163.9	52382	381.8411	-2.7	3
89	9.78E+05	1155.5249	TLLGTPHP	171.11	55925	386.1817	-1.4	3
86	9.65E+05	1211.5986	TLLGNLKHHP	186.54	63436	404.8728	-1.8	3
86	9.64E+05	1241.5881	TLLGWKAHP	173.28	56993	414.8693	-1.6	3
88	9.58E+05	1253.5518	TLLGWPNHHP	228.98	83074	418.8576	-0.6	3
87	9.58E+05	1253.5518	TLLGNPWHP	229.51	83326	418.8572	-1.5	3
88	9.54E+05	1267.6038	TLLGKWPHP	192.63	66330	423.5414	-1.2	3
84	9.49E+05	1227.5935	TLLGDKKHHP	86.42	17557	410.2041	-2.5	3
91	9.45E+05	1254.5356	TLLGPDWHP	260.34	96777	419.1854	-1	3
85	9.31E+05	1244.5876	TLLGNWPHKp	236.27	86477	415.8695	-0.8	3
86	9.26E+05	1169.5518	TLLGNAKHHP	119.88	32919	390.8573	-1.4	3
87	9.22E+05	1231.5925	TLLGWTPHKp	231.8	84421	411.5375	-1.6	3
89	9.15E+05	1213.5205	TLLGWNHHP	222	79833	405.5135	-1.5	3
88	9.13E+05	1205.5405	TLLGWDKHGp	245.54	90614	402.8538	-0.7	3
88	8.85E+05	1213.5205	TLLGWNHHGp	221.07	79409	405.5136	-1.3	3
85	8.84E+05	1106.4932	TLLGTATHHP	235.01	85915	554.2534	-0.8	2
89	8.73E+05	1257.5466	TLLGNTWHP	221.67	79682	420.1888	-1.6	3
88	8.70E+05	1140.5251	TLLGNVAHHP	213.85	76068	381.1823	0	3
91	8.52E+05	1161.5505	TLLGWAKHGp	241.71	88927	388.1902	-1.5	3
88	8.44E+05	1142.5044	TLLGANTHHP	160.56	50771	381.8414	-1.8	3
87	8.31E+05	1171.5198	TLLGDTVHP	223.38	80464	391.5132	-1.6	3
86	8.20E+05	1223.4895	TLLGDDHHHP	153.92	47591	408.8365	-1.5	3
91	7.97E+05	1227.5725	TLLGWKHHP	177.35	58966	410.1979	-0.6	3
80	7.85E+05	1128.4888	TLLGNHTHGp	161.03	51003	377.1696	-1.5	3

88	7.83E+05	1196.5303	TLLGPGWHP	221.03	79391	399.8501	-1.5	3
95	7.69E+05	1272.51	TLLGWDHHP	264	98087	425.1768	-1.2	3
94	7.69E+05	1272.51	TLLGWDHDP	264.53	98276	425.1765	-1.8	3
90	7.58E+05	1214.5044	TLLGGDWHP	253.37	94010	405.8417	-1	3
89	7.48E+05	1156.4836	TLLGANDHHP	175.82	58235	386.5011	-2	3
87	7.45E+05	1069.488	TLLGGVGHHP	189.05	64625	357.5024	-2.4	3
89	7.38E+05	1027.4412	TLLGGGHHGP	153.68	47479	343.4872	-1.3	3
89	7.32E+05	1164.5	TLLGHNGHHP	98.83	23079	389.1732	-1.9	3
83	7.31E+05	1156.52	TLLGHDKHGP	123.53	34594	386.5133	-1.7	3
86	7.11E+05	1253.5518	TLLGNWPHHP	236.78	86713	418.8575	-1	3
86	7.07E+05	1221.5466	TLLGLHDHHP	177.27	58928	408.1893	-0.4	3
84	7.06E+05	1191.5613	TLLGWKTHGP	226.09	81719	398.1934	-2.4	3
85	7.03E+05	1156.4836	TLLGNADHHP	180.93	60722	579.2484	-1.2	2
83	6.97E+05	1227.5725	TLLGKWHHP	180.3	60411	410.1978	-0.8	3
89	6.88E+05	1200.5251	TLLGWTTHHP	231.61	84320	401.1819	-1.1	3
91	6.78E+05	1254.5356	TLLGDPWHP	261.2	97073	419.1858	0	3
86	6.77E+05	1223.4895	TLLGDHDHHP	150.97	46202	408.8362	-2.1	3
82	6.74E+05	1138.5095	TLLGANPHHP	165.43	53142	380.5095	-2.4	3
88	6.45E+05	1293.5579	TLLGHNWHP	170.9	55825	432.193	-0.5	3
89	6.25E+05	1141.5093	TLLGAVDHHP	218.31	78112	381.5099	-1.1	3
84	6.25E+05	1213.5205	TLLGWHNHGP	218.56	78230	405.5139	-0.5	3
89	6.22E+05	1212.5828	TLLGKLDHHP	191.91	65999	405.2008	-1.9	3
81	6.18E+05	1128.4888	TLLGNTTHHP	158.28	49679	377.1698	-1	3
82	6.15E+05	1076.4614	TLLGWGGHGP	288.75	106480	539.2385	0.9	2
84	5.99E+05	1141.4841	TLLGNNGHHP	144.81	43398	381.5018	-0.4	3
92	5.97E+05	1163.5413	TLLGLGHHHP	164.14	52501	388.8538	-1.6	3
90	5.88E+05	1172.5151	TLLGNTTHHP	163.29	52098	391.8449	-1.9	3
89	5.82E+05	1151.5049	TLLGGHTHHP	107.92	27332	384.8414	-2.2	3
81	5.80E+05	1183.531	TLLGNNVHHP	167.98	54361	395.517	-1.6	3
88	5.78E+05	1228.52	TLLGWADHHP	259.65	96474	410.5139	0	3
86	5.76E+05	1118.5659	TLLGTKLHGP	216.5	77286	373.862	-1.5	3
87	5.68E+05	1088.5554	TLLGLAKHGP	226.55	81958	363.8585	-1.6	3
87	5.67E+05	1129.5457	TLLGAPKHHP	162.24	51611	377.5217	-2.2	3
88	5.58E+05	1248.5828	TLLGWKTHHP	217.83	77886	417.2007	-2	3
91	5.55E+05	1257.5466	TLLGWNTTHHP	224.08	80782	420.1892	-0.6	3
91	5.45E+05	1294.542	TLLGWDHHP	205.53	72394	432.5206	-1.5	3
86	5.45E+05	1270.542	TLLGNNVHHP	215.89	76998	424.5209	-0.9	3
88	5.41E+05	1154.5771	TLLGLKGHHP	159.18	50111	385.8657	-1.6	3
82	5.39E+05	1235.5735	TLLGKHNHHP	80.7	15234	412.8647	-0.9	3
89	5.37E+05	1269.6194	TLLGKVVHHP	213.28	75840	424.2133	-1.1	3
88	5.37E+05	1210.5459	TLLGAPWHP	232.32	84655	404.5222	-0.9	3
92	5.18E+05	1244.5515	TLLGTTWHP	225.44	81419	415.8576	-0.4	3

91	5.18E+05	1342.5784	TLLGWNWHHp	275.04	101909	448.5329	-1.1	3
87	5.18E+05	1221.5466	TLLGHLDHHP	189.71	64935	408.189	-1.1	3
89	5.10E+05	1257.5466	TLLGTWNHHP	224.97	81205	420.1891	-1	3
84	5.10E+05	1176.5464	TLLGNTKHNp	151.89	46625	393.1883	-2.8	3
86	5.09E+05	1156.499	TLLGGHWHGp	230.73	83907	386.5068	-0.5	3
89	5.07E+05	1269.6194	TLLGKVVHHP	211.61	75063	424.2131	-1.6	3
86	5.05E+05	1221.5466	TLLGDLHHP	205.04	72119	408.1884	-2.6	3
86	4.90E+05	1159.5562	TLLGTKPHNp	160.6	50790	387.5255	-1.4	3
84	4.90E+05	1253.5518	TLLGPNWHHP	224.6	81032	418.8572	-1.6	3
92	4.89E+05	1132.5452	TLLGDLKHGp	257.9	95805	378.522	-0.9	3
85	4.88E+05	1253.5518	TLLGWNPHHP	230.66	83878	418.8579	0	3
84	4.70E+05	1141.5093	TLLGDLGHHP	256.18	95094	381.51	-0.9	3
82	4.59E+05	1129.5093	TLLGTTAHP	169.26	55005	377.5096	-2	3
87	4.58E+05	1294.542	TLLGDWHHP	211.24	74897	432.5209	-0.9	3
86	4.53E+05	1141.5093	TLLGGDLHHP	227.68	82467	381.51	-1	3
82	4.47E+05	1248.5828	TLLGNTKHWP	237.51	87050	417.2013	-0.5	3
91	4.35E+05	1143.4885	TLLGDATHHP	217.19	77606	382.1695	-1.6	3
88	4.33E+05	1163.551	TLLGTTKHNp	155.79	48493	388.8573	-0.8	3
85	4.26E+05	1118.5659	TLLGLTKHGp	219.21	78526	373.8625	-0.2	3
88	4.18E+05	1250.552	TLLGAWHHP	189.62	64933	417.8572	-1.7	3
86	4.04E+05	1142.5044	TLLGTANHP	161.97	51451	381.8414	-1.8	3
89	3.95E+05	1214.5044	TLLGDWGHHP	257.76	95746	405.8416	-1.3	3
85	3.91E+05	1250.552	TLLGWAHHP	173.57	57141	417.8575	-1.1	3
90	3.86E+05	1165.582	TLLGPLVHHP	241.04	88631	389.5344	-0.7	3
88	3.83E+05	1205.5405	TLLGWKDHP	238.66	87575	402.8539	-0.7	3
84	3.75E+05	1269.583	TLLGLNVHHP	265.63	98676	424.2011	-1.1	3
92	3.70E+05	1228.52	TLLGDAWHP	271.39	100674	410.5132	-1.9	3
88	3.68E+05	1141.5093	TLLGDVAHP	254.14	94287	381.5095	-2.4	3
89	3.65E+05	1074.5398	TLLGLKGHP	207.92	73418	359.1868	-1.2	3
85	3.58E+05	1258.5308	TLLGDWTHHP	278.67	103159	420.5172	-0.7	3
82	3.55E+05	1165.5205	TLLGATHHP	118.63	32341	389.5132	-2.2	3
87	3.52E+05	1185.5354	TLLGDTLHP	257.06	95437	396.1852	-1.4	3
81	3.46E+05	1164.5251	TLLGPNPHHP	170.27	55516	389.1821	-0.5	3
82	3.31E+05	1142.5044	TLLGATNHP	161.4	51183	381.8413	-2.1	3
90	3.23E+05	1199.5146	TLLGDLHHP	264.16	98150	400.845	-1.2	3
82	3.23E+05	1272.51	TLLGDVHHP	269.85	100149	425.1766	-1.7	3
87	3.22E+05	1200.5251	TLLGTWGHHP	230.45	83779	401.1823	0	3
81	3.15E+05	1163.5413	TLLGLHGHP	156.94	49036	388.8531	-3.3	3
89	3.03E+05	1325.5881	TLLGWPHHP	285.39	105366	442.8696	-0.9	3
84	3.00E+05	1127.4937	TLLGHVDHP	207.95	73425	376.838	-1.4	3
84	2.96E+05	1219.5562	TLLGWAKHP	258.65	96110	407.5256	-1	3
90	2.91E+05	1269.583	TLLGWNLHP	257.44	95599	424.2015	-0.2	3

83	2.88E+05	1133.5405	TLLGTKAHNp	153.99	47624	378.8531	-2.6	3
84	2.87E+05	1221.5466	TLLGDHLHhp	183.44	61950	408.1886	-2.2	3
82	2.87E+05	1228.52	TLLGWDAAHhp	257.27	95525	410.5136	-0.9	3
90	2.84E+05	1183.5562	TLLGLDVHhp	266.87	99110	395.5255	-1.2	3
84	2.79E+05	1274.6711	TLLGLKKHwp	219.11	78522	425.8972	-1	3
83	2.77E+05	1199.5146	TLLGLDHHDp	255.86	94964	400.8453	-0.4	3
94	2.74E+05	1214.5408	TLLGWATHHp	270.53	100389	405.8538	-0.9	3
87	2.70E+05	1227.5361	TLLGAWNHHp	229.33	83242	410.1855	-1.1	3
91	2.64E+05	1125.5508	TLLGAVVHHp	243.92	89925	376.1904	-1.2	3
88	2.59E+05	1159.5925	TLLGKVVHhp	227.65	82454	387.5375	-1.5	3
89	2.57E+05	1199.5146	TLLGLDDHHp	252.24	93500	400.8448	-1.8	3
88	2.54E+05	1270.542	TLLGNWHHhp	219.94	78877	424.521	-0.6	3
88	2.53E+05	1104.5139	TLLGAVTHNp	276.93	102593	553.264	-0.4	2
82	2.44E+05	1236.5615	TLLGPWPHhp	238.45	87482	413.1934	-2.6	3
90	2.22E+05	1203.5977	TLLGWKLHGp	264.04	98102	402.2058	-1.8	3
93	2.19E+05	1272.51	TLLGDWDHhp	268.64	99751	425.177	-0.7	3
90	2.19E+05	1228.52	TLLGDWAHhp	281.07	103956	410.5138	-0.3	3
93	2.10E+05	1076.4827	TLLGAATHNp	239.44	87921	539.2476	-2	2
88	2.10E+05	1256.5879	TLLGLTWHHp	263.58	97939	419.869	-2	3
82	2.05E+05	1256.5515	TLLGWDVHhp	274.89	101853	419.8571	-1.5	3
85	2.02E+05	1212.5615	TLLGWGLHhp	273.94	101550	405.1941	-1	3
90	2.01E+05	1102.571	TLLGKVVHGp	233.75	85310	368.5306	-0.9	3
81	1.98E+05	1249.5667	TLLGWKTHDp	234.08	85469	417.5288	-1.6	3
93	1.94E+05	1125.5508	TLLGLVGHHp	245.85	90751	376.1904	-1.2	3
89	1.91E+05	1097.5193	TLLGALHHGp	228.94	83061	366.8464	-1.8	3
80	1.85E+05	1143.4521	TLLGGHDHDp	206.57	72834	382.1579	-0.2	3
88	1.81E+05	1233.6082	TLLGTKVHwp	260.09	96650	412.2092	-1.9	3
87	1.81E+05	1253.5518	TLLGPWHHNp	223.13	80350	418.8578	-0.1	3
90	1.73E+05	1125.5505	TLLGLVHHGp	242.2	89148	376.1902	-1.6	3
85	1.65E+05	1198.5459	TLLGHVVGp	256.81	95344	400.5222	-1	3
88	1.62E+05	1183.5562	TLLGLVDHHp	261.78	97298	395.5254	-1.6	3
85	1.58E+05	1257.5466	TLLGHTWHNp	219.55	78690	420.1893	-0.4	3
90	1.43E+05	1272.51	TLLGDWHHDp	266.69	99053	425.1761	-2.7	3
82	1.35E+05	1103.5188	TLLGLDVHGp	369.86	122788	552.7668	0.3	2
86	1.30E+05	1083.5037	TLLGHLGHGp	220.27	79040	362.175	-0.3	3
88	1.17E+05	1185.5354	TLLGDLTHHp	284.25	105008	396.1852	-1.5	3
82	1.16E+05	1292.5991	TLLGWHLHhp	211.75	75128	431.8735	-0.3	3
82	1.07E+05	1152.5615	TLLGAKPHhp	112.96	29672	385.194	-1.3	3

## 6.4. Acknowledgements

The work was supported by the Defense Advanced Research Projects Agency (Award No. 023504-001) for B. L. P. and the MIT/NIGMS Biotechnology Training Program (T32GM008334-28) for A. R. L. We thank Dr. Zachary P. Gates and Mr. Chi Zhang for insightful discussions throughout the course of the work.

## 6.5. References

- (1) Adams, A.; Cowper, B.; Morgan, R.; Premdjee, B.; Caddick, S.; Macmillan, D. *Angew. Chem. Int. Ed.* **2013**, *52*, 13062–13066.
- (2) Huang, Y.; Fang, G.; Liu, L. *Natl. Sci. Rev.* **2016**, *3*, 107–116.
- (3) Thom, J.; Anderson, D.; McGregor, J.; Cotton, G. *Bioconjug. Chem.* **2011**, *22*, 1017–1020.
- (4) Fang, G.; Li, Y.; Shen, F.; Huang, Y.; Li, J.; Lin, Y.; Cui, H.; Liu, L. *Angew. Chem. Int. Ed.* **2011**, *50*, 7645–7649.
- (5) Fang, G.; Wang, J.; Liu, L. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 10347–10350.
- (6) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science*. **1994**, *266*, 776–778.
- (7) Ramil, C.; An, P.; Yu, Z.; Lin, Q. *J. Am. Chem. Soc.* **2016**, *138*, 5499–5502.
- (8) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M.; Van Voorhis, T.; Pentelute, B. *Nat. Chem.* **2015**, *8*, 120–128.
- (9) Collins, K.; Gensch, T.; Glorius, F. *Nat. Chem.* **2014**, *6*, 859–871.
- (10) Bantscheff, M.; Lemeer, S.; Savitski, M.; Kuster, B. *Anal. Bioanal. Chem.* **2012**, *404*, 939–965.
- (11) Lam, K.; Lebl, M.; Krchnak, V. *Chem. Rev.* **1997**, *97*, 411–418.
- (12) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342.
- (13) Ma, B.; Zhang, K.; Liang, C. *J. Comput. Syst. Sci.* **2005**, *70*, 418–430.
- (14) Kang, J.; Richardson, J.; Macmillan, D. *ChemComm* **2009**, *6*, 407–409.
- (15) Pannifer, A.; Yian Wong, T.; Schwarzenbacher, R.; Renatus, M.; Petosa, C.; Bienkowska, J.; Lacy, D.; Collier, J.; Park, S.; Leppla, S.; Hanna, P.; Liddington, R. *Nature* **2001**, *414*, 229–233.
- (16) Van der Walt, S.; Colbert, C.; Varoquaux, G. *Comput. Sci. Eng.* **2011**, *13*, 22–30.
- (17) McKinney, W. *Proceedings of the 9th Python in Science Conference.* **2010**, 51–56.
- (18) <http://www.scipy.org/>
- (19) Hunter, J. *Comput. Sci. Eng.* **2007**, *9*, 90–95.