Synthetic Tools for the Preparation of Modified Histones

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

John C. Shimko, M.S.

Graduate Program in The Ohio State Biochemistry Program

The Ohio State University

2011

Dissertation Committee:

Jennifer J. Ottesen, Advisor

Ross E. Dalbey

Venkat Gopalan

Zhengrong R. Wu

Thomas G. Wilson
Abstract

The eukaryotic genome is organized into nucleosomes consisting of 146 bp of DNA wrapped around an octamer of histone proteins, two copies each of H2A, H2B, H3, and H4. Post-translational modification (PTM) of histones perturbs nucleosome structure and dynamics thereby regulating important biological processes including transcription, replication and DNA repair. To understand these processes, we have developed synthetic tools for the preparation of homogenous samples of modified histones.

We established a novel ligation-desulfurization system for the preparation of modified histone H4 proteins enabling acetylation and phosphorylation adjacent to the C-terminus of the protein while retaining the native protein sequence. Modified H4 proteins were reconstituted into nucleosomes and nucleosome arrays. The effect of lysine 77 and 79 acetylation on nucleosome array stability was assessed. Furthermore, we demonstrated that the simultaneous incorporation of eight acetylated lysines within the LRS and dyad regions of the nucleosome does not significantly impact the structure or stability of the nucleosome.

We introduced the total synthesis of histone H3 acetylated at lysine 56 (H3-K56ac) via the chemoselective condensation of three peptide segments prepared by manual solid...
phase peptide synthesis (SPPS) with Boc chemistry. Non-native cysteine residues at sites of ligation were converted to native alanine residues by free-radical-desulfurization. Reconstitution and characterization of H3-K56ac semi-synthetic nucleosomes revealed that the introduced modification increases DNA site-accessibility and protein invasion of the nucleosome.

We developed a reversible protection strategy for 3,4-diaminobenzoic acid (Dbz), a Fmoc compatible thioester precursor, which allows for the use of highly activating conditions and acetyl capping during Fmoc peptide synthesis. Dbz protection maximizes product yield while minimizing the formation of deletion products. Further, we demonstrated the novel site-specific derivitization of the unsubstituted Dbz amine with biophysical probes including biotin and fluorophores.

Finally, we adapted our initial total synthesis strategy for modified histone H3 to allow for the automated synthesis of a library of ~45 residue-modified peptide segments in high yield and purity by Fmoc chemistry. The combination of automated peptide synthesis protocols with sequential NCL allows for a widely accessible strategy for the combinatorial preparation of differentially-modified histone proteins suitable for biophysical characterization.
Dedication

Lissa, Suzka, Serena, Nathaniel and Zen

Order not defined.
Acknowledgments

First and foremost I must thank my advisor Dr. Jen Ottesen for her support and guidance throughout my graduate school career. Her limitless understanding and creativity have played a key role in allowing me to reach this point, and her mentorship has helped to shape the scientist that I have become.

I would like to thank all members of the Ottesen and Poirier labs for their continuous support, assistance, and guidance throughout my graduate career. In particular, I would like to thank Michelle Ferdinand (not Taylor) for helping me maintain my high levels of sanity and zen, keeping me caffeinated, and reminding me that there is a world outside, even if our lab does not have windows. Thank you to C.J. Howard for having good taste in movies and for always keeping the project moving forward. Thank you to Sarah Dreher for never tiring of spotting one more MALDI plate and for the cookie drawer. Thank you to Santosh Mahto for teaching me about cricket. A final thank you to Bryon Mahler for helping me get in to “writing mode”, and for never getting tired of long, slow walks around campus that were never as quiet as he would have liked.

A special thank you to the makers of Civilization, Batman, and Naruto; without you I would have graduated last quarter.
Vita

1990.................................................................Mechanicsburg Area Senior High
1990-1993 ......................................................University of Pennsylvania
1993-1995 ..........................................................Juniata College
1995.................................................................B.S. Chemistry, Juniata College
1995-1998 ...........................................................Bucknell University
1999-2001 ..........................................................Research Associate, Message Pharmaceuticals
2005-2011 ..........................................................Graduate Teaching and Research Associate,
The Ohio State University
2008.................................................................M.S. Biochemistry, The Ohio State University

Publications


**Fields of Study**

**Major Field:** The Ohio State Biochemistry Program
Table of Contents

Abstract ............................................................................................................................. ii
Dedication .......................................................................................................................... iv
Acknowledgments .............................................................................................................. v
Vita ....................................................................................................................................... vi
Publications ...................................................................................................................... vi
Fields of Study ................................................................................................................. vii
Table of Contents ............................................................................................................ viii
List of Tables .................................................................................................................... xv
List of Figures .................................................................................................................. xvi
List of Abbreviations ....................................................................................................... xx

Chapter 1: Introduction ................................................................................................. 1

Genetic Substitutions as Modification Mimics ............................................................... 3
Cysteine Alkylation to Introduce Modification Mimics ................................................ 6
Codon Suppression .......................................................................................................... 9
Peptide Synthesis and Native Chemical Ligation .......................................................... 13
Sequential NCL for the Total Synthesis of H3-K56ac(R40C,S96C,C110A)........... 76
Total Chemical Synthesis of H3-K56ac(R40C,S96C,C110A)........................................ 82
Reduction of Oxidized Methionine in H3-K56ac(R40C,S96C,C110A)................. 84
Characterization of Synthetic H3-K56ac(R40C,S96C,C110A)........................... 86
Total Synthesis of H3-K56ac(C110A)................................................................. 89
Characterization of Synthetic H3-K56ac(C110A)............................................. 99
Conclusions............................................................................................................. 105
Acknowledgements............................................................................................... 108

Chapter 4: Reversible Protection and Substitution of Dbz................................. 109
Introduction............................................................................................................ 109
Experimental Methods.......................................................................................... 113
General Automated Peptide Synthesis Protocols.................................................. 113
Synthesis of H4N and H4C peptides on unprotected Dbz resin.......................... 113
Synthesis of LYRAGA and LYRAGF Peptides....................................................... 114
Synthesis of H4N and H4C Peptides on Fmoc-Dbz(Alloc) Resin....................... 114
Determination of Amino Acid Loading on Dbz(Alloc) Resin............................... 115
Alloc Protection of Amino Acid Substituted Resin............................................. 116
Synthesis of 3’-Fmoc_Dbz).................................................................................... 117
Preparation of 3’-Fmoc_Dbz(Alloc) Resin......................................................... 117
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alloc Deprotection</td>
<td>118</td>
</tr>
<tr>
<td>Nbz conversion</td>
<td>118</td>
</tr>
<tr>
<td>Automated Nbz Conversion</td>
<td>119</td>
</tr>
<tr>
<td>Peptide Cleavage</td>
<td>119</td>
</tr>
<tr>
<td>Thioester Preparation</td>
<td>120</td>
</tr>
<tr>
<td>Determining Rate of Fmoc-Gly Coupling to Boc-LYRGA-Dbz</td>
<td>120</td>
</tr>
<tr>
<td>Determining Rate of Biotin Coupling to Boc-LYRGA-Dbz</td>
<td>121</td>
</tr>
<tr>
<td>Synthesis of H3T-Dbz(Gly) and H3T-Dbz(Biotin) Peptides</td>
<td>122</td>
</tr>
<tr>
<td>Antibody Pulldown</td>
<td>122</td>
</tr>
<tr>
<td>Synthesis of TAMRA-coupled H3T Peptides</td>
<td>123</td>
</tr>
<tr>
<td>Synthesis of TAMRA-H3T-Dbz(biotin) Peptide</td>
<td>124</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>125</td>
</tr>
<tr>
<td>Automated Alloc Deprotection and Nbz Conversion</td>
<td>133</td>
</tr>
<tr>
<td>Exploiting the Activity of Dbz for Peptide Modification</td>
<td>138</td>
</tr>
<tr>
<td>Conclusions</td>
<td>155</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>157</td>
</tr>
<tr>
<td>Chapter 5: Preparation of Synthetic Modified Histones by Fmoc SPPS</td>
<td>158</td>
</tr>
<tr>
<td>Introduction</td>
<td>158</td>
</tr>
<tr>
<td>Experimental Methods</td>
<td>162</td>
</tr>
</tbody>
</table>

xii
Synthesis of H3 C-terminal Peptides (Residues 91-135) ......................................... 162
Alloc Protection of Amino Acid Substituted Resin...................................................... 163
Synthesis of H3 Middle Peptide Segments (Residues 47-90) ................................... 163
Synthesis of H3 N-terminal Peptide Segments (Residues 1-46) ............................ 164
Removal of the Alloc Protecting Group ................................................................... 165
Conversion of Peptide-Dbz to the Nbz Form ............................................................... 166
Side Chain Deprotection and Cleavage from Resin .................................................... 166
Peptide Purification ..................................................................................................... 167
Ligation to Prepare Synthetic Truncated H3(A47C,S96C) Residues 47-135 ......... 167
Ligation to Prepare Full Length H3(A47C,A91C,C110A) ........................................... 168
Desulfurization of Synthetic H3 Proteins ................................................................. 169
Results and Discussion ......................................................................................... 171
Synthesis of C-terminal Peptide Segments ............................................................... 173
Preparation of Amino Acid Substituted Dbz(Alloc) Resins ..................................... 183
Synthesis of Middle Peptide Segments ................................................................. 185
Synthesis of N-terminal Peptide Segments .............................................................. 188
Total Chemical Synthesis of Histone H3 Proteins .................................................. 192
Total Synthesis of H3-K56ac,K115ac,K122ac(A47C,A91C) ...................................... 195
Conclusions ............................................................................................................ 198
Acknowledgements ................................................................. 201

Chapter 6: Conclusions .............................................................. 202

References .................................................................................. 210

Appendix A: MALDI-TOF MS of LYRAGX-Dbz-R Products ................. 223

Appendix B: RP-HPLC analysis of amino acid loading on Dbz(Alloc) resin .......... 224
List of Tables

Table 1. Synthetic peptides for EPL of histone H4 ................................................................. 36
Table 2. Peptides for the total synthesis of H3-K56ac ............................................................ 76
Table 3. Sequence of prepared peptides .............................................................................. 126
Table 4. Percent loading of amino acids on Dbz(Alloc) resin ............................................. 138
Table 5. Sequence of prepared synthetic peptide segments .............................................. 172
List of Figures

Figure 1. Structure of the nucleosome ................................................................. 2
Figure 2. Common genetic substitutions mimic PTMs ........................................... 4
Figure 3. Cysteine alkylation to introduce modification mimics ............................ 7
Figure 4. Scheme for introducing modified amino acids by codon suppression ....... 9
Figure 5. Native chemical ligation scheme .......................................................... 14
Figure 6. EPL of a synthetic N-terminal peptide thioester to a recombinant protein ...... 15
Figure 7. Ligation-desulfurization scheme to generate modified H4 ...................... 23
Figure 8. Characterization of modified H4p peptides .......................................... 37
Figure 9. SDS-PAGE analysis of EPL to yield modified H4(A76C) ......................... 38
Figure 10. MALDI-TOF MS analysis of modified H4(A76C) desulfurization .......... 40
Figure 11. RP-HPLC characterization of modified H4 proteins ............................. 42
Figure 12. Characterization of partially purified H4-pT80 .................................. 43
Figure 13. SDS-PAGE of semi-synthetic modified histone octamers .................... 45
Figure 14. MALDI-TOF MS of semi-synthetic modified histone octamers ............. 46
Figure 15. Crude H4-K77ac,K79ac and H4-pT80 reconstitutions .......................... 47
Figure 16. Sucrose gradient purification of semi-synthetic nucleosomes ................ 49
Figure 17. Sucrose gradient-purified modified semi-synthetic nucleosomes .......... 50
Figure 18. Sucrose gradient-based purification of 17-mer nucleosome arrays .......... 52
Figure 19. AFM image of WT and H4-K77ac,K79ac nucleosome array .......................... 53
Figure 20. Force extension of modified semi-synthetic nucleosome arrays .................. 54
Figure 21. Reconstitution of highly acetylated nucleosomes .................................. 57
Figure 22. Scheme for total synthesis by sequential NCL ........................................ 63
Figure 23. Total synthesis scheme for H3-K56ac(R40C,S96C,C110A) ....................... 75
Figure 24. Characterization of purified peptides N1 and C1 .................................... 78
Figure 25. Comparison of RP-HPLC solvent systems for partial M1 synthesis ........ 79
Figure 26. MALDI-TOF MS of partial M1 peptide (H3-K56ac residues 52-95) .......... 80
Figure 27. Characterization of purified peptide M1 .................................................. 82
Figure 28. Characterization of ligation product M1C1 ............................................. 83
Figure 29. Reduction of oxidized Met in H3-K56ac(R40C,S96C,C110A) .................. 84
Figure 30. Analysis of purified H3-K56ac(R40C,S96C,C110A) .............................. 85
Figure 31. Native PAGE analysis of crude nucleosome reconstitutions ..................... 87
Figure 32. FRET efficiency of H3-K56ac(R40C,S96C,C110A) .............................. 89
Figure 33. Total synthesis scheme for the preparation of H3-K56ac(C110A) .......... 90
Figure 34. Analysis of purified peptides N2, M2 and C2 ........................................ 92
Figure 35. Characterization of M2C2 ligation product .......................................... 95
Figure 36. Crude ligation mixture for H3-K56ac(A47C,A91C,C110A) .................... 97
Figure 37. Desulfurization to yield H3-K56ac ....................................................... 97
Figure 38. Characterization of H3-K56ac(C110A) ................................................. 99
Figure 39. Native PAGE of sucrose gradient purified nucleosomes ....................... 100
Figure 40. FRET efficiency of purified nucleosomes ............................................ 101
Figure 63. Scheme for combinatorial library of modified histone H3 ..................... 161
Figure 64. Characterization of peptide C3 ............................................................... 174
Figure 65. Characterization of peptide C3N ............................................................. 175
Figure 66. Characterization of peptide C4 purification ........................................... 176
Figure 67. Crude synthesis of M4M on Rink and Dbz resins .................................. 178
Figure 68. RP-HPLC analysis of crude M4M-Dbz and M5M-Dbz ......................... 179
Figure 69. Characterization of pure M4M-Nbz and M5M-Nbz peptides ................ 180
Figure 70. Pseudoproline dipeptide decreases formation of truncated product ....... 182
Figure 71. Schemes for the preparation of Met-Dbz(Alloc) resin ............................ 184
Figure 72. Characterization of M4N and M5N purification ..................................... 186
Figure 73. Analysis of M6-Dbz(Alloc) crude peptide mixture .............................. 188
Figure 74. Characterization of N3-Nbz purification .............................................. 190
Figure 75. Crude syntheses of N-terminal peptides on Val-Dbz(Alloc) resin .......... 191
Figure 76. Total synthesis scheme for the preparation of histone H3 ................. 192
Figure 77. MALDI-TOF MS analysis of synthetic H3 proteins ............................. 194
Figure 78. Characterization of H3-K56ac,K115ac,K122ac first ligation product ....... 196
Figure 79. Characterization of H3-K56ac,K115ac,K122ac(A47C, A91C) .................. 197
List of Abbreviations

Amino acids are referred to by the appropriate one or three letter code. Histones H2A, H2B, H3, and H4 are referred to as H2A, H2B, H3 and H4 respectively.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-Cl-HOBt</td>
<td>6-chloro-1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>Ac</td>
<td>Acetylated</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>Alloc</td>
<td>Allyloxycarbonyl</td>
</tr>
<tr>
<td>BME</td>
<td>$\beta$-mercaptoethanol</td>
</tr>
<tr>
<td>Boc</td>
<td>t-butoxycarbonyl</td>
</tr>
<tr>
<td>Bom</td>
<td>Benzyloxymethyl</td>
</tr>
<tr>
<td>BZA</td>
<td>Benzamidine</td>
</tr>
<tr>
<td>Dbz</td>
<td>3,4-diaminobenzoic acid</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DIEA</td>
<td>N,N-diisopropylethylamine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMS</td>
<td>Dimethylsulfide</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDT</td>
<td>Ethane dithiol</td>
</tr>
<tr>
<td>xx</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>EDTA</td>
<td>2,2',2''',2'''-(ethane-1,2-diyl)hydrotritiro)tetraacetic acid</td>
</tr>
<tr>
<td>EPL</td>
<td>Expressed protein ligation</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-fluoronylmethoxycarbonyl</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>Gdn</td>
<td>Guanidine</td>
</tr>
<tr>
<td>HATU</td>
<td>2-(7-aza-1-H-benzotriazol-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphate</td>
</tr>
<tr>
<td>HBTU</td>
<td>2-(1-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate</td>
</tr>
<tr>
<td>HCTU</td>
<td>2-(6-chloro-1-H-benzotriazol-1-yl-1,1,3,3-tetramethyluronium hexafluorophosphate</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>HEPES</td>
<td>2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid</td>
</tr>
<tr>
<td>HO</td>
<td>Histone octamer</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix assisted laser desorption ionization-time of flight</td>
</tr>
<tr>
<td>MBHA</td>
<td>4-methylbenzhydrylamine hydrochloride salt resin</td>
</tr>
<tr>
<td>Me</td>
<td>Methylated</td>
</tr>
<tr>
<td>Met(O)</td>
<td>Oxidized methionine</td>
</tr>
<tr>
<td>MESNA</td>
<td>Mercaptoethysulfonate sodium salt</td>
</tr>
<tr>
<td>MPAA</td>
<td>4-mercaptophenylacetic acid</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>Nbz</td>
<td>N-acyl-benzimidazolinone</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NCL</td>
<td>Native chemical ligation</td>
</tr>
<tr>
<td>Nle</td>
<td>Norleucine</td>
</tr>
<tr>
<td>NMP</td>
<td>N-methylpyrrolidone</td>
</tr>
<tr>
<td>Nuc</td>
<td>Nucleosome</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PAM</td>
<td>4-hydroxymethyl-phenylacetamidomethyl resin</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>Reverse phase-high performance liquid chromatography</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecylsulfate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SPPS</td>
<td>Solid phase peptide synthesis</td>
</tr>
<tr>
<td>TAMRA</td>
<td>6-(tetramethylrhodamine-5-(and-6-carboxamido) hexanoic acid</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td>TEMED</td>
<td>1,2-bis(dimethylamino)ethane</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>TFE</td>
<td>Trifluoroethanol</td>
</tr>
<tr>
<td>Thz</td>
<td>Thiazolidine</td>
</tr>
<tr>
<td>TIS</td>
<td>Triisopropylsilane</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-hydroxymethyl-propane-1,3-diol</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VA-044US</td>
<td>2,2’-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

The nucleosome is the basic unit of chromatin consisting of ~146 bp of DNA wrapped around an octamer histone proteins, two copies each of the core histone proteins: H2A, H2B, H3 and H4 (Fig. 1).\textsuperscript{1,2} The structure of the nucleosome modulates access to the DNA and attenuates important cellular processes including transcription, replication, and DNA repair. Nucleosome structure is dynamic and highly regulated to allow access or protection of DNA as necessitated by the requirements of the cell. Nucleosome structural dynamics are primarily regulated by ATP-dependent chromatin remodeling proteins and post-translational modification (PTM) of the histone octamer.\textsuperscript{3,4}

Histone acetylation was the first histone PTM to be identified.\textsuperscript{5,6} In the years since, an ever increasing number and class of modifications have been identified. Initially, modifications were identified in bulk chromatin and PTMs that were studied were high abundance modifications primarily located within the unstructured histone tails. With the advent of more sensitive mass spectrometry tools as well as the enrichment of specific sub-categories of bulk chromatin, less ubiquitous modifications, which may themselves be cell cycle dependent, have been identified throughout the four core histone proteins.\textsuperscript{7}
It is the goal of the presented work to develop synthetic tools for the preparation of an array of homogeneous modified histone proteins. A successful strategy would be robust enough to yield quantities of modified protein suitable for characterization and be flexible enough to allow for the site-specific incorporation of a multitude of modifications throughout the protein sequence. The direct modification of recombinant histones by histone modifying enzymes is unlikely to yield a homogeneous sample due to incomplete and/or extraneous modification. Additionally, a specific modifying enzyme has not been
characterized for all identified histone modifications which necessarily limits the subset of modifications which may be introduced.

This chapter describes a variety of valuable chemical and biological techniques that have successfully been employed for the preparation of homogeneous modified histones, including genetic substitution, cysteine alkylation, codon suppression and native chemical ligation (NCL). A summary of the various techniques and their role in the preparation of modified histone proteins is presented below. Despite their ability to prepare homogeneous samples, each technique has a unique set of limitations which render them incompatible with the site-specific incorporation of various PTMs throughout the entire sequence of a single histone protein.

Genetic Substitutions as Modification Mimics

Genetic amino acid substitution is a common method for assessing the possible impact of PTMs in vivo. Amino acid substitution may eliminate dynamic post-translational modification and may partially mimic the structure or electrostatic properties of the modified amino acid of interest (Fig. 2). The substitution of lysine with glutamine is a commonly studied genetic manipulation. Glutamine substitution mimics the charge neutralization of lysine acetylation, but does not completely replicate the hydrogen bonding or steric properties of acetylation. The semi-conservative lysine to arginine
Figure 2. Common genetic substitutions mimic PTMs

Commonly employed genetic substitutions of acetylysine (top) and phosphorylated serine or threonine (bottom).

substitution may be employed to eliminate the possibility of acetylation while maintaining the native positive charge, but does lead to the introduction of significant steric bulk at the site of substitution. Glutamate is typically used as a substitute to mimic
serine or threonine phosphorylation. This substitution does introduce negative charge and increases steric bulk, but it lacks the shape and charge distribution of the native modification.

Although they are not structurally identical, substitutions have been a valuable tool in assessing the possible impact of histone PTMs in vivo. The Lys to Gln substitution has been extensively studied in histones⁸⁻¹⁰ and the observed impact has been validated in several cases by in vitro characterization of histones exhibiting the correlating modifications, particularly in the unstructured histone tails.¹¹ However, glutamine substitution does not accurately describe the effects of acetylation in all cases.¹²

The observed discrepancies between amino acid substitutions and the corresponding modifications emphasize the need to develop tools for the generation of homogeneous samples which incorporate the cognate modifications to be studied. Furthermore, given the variety of PTMs that exist in histones, the appearance of multiple modifications on a single protein, and the increasing evidence of specific cross-talk between modifications,¹³ the generation of homogeneous samples must be robust enough to allow for the site-specific incorporation of multiple cognate modifications within a single histone protein or octamer. A multitude of chemical and biological tools have previously been reported for the generation of homogeneous samples.
Cysteine Alkylation to Introduce Modification Mimics

Relative to other natural amino acids, the cysteine side chain sulfhydryl has unique chemical reactivity, which may be exploited to allow for site-specific modification. Cysteine reactivity is commonly used to introduce moieties which allow for direct characterization of the modified molecule or complex. Few native cysteine residues are present in the core histone proteins. Histone H3-C110 is the only native cysteine residue found in the core histones of the commonly studied *Xenopus laevis*,¹⁴ and as such, the site-specific incorporation of Cys residues allows for a controlled introduction of chemical moieties. In the case of histones, cysteine modification has been employed for the introduction of various biochemical tools including spin labels,¹⁵,¹⁶ fluorophores,¹⁷,¹⁸ cross-linkers,¹⁹,²⁰ and footprinting reagents¹⁰ which have allowed for biophysical characterization of nucleosomes.

The unique chemistry of the cysteine sulfhydryl group has recently been exploited to introduce modification mimics that resemble native modifications more closely than amino acid substitution. Initially developed by Shokat and co-workers, analogs of methylated lysines have been generated by the reaction of genetically encoded cysteine side chains with N-methyl-aminoethylhalide derivatives throughout the sequence (Fig. 3).²¹ The resultant methylated-lysine mimics differ structurally from their natural counterparts by replacement of the γ methylene with sulfur, which adds steric bulk and increases the length of the side chain by ~0.28Å, while perturbing the pKa by -1.1 unit.
Antibodies for methylated histones H3 and H4 have shown specific recognition to the corresponding lysine mimic in all tested cases except for the H3-K9me\(^2\) antibody.\(^{21}\)

Cysteine alkylation to mimic methylated lysines has been exploited to study the role of H3K36 mono-, di- and tri-methylation in regulation of the Rpd3S histone deacetylase complex with nucleosomes,\(^{22}\) as well as to study the binding of ING4 with H3-K4me\(^3\) which is required to suppress growth of T47D breast cancer cells.\(^{23}\)

Figure 3. Cysteine alkylation to introduce modification mimics

Common pathways for the chemical modification of cysteine to prepare modified lysine mimics.\(^{21,24}\)
Cysteine alkylation is not suitable for the preparation of the thioether analog of acetyllysine. However, alkylation of cysteine with methylthiocarbonyl-aziridine to generate an acetylated lysine mimic suitable for biochemical characterization has recently been reported.\textsuperscript{24} In addition to the methylene to sulfur substitution at the $\gamma$ position, the acetylation mimic also introduces a thiomethyl group in place of the native methyl moiety in order to minimize observed side products. Synthesized peptides containing the acetyl-lysine mimic were recognized by antibodies and proteins specific for acetylated H3 and H4 tails. Preliminarily, the acetylation mimic appears to be resistant to histone deacetylases (HDACs) and therefore may prove useful for biochemical assay of nucleosomes in the presences of HDACs.\textsuperscript{24}

Cysteine alkylation to generate mimics of PTMs is dependent on the successful overexpression of recombinant proteins which contain a cysteine substitution at the site of interest. Consequently, an important limitation of this technique is that only a single class of modification may be introduced within a single molecule due to the requirement for complete alkylation of all cysteine residues necessary to yield a homogeneous sample. Highly modified histone proteins often contain varying degrees of methylation and acetylation within a single molecule, and therefore cysteine alkylation is limited as a tool for the study of common biologically relevant modified states.
Codon Suppression

Codon suppression is a powerful tool for the introduction of modified amino acids into an expressed protein. The technique originally developed by Schultz and co-workers utilizes the rare Amber stop codon to genetically control the introduction of modified amino acids in a sequence dependent manner (Fig. 4). Bacterial overexpression coupled with codon suppression allows for the rapid generation of appreciable amounts of desired protein in high yield and purity limited only by the physical properties of the protein being generated.

Figure 4. Scheme for introducing modified amino acids by codon suppression
Codon suppression involves the evolution of an amino-acyl tRNA synthetase which specifically and efficiently acylates tRNA which recognizes a non-abundant codon. The anber codon (tRNA_{CUA}) is the most commonly employed target. The evolved synthetase/tRNA pair must not only be specific for acylation of the desired modified amino acid, but must be orthogonal to the acylation machinery of the host cell to eliminate mis-acylation. Once evolved, the synthetase/tRNA pair may be expressed within a host cell to allow site-specific incorporation of the modified amino acid in a sequence dependent manner. An additional consideration is the bioavailability and stability of the modified amino acid which is to be incorporated.

Although a large number of non-native amino acids have been introduced into proteins through codon suppression,\(^{26}\) the evolution of synthetase/tRNA pairs for the introduction of modified amino acids commonly found in histones (acetylation, methylation and phosphorylation) have only recently been developed. A pyrrolysyl-tRNA synthetase/tRNA_{CUA} pair for the site-specific introduction of acetylated lysine was recently reported.\(^ {27}\) This codon suppression technique was used to generate milligram quantities of the H3-K56ac protein, which were characterized for the impact of K56 acetylation on DNA accessibility and chromatin remodeling enzymes.

Mono-methylation of lysine residues is abundant in histone proteins, and a method for the introduction of mono-methylated lysines through codon suppression has recently been reported.\(^ {28}\) Development of a specific pyrrolysyl-tRNA synthetase/tRNA_{CUA} pair was
hampered by the inability of the active site to reliably differentiate between the desired mono-methylated species and unmodified lysine due to their similar size and shape. The authors therefore developed a synthetase / tRNA pair for the site specific introduction of $N^\varepsilon$-Boc-$N^\varepsilon$-methyl-$l$-lysine, which may subsequently be converted to the desired mono-methylated species by TFA deprotection. More recently a similar system has been reported which introduces monomethylated lysine with $N^\varepsilon$-allyl protection.\textsuperscript{29}

To date, a synthetase/tRNA pair has not been developed for the direct incorporation of native di- and tri-methylated lysines, however an alternative codon suppression scheme has been reported which involves the oxidative elimination of genetically encoded phenylselenocysteine to yield dehydroalanine.\textsuperscript{30} Dehydroalanine residues may be specifically targeted for thiol addition to introduce a mimic which contains the methylene to sulfur mutation at the $\beta$ carbon.\textsuperscript{31} The conversion of dehydroalanine to acetylated- and methylated- lysine mimics has been conducted.\textsuperscript{32} Additionally, this chemistry may be employed to introduce a large variety of modifications from lipids to sugar molecules, limited only by the availability of a compatible thiol moiety. It is important to note that thiol addition to dehydroalanine is not stereospecific and results in a racemic mixture of the desired modified side chain.

Phosphorylation is a particularly important PTM that has been detected throughout the nucleosome. Recently, a codon suppression system for the site-specific incorporation of phosphoserine was introduced.\textsuperscript{33} This development is particularly remarkable due to the
need for the co-expression of a phosphoserine importer as well as the highly labile nature of the modification within the cellular environment necessitating genetic deletion of endogenous phosphatases. Previously, hydrolysis stable phosphorylation mimics had been engineered for site-specific incorporation,\textsuperscript{34} but these mimics do not faithfully reproduce the steric or electrostatic properties of phosphorylation.

Codon suppression as a tool for generating modified histones suitable for biophysical characterization is limited by the types of modifications that may be introduced. Additionally, the incorporation of multiple classes of modifications into a single protein requires the evolution of synthetase/tRNA pairs which are orthogonal to each other as well as the native cellular expression system. The successful development of orthogonal pairs has recently been reported to allow for the introduction of different classes of modified amino acids within the same molecule.\textsuperscript{35} However, the present generation acetyl- and methyl-lysine codon suppression techniques are derived from the same pyrolysyl-tRNA system and are therefore not amenable to co-expression. In addition, while reasonable quantities of singly substituted proteins are readily available, there are practical limitations on the introduction of multiple instances of the same modification in a single protein. Despite its limitations, codon suppression is a valuable tool for generating mg quantities of homogeneous modified proteins.
Peptide Synthesis and Native Chemical Ligation

The development of solid phase peptide synthesis (SPPS) has allowed for the production of modified peptides which contain an array of PTMs. The power of synthesis lies in the ability to control the introduction of chemical moieties at each step within the limits of chemistry and orthogonal protection. It is therefore possible to synthesize a peptide which contains multiple modified or non-native amino acids throughout the sequence. Synthetic peptides may be used to directly study modification states within a given region of a protein and assess sequence specific binding interactions. Several seminal studies have used synthetic peptides to determine the specificity of histone-modifying enzymes\textsuperscript{23, 36-42} as well as the structural determination of binding interactions.\textsuperscript{43-46} The primary limitation of SPPS lies in the <100% reactivity at each step which necessitates a practical limit of ~50 residues as the size of the synthetic unit.

Native chemical ligation (NCL) is the chemoselective condensation of a 1,2-aminothiol, typically an N-terminal cysteine, with a C-terminal α-thioester to yield a native peptide amide bond (Fig. 5).\textsuperscript{47} Following ligation, the 1,2-aminothiol is retained C-terminal to the newly formed bond in its original state. NCL is a powerful tool that significantly increases the size limitation inherent in classical SPPS and may be exploited to allow for the generation of species not readily accessible to recombinant techniques.
Expressed protein ligation (EPL) expands on the concept of NCL by allowing for the combination of a synthetic peptide with an overexpressed protein segment resulting in a semi-synthetic unit that may not be amenable to direct synthesis. The recombinant segment may be engineered to contain an N-terminal Cys or expressed as a fusion protein with an introduced intein domain that would allow for the generation of a C-terminal thioester on the protein of interest. EPL has greatly expanded the role of peptide chemistry in studying biological relevant species.

Figure 5. Native chemical ligation scheme
EPL has been of particular value in establishing the role of PTMs in the context of histone proteins and their impact on nucleosome. Initial studies involved the incorporation of a modified synthetic thioester peptide ligated to a C-terminal recombinant protein with an N-terminal Cys residue (Fig. 6).

The first reported use of EPL to generate a modified histone was the preparation of the H3-pS10,T32C protein. The synthetic N-terminal peptide fragment bearing phosphorylated serine and a C-terminal thioester was combined with recombinant truncated histone with an N-terminal Cys to yield the full length protein. Semi-synthetic
H3-pS10,T32C protein was subsequently reconstituted into nucleosome arrays and characterized. EPL has also been used to generate modified H3, H4,\textsuperscript{50-52} and H2B proteins.\textsuperscript{53}

The Ottesen laboratory has recently reported an EPL strategy for the incorporation of modified amino acids within the C-terminal domain of histone H3 utilizing the native cysteine at position 110.\textsuperscript{12,54} In this method, modified amino acids are incorporated by SPPS into the terminal peptide which bears an N-terminal cysteine. The N-terminal protein segment is expressed as a fusion protein upstream of a functional intein which upon thiolysis will yield a thioester. Ligation of the prepared segments yields the desired modified protein. A similar strategy has also been reported for the preparation of modified histone H2B.\textsuperscript{55}

The quantity and variety of modifications that may be introduced into a specific region of a protein by NCL and EPL is limited primarily by the peptide chemistry used to prepare the modified segment. As such, the practical limit of ligation is the ability to generate homogenous synthetic peptide in a yield suitable for ligation as well as the careful selection of ligation sites that do not directly impact the properties of the protein product. The inherent limitation of synthetic peptide size restricts the usefulness of EPL to generating histones modified adjacent to the termini and makes study of modifications in the core folded domain impractical.
Developing Synthetic Tools for the Preparation of Modified Histones

The presented summary of modern biochemical tools for the preparation of site-specifically modified histone proteins highlights the strengths and weaknesses of the various techniques. As evidenced by the work described, each of the presented methods is amenable to the generation of a specific and limited subset of modified histones. We have sought to develop a more robust method which would allow for the preparation of homogeneous site-specifically modified histone proteins which is not limited by number, class or location of modifications that may be introduced.

In Chapter 2, we describe the adaptation of the successful EPL strategy for the semi-synthesis of modified H3 proteins\textsuperscript{12, 54} to a novel method for the preparation of modified H4 proteins.\textsuperscript{56} The lack of a native cysteine residue in H4 necessitated the use of a ligation-desulfurization strategy. The combination of EPL with free-radical desulfurization generates full length modified H4 protein exploiting a native alanine residue as the site of ligation. We have successfully used this strategy to prepare H4 proteins site-specifically acetylated or phosphorylated adjacent to the C-terminus.

EPL is a valuable technique for the preparation of modified proteins when the modification of interest is positioned near the protein terminus. However, in order to study histone PTMs that lie within the folded core distal to the termini, or proteins that
are modified in different regions throughout the sequence, a simple one-step ligation strategy such as EPL may not be sufficient. In Chapter 3, we discuss our development of a sequential native chemical ligation technique which allows for the preparation of histone proteins site-specifically modified throughout the sequence. In this strategy we employ the ordered condensation of three peptide segments in two independent ligation steps to yield the final full length modified product. In our initial study, we used peptide segments prepared manually by Boc in situ neutralization\textsuperscript{57} for the preparation of histone H3 acetylated at lysine 56 (H3-K56ac).

The success of our total synthesis strategy for the preparation of histone H3-K56ac was tempered by the low synthetic yields of the middle peptide segment and the incompatibility of Boc chemistry with automated synthetic protocols. Additionally, the acid labile nature of phosphorylated and tri-methylated amino acids renders these modifications inaccessible to preparation by Boc synthesis limiting the general applicability of our initial strategy. Because of these limitations, we sought to develop synthetic approaches compatible with accessible Fmoc SPPS techniques for the preparation of peptide segments.

The first challenge was the preparation of histone peptide thioesters by Fmoc SPPS. In Chapter 4 we discuss the development of a reversible protection strategy for the Fmoc compatible thioester precursor 3,4-diaminobenzoic acid (Dbz) which was recently introduced by the Dawson lab.\textsuperscript{58} Protection of the free amine of Dbz during solid phase
peptide synthesis allows the use of optimized synthetic chemistry, including highly reactive amino acid activating agents as well as acetyl capping. The use of optimized chemistry maximizes yield and minimizes deletion products which are difficult to resolve from the desired product during purification. Further, we develop a method to exploit the free amine of a peptide-Dbz species for the introduction of biochemical and biophysical probes including fluorophores and biotin.

As a final step in our effort to develop a general use method for the routine synthesis of highly modified histone proteins, we have combined the sequential native chemical ligation strategy first presented in Chapter 3 with the Dbz(Alloc) thioester precursor detailed in Chapter 4 to allow for the automated synthesis of peptide segments by Fmoc chemistry suitable for use in the total chemical synthesis of histone H3. In addition to the benefits of automation and accessibility to acid sensitive modification, Fmoc is the preferred chemistry of the majority of synthesis labs and further increases the general applicability of the introduced sequential NCL total synthesis scheme.
Chapter 2: Preparation of Modified Histone H4 by Expressed Protein Ligation

Introduction

The LRS region of the nucleosome has been defined functionally by the identification of a cluster of modifications which perturb rDNA silencing. The LRS region includes an L1L2 motif formed by the interaction of the H3 L1 and H4 L2 loops adjacent to the DNA-histone interface. Several histone modifications in the LRS region have been identified by mass spectrometry. H4-K77 and H4-K79 can be simultaneously acetylated in bovine histones and H4-T80 is reported to be phosphorylated in low abundance. H4-K77ac has been identified as a low-abundance modification associated with ASF1 chaperone complexes. Furthermore, yeast mutational studies suggest a complex role for H4-K77ac, H4-K79ac, and H4-pT80 in transcriptional silencing.

H4-K91 within the αC helix near the C-terminus has been reported as a site of acetylation and ubiquitination at an important histone-histone interface. This residue is thought to stabilize the histone octamer via a conserved salt bridge across the interface of the H3-H4 tetramer and the H2A-H2B dimer. Studies of the H4-K91A mutant in yeast suggest a strong role for this modification in DNA repair and in silencing at telomeres,
likely due to an impact on chromatin assembly resulting in disruptions of heterochromatin structure.9

Our laboratory has previously developed an expressed protein ligation (EPL) strategy for the incorporation of modified amino acids into the C-terminal domain of histone H3 utilizing the native cysteine at position 110.12, 54 In light of the recent identification of post-translational modification within the L2 and αC regions of the C-terminal portion of histone H4,60, 61 we believed a similar approach would be appropriate for the preparation of modified histone H4.

Histone H4 does not contain a native Cys residue and a standard EPL approach would consequently lead to the introduction of a non-native residue at the ligation site. Proteins without a native cysteine and are not tolerant of the introduction of non-native Cys residues, were originally not good candidates for semi-synthesis by EPL. Desulfurization of cysteine to yield alanine, a more commonly observed amino acid, vastly expands the possible sites of ligation within a protein target. Several different chemical methods have been introduced for the post-ligation desulfurization of cysteine.66, 67 Selenocysteine has also been shown to function as a suitable nucleophile for NCL. Following ligation, the introduced selenocysteine may be converted to alanine by deselenization in a reducing environment with the addition of excess TCEP.68
As initially described, the requirements of NCL are an N-terminal Cys and a C-terminal α-thioester;\textsuperscript{47} however, the N-terminal moiety is not limited to Cys, but may be any 1,2-aminothiol which allows for the initial nucleophilic attack and subsequent S to N acyl shift necessary for amide bond formation. The development of various thiol-substituted amino acid derivatives has vastly increased the possible sites of ligation. To date, the successful use of 1,2-aminothiol derivatives of valine,\textsuperscript{69} phenylalanine,\textsuperscript{70} leucine,\textsuperscript{71} lysine,\textsuperscript{72} and proline\textsuperscript{73} for NCL have been reported. In particular, the use of the valine and proline allows for ligation with nucleophilic thiols that are themselves poor thioester targets for NCL.\textsuperscript{74}

Post-ligation desulfurization of a non-native cysteine to alanine allows for the identification of alanine as a potential ligation site. Herein, we present a ligation-desulfurization strategy which would allow for the preparation of H4 proteins modified proximal to the C-terminus (Fig. 7).

We describe the successful application of our ligation-desulfurization strategy to prepare semi-synthetic proteins modified within the C-terminal domain of histone H4. We have introduced acetylation of lysines 77 and 79 and phosphorylation of threonine 80 within the LRS region as well as acetylation of lysine 91 in the αC helix at the dimer-tetramer interface. We subsequently reconstituted the modified semi-synthetic proteins into octamers, nucleosomes, and in the case of H4-K77ac,K79ac arrays of 17 discrete nucleosomes positioned on a single piece of DNA.
Finally, we have combined histone H3 acetylated within the nucleosome dyad at lysines 115 and 122 and histone H4 acetylated at lysines 77 and 79 within the LRS region to prepare highly modified semi-synthetic nucleosomes. Given the two-fold symmetry of the octamer, each prepared nucleosome includes eight distinct sites of acetylation at the histone-DNA interface. This work represents the first direct study of a nucleosome bearing this large number of modifications within the folded core of the nucleosome, and further demonstrates the power of EPL as a tool for the preparation of highly modified homogeneous samples suitable for characterization.
Experimental Methods

H4(1-75) Thioester Preparation

Forward GGTGGTGGAGGTGAAGGAGATACATATGTCGTGG and reverse GTGGTGGCTCTCAGCAGTGGTTCTGGTAACAGC DNA primers were used to amplify by PCR the DNA sequence encoding histone H4 residues 1-75 from a plasmid encoding wild-type histone H4 and to introduce Nde1 and Sap1 restriction sites suitable for proper insertion into the pTXB1 plasmid (NEB). Following digestion of the PCR product and pTXB1 plasmid, the resulting mixtures were ligated, and transformed into E. coli DH5α cells. Plasmids were purified from individual colonies and screened by sequencing to identify plasmids which contain the desired insert.

The H4(1-75)-intein-CBD fusion protein was expressed in E. coli BL21(DE3) cells and accumulated in inclusion bodies upon induction with 0.2 mM IPTG. Induced cells were incubated for 4 hours at 37 °C and the resultant cell mass was pelleted by centrifugation. Pelleted cells were resuspended with 25 mM HEPES pH 7.5, 1 M NaCl and 1 mM EDTA. PMSF was added to a final concentration of 1 mM and the resuspended cells were frozen with liquid nitrogen and stored at -80 °C.

The frozen cell reuspension was thawed and cells were lysed by passing through a French press three to four times. The inclusion bodies which contain the H4(1-75)-intein-CBD
fusion protein were separated from soluble proteins by centrifugation at 20,000 xg for 15 minutes at 4 °C. The precipitated inclusion bodies were washed twice with buffer (25 mM HEPES pH 7.5, 1 M NaCl, 1 mM EDTA, 1 % Triton-X) and pelleted with centrifugation at 10,000 xg for 10 minutes at 4 °C. Pellet was washed once in the same buffer without Triton-X and the remaining insoluble proteins were pelleted by centrifugation as described. The pellet was solubilized with DMSO for 30 minutes with mixing and resuspended with 6 M urea, 25 mM HEPES pH 7.5, 500 mM NaCl, 1 mM EDTA. The reuspension was vortexted for 1 hour at 25 °C and centrifuged to pellet insoluble material. The supernatant was decanted and dialyzed against 6 M urea, 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA.

Dialyzed protein was purified by ion-exchange over a Sepharose fast flow (SP-FF) ion-exchange column (GE Healthcare). The crude sample was loaded directly onto the column and the partially purified H4(1-75)-intein-CBD protein was eluted with 6 M urea, 25 mM HEPES pH 7.5, 500 mM NaCl, 1 mM EDTA. The intein was refolded by dialysis overnight into 25 mM HEPES pH 7.5, 1 M NaCl, 1 mM EDTA. Sample removed from the dialysis tubing was centrifuged to remove insoluble aggregates from the supernatant. If a significant amount of misfolded H4-intein-CBD was detected in the pellet, it was refolded as previously described.

Thiolysis of the soluble protein fraction was initiated by the addition of 100 mM MESNA and allowed to proceed for 18-24 hours at 4 °C. Once thiolysis was complete, the buffer
components were adjusted to reach final conditions of 6 M ultrapure guanidine, 25 mM HEPES pH 7.5, 1 M NaCl, 1 mM EDTA, 50 mM MESNA and concentrated to > 1 mg/mL H4(1-75) thioester as determined by UV-Vis. The crude thioester sample was flash frozen with liquid nitrogen and stored at -80 °C for use in ligation without further purification.

Determining Concentration of H4(1-75) Thioester by UV

The post-thiolysis H4(1-75) thioester mixture includes one equivalent of intein-CBD as well as residual intact H4(1-75)-intein-CBD. H4(1-75) thioester contains a single UV active tyrosine and the Abs$_{280}$ of this mixture is dominated by aromatic residues in the intein-CBD fusion domain. As such, the H4(1-75) thioester concentration may be determined by a combination of SDS-PAGE analysis of protein ratios in the crude thiolysis mixture and UV spectroscopy exploiting the intein-CBD domain extinction coefficient of 63260 cm$^{-1}$ M$^{-1}$.

Peptide Synthesis and Purification

All peptides (Table 1) were synthesized on Fmoc-Gly-Wang LL resin (Novabiochem) with automation on an Apex 396 synthesizer (Aapptec) with Fmoc compatible chemistry.
For a typical coupling cycle, resin was deprotected with 20% (v/v) piperidine in NMP and washed. Amino acids (4.4 eq.) were coupled for 1 hour with 4 eq. HBTU, 4 eq. HOBt, and 8 eq. DIEA in NMP. The N-terminal cysteine was introduced as Boc-thiazolidine-4-carboxylic acid (Bachem) to minimize racemization. Phosphorylated threonine (2.2 eq.) was manually introduced as the N-α-Fmoc-O-benzyl-L-phosphothreonine (Novabiochem) species with reduced excess to conserve amino acid.

Non-phosphorylated peptides were deprotected and cleaved from resin with 94% TFA, 2.5% H₂O, 2.5% EDT, 1% TIS. The H4p-pT80 peptide was deprotected and cleaved with reagent K: 82.5% TFA, 5% phenol, 5% thioanisole, 5% H₂O, 2.5% EDT. Crude cleavage mixtures were concentrated by evaporation and precipitated with cold anhydrous diethyl ether. Following centrifugation, the pellet was washed with ether, resuspended in 50% acetonitrile / H₂O with 0.1% TFA and lyophilized.

Lyophilized peptide was resuspended in 40% acetonitrile / 60% H₂O and ring opening was initiated with the addition of methoxylamine HCl to a final concentration of 0.4 M and the pH was verified to be ~4. Peptide may also be resuspended in non-organic conditions such as 3 M ultrapure guanidine, 25 mM HEPES, 1 M NaCl, 1 mM EDTA and 0.4 M methoxylamine HCl at pH 4. The reaction was allowed to proceed at 25 °C with nutation until full conversion to the unmasked cysteine was confirmed by MALDI-TOF MS. Ring opening typically was complete within 4-6 hours.
All peptides were purified by RP-HPLC over a Supelco Discovery Bio Wide Pore C18 preparatory column with a gradient of 23-36% acetonitrile / 0.1% TFA. Individual fractions were assayed by MALDI-TOF MS and fractions deemed to be > 95% pure were pooled. The purity and identity of the pool was confirmed by RP-HPLC and MALDI-TOF MS, and the final pool was lyophilized. After purification, 20 mg of H4p-K77ac,K79ac ([M+H]⁺ observed 3102.1 m/z; expected 3103.6 m/z), 35 mg H4p-K91ac ([M+H]⁺ observed 3061.0 m/z; expected 3061.6 m/z), and 18 mg H4p-pT80 ([M+H]⁺ observed 3099.6 m/z; expected 3099.4 m/z) peptides were obtained.

**Ligation to Generate H4(A76C) Semi-Synthetic Modified Proteins**

An excess of lyophilized and purified H4(76-102) modified peptide (5-10 eq.) was directly resuspended with crude H4(1-75) thioester. In a typical reaction, 6 mg of peptide was resuspended with 1.2 mg of thioester in 6 M ultrapure guanidine, 25 mM HEPES pH 7.5, 1 M NaCl, 1 mM EDTA, 50 mM MESNA and ligation proceeded with nutation at 25 °C. Ligation progress was monitored by SDS-PAGE until no further product formation was observed and was typically complete within 6 hours.
Desulfurization of Modified H4 Semi-Synthetic Proteins

Desulfurization of H4(A76C) modified proteins to yield the native alanine residue were completed by modifying previously reported protocols. Briefly, a solution of 1.2 M TCEP was directly added to the crude ligation mixture to a final concentration of 300 mM, while maintaining a minimum concentration of 4 M guanidine, and the desired pH of 7.5 was verified. The mixture was sparged with argon for 30 minutes. The free-radical initiator VA-044US (Wako Chemicals) was added to a final concentration of 10 mM and the resultant mixture was incubated at 42 °C. Desulfurization progress was monitored by MALDI-TOF MS and was typically complete within 4-6 hours.

RP-HPLC Purification of Semi-Synthetic Acetylated H4

Crude H4-K77ac,K79ac and H4-K91ac desulfurization mixtures were adjusted to 32% acetonitrile / 0.1% TFA and centrifuged to remove particulates. Precipitated material was assayed for the presence of full-length modified H4 and resuspended as necessary with a 32% acetonitrile, 4 M guanidine solution until all full-length H4 protein was solubilized. Supernatants were purified by RP-HPLC over a Supelco Discovery Bio Wide Pore C18 column with a gradient of 32-59% acetonitrile / 0.1% TFA. Individual fractions were assayed by MALDI-TOF MS and fractions deemed to be > 95% pure were pooled. The
purity and identity of the pool was confirmed by RP-HPLC and MALDI-TOF MS, and the final pool was lyophilized.

**Ion-Exchange Purification of H4-pT80**

Crude desulfurization mixture, typically the H4-pT80 mixture, was extensively dialyzed versus water to remove salts. The dialyzed sample, including all precipitation, was lyophilized and subsequently resuspended with 10 mM Tris pH 9.0, 7 M urea, 100 mM NaCl, 1.0 mM EDTA, 5 mM BME. Resuspended sample was purified over a TSKgel SP-5PW column (TOSOH Biosciences) with a gradient of 100-600 mM NaCl. Full-length H4(1-102) proteins do not fully resolve from the truncated H4(1-75) species under these conditions. Individual fractions were analyzed by SDS-PAGE and fractions determined to contain > 50% H4(1-102) over H4(1-75) were pooled and extensively dialyzed versus water. Following dialysis, the final pool was lyophilized and analyzed by RP-HPLC and SDS-PAGE to quantify the ratio of full-length to truncated H4. Product identity was verified by MALDI-TOF MS.

**Histone Octamer Refolding and Purification**

Recombinant histones were expressed and purified as previously described. Histones H2A, H2B, H3 and either H4, H4-K77ac,K79ac, H4-pT80 or H4-K91ac were combined
at equal molar ratios and refolded as previously described. Briefly, lyophilized histone aliquots were resuspended with 6 M ultrapure guanidine HCl, 20 mM Tris pH 7.5 and 10 mM DTT and loaded into an engineered 50 µl dialysis chamber which was placed in a larger dialysis tube containing 50 ml of the same buffer. The larger dialysis tube was extensively dialyzed against 10 mM Tris pH 7.5, 2 M NaCl, 1 mM EDTA, 5 mM BME at 4 °C for 32 hours with two buffer changes to allow for gentle dialysis of the histone sample. Refolded octamer was reclaimed from the dialysis chamber and stored at 4°C on ice.

Octamers were purified on a Superdex 200 10/300 GL column (GE Biosystems) with a running buffer of 10 mM Tris pH 7.5, 2 M NaCl, 1 mM EDTA and 5 mM BME. Purifications were typically done at 4 °C, but were also conducted at 25 °C. Fractions were assayed by SDS-PAGE, pooled, concentrated to > 1.3 mg/mL and stored at 4°C on ice. The purity of each octamer was confirmed by SDS-PAGE and mass spectrometry.

Reconstitution of Mononucleosomes

Mononucleosome reconstitutions were conducted as previously described. Briefly, refolded octamers were combined with 1.3 or 1.4 molar equivalents of mp2 DNA which consisted of a mixture of labeled (5’-Cy5, 3’-Cy3) and unlabeled DNA in 50 µl of 0.5x TE pH 8.0, 2 M NaCl and 1 mM BZA. The sample was loaded into an engineered 50
μl dialysis chamber which was placed in a larger dialysis tube containing 100 ml of the same buffer and dialyzed for a total of 24 hours with one buffer change against 5 mM Tris pH 8.0, 0.5 mM EDTA, 1 mM benzamidine at 4 °C. The 50 μl sample was extracted from the dialysis button and stored at 4 °C on ice. Reconstituted mononucleosomes were assayed for quality by native PAGE without further purification.

Sucrose Gradient-Purification of Mononucleosomes

Nucleosomes were purified on a 5-30% (w/v) sucrose gradient by ultracentrifugation. Gradients were prepared at 25 °C and chilled for 1 hour at 4 °C. Crude reconstituted nucleosome samples were applied to the top of the gradient. The samples were centrifuged at 41,000 rpm at 4 °C for 22-24 hours. Gradients were fractionated and individual fractions were assayed by native PAGE.

Preparation of H4-K77ac,K79ac Nucleosome Arrays

Unmodified and modified nucleosome arrays were prepared as previously described. Briefly, 3 μg of plasmid DNA containing 17 tandem repeats of the high affinity 601 nucleosome positioning sequence separated by 30 bp of linker DNA was combined with 15 μg of buffering DNA, 12 μg of purified histone octamer in an engineered dialysis
Arrays were reconstituted by salt double-dialysis against 5 mM Tris pH 8.0, 1 mM benzamidine, and 0.5 mM EDTA at 4 °C. Recovered arrays were purified on a 5–40% sucrose gradient by ultracentrifugation for 6.5 hours at 4 °C. Gradients were fractionated and individual fractions were initially analyzed on a 0.8% agarose gel. Fractions containing saturated nucleosome arrays were pooled and assayed by electrophoretic mobility shift on an acrylamide / agarose composite gel.

AFM Imaging of Nucleosome Arrays

AFM imaging of prepared nucleosome arrays was conducted by Marek Simon in the laboratory of Dr. Michael Poirier (OSU) as previously described. Briefly, 50 μL of poly-D-lysine at 10 ng/μL was deposited on freshly cleaved mica surface and incubated for 90 seconds, washed twice with 150 μL of ultra-pure water (Gibco), and dried under very gentle Nitrogen flow. 20 μL of 1 nM unmodified or H4-K77ac,K79ac purified nucleosome arrays in 2 mM Tris, pH 8.0, 0.2 mM EDTA were deposited on the treated mica surface, incubated for 5 min, washed with 200 μL ultra-pure water, and gently dried with nitrogen. Imaging was done in tapping mode with a scan rate of 1.7 Hz. The number of nucleosomes within each array was determined manually.
Single Molecule Force-Extension of Nucleosome Arrays

Force-extension measurements were conducted by Marek Simon in the laboratory of Dr. Michael Poirier (OSU) as previously described.\textsuperscript{56} Briefly, force-extension measurements were completed on single nucleosome arrays tethered to antidigoxigenin-coated cover glass slides and streptavidin-coated magnetic Dynabeads (Invitrogen) in lab-built flow cells on magnetic tweezers. We determined the number of nucleosomes within a single array by detecting the number of steps in the array extension as the force is increased from 7.5 to 29 pN. The force was initially increased to 7.5 pN over 20 seconds, then increased from 7.5 to 29 pN over 100 seconds, and then held at 29 pN for 130 seconds. The tension was relaxed to 0.1 pN and held at this force for 4 minutes. This cycle was repeated four additional times. All bead heights were measured relative to a nearby bead fixed to the surface.

The step number and size was determined by Matlab analysis of each time series of the nucleosome array length. The steps were detected by calculating the convolution of a 31-point step function with 31 data points centered about the time point of interest. This calculation resulted in a time series with a peak centered about each step. The number of peaks was determined and the center of each peak was located. The step size was then determined from the difference between the average of the 10 points before and after the step. The two adjacent points before and after the step were ignored when calculating step sizes.
Results and Discussion

Our laboratory has previously reported an EPL strategy for the generation of semi-synthetic histone H3 proteins which are precisely modified within the C-terminal domain. Typically, EPL introduces a cysteine residue at the site of ligation. However, histone H4 does not contain a native cysteine residue. To develop an analogous strategy for the introduction of post-translational modifications within the C-terminus of histone H4, it was essential to couple EPL with post-ligation desulfurization which would convert the ligation site cysteine to alanine, present in the native sequence of the protein.

There are several alanines proximal to the C-terminus of H4 that would be amenable for use in ligation. We have chosen the His75-Ala76 junction as a potential ligation site, as the histidine α-thioester has been shown to have good ligation kinetics. Additionally, Ala76 is positioned 27 residues away from the C-terminus minimizing the size of the synthetic peptide segment, yet allowing the introduction of modifications of interest.

The H4(1-75)-intein-CBD fusion protein was generated by standard molecular cloning techniques by introducing the coding sequence for histone H4 residues 1-75 upstream of the GyrA intein using the commercial available pTXB1 plasmid (NEB). The GyrA intein is preferable for this scheme as it reliably refolds following denaturation. Previous work has shown that recombinant histone-intein fusions may express as soluble
proteins, or they may accumulate in the insoluble fraction post-lysis and therefore must be denatured prior to isolation and purification.

All peptides were synthesized with standard Fmoc chemistry protocols with automation on an Apex 396 (Aapptec) synthesizer (Table 1). The N-terminal cysteine residue was introduced as thiazolidine to eliminate the possibility of cysteine racemization during coupling. Following peptide cleavage, thiazolidine was converted to cysteine by the addition of 0.4 M methoxylamine at a pH of 4. Peptides were purified by RP-HPLC and peptide purity and identity was verified by RP-HPLC and MALDI-TOF MS. Following purification, 20 mg of H4p-K77ac,K79ac ([M+H]+ observed 3102.1 m/z; expected 3103.6 m/z), 35 mg H4p-K91ac ([M+H]+ observed 3061.0 m/z; expected 3061.6 m/z), and 18 mg H4p-pT80 ([M+H]+ observed 3099.6 m/z; expected 3099.4 m/z) were obtained (Fig. 8).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4p-K77ac,K79ac</td>
<td>H4(76-102) A76Thz, K77ac,K79ac</td>
<td>(Thz)(Kac)R(Kac)TVTAMDVVYALKRQGRTLYGFGG</td>
</tr>
<tr>
<td>H4p-pT80</td>
<td>H4(76-102) A76Thz, pT80</td>
<td>(Thz)KRK(pT)VTAMDVYALKRQGRTLYGFGG</td>
</tr>
<tr>
<td>H4p-K91ac</td>
<td>H4(76-102) A76Thz, K91ac</td>
<td>(Thz)KRKTVTAMDVVYAL(Kac)RQRGRTLYGFGG</td>
</tr>
</tbody>
</table>

Table 1. Synthetic peptides for EPL of histone H4
Figure 8. Characterization of modified H4p peptides

RP-HPLC (left) and MALDI-TOF MS (right) analysis of pure H4p peptide pools. Peptides were resolved by RP-HPLC on a H2O / acetonitrile + 0.1% TFA gradient as follows: H4p-K77ac,K79ac (18-41%), H4p-K91ac (29-47%) and H4p-pT80 (25-39%). Expected and observed [M+H]+ for each peptide as described.
Lyophilized peptide was directly resuspended with the crude thiolysis mixture which includes the H4(1-75) thioester in a denaturing environment, typically 6 M guanidine. In a typical ligation, 6 mg of peptide was resuspended with 1.2 mg of H4(1-75) thioester and the ligation was allowed to proceed at 25 °C with nutation. Ligation progress was monitored by SDS-PAGE following TCA precipitation of individual time points and was typically complete within 6 hours (Fig. 9).

**Figure 9.** SDS-PAGE analysis of EPL to yield modified H4(A76C)

SDS-PAGE analysis of ligation progress for H4(1-75) thioester with corresponding modified H4p peptides as described. The (-) time point indicates crude thiolysis mixture prior to addition of peptide and the modified H4(A76C) product is identified by an arrow.
Desulfurization of Modified H4(A76C) Proteins

After ligation to yield the desired site-specifically modified H4 proteins, cysteine 76 at the site of ligation was converted to the native alanine residue by the Danishefsky free-radical-desulfurization method. Free-radical-desulfurization is a solution based method that is compatible with typical ligation conditions in aqueous buffers and quantitatively generates the desired product with minimal side reactions; the co-thiol required for ligation may serve as a proton source in the reaction. A comparative study of the Danishefsky method with more classical metal-catalyzed methods has been completed on modified histone H2B proteins generated by EPL. This study finds that the rate and yield of the free-radical-based method far exceeds that of the metal-based technique.

For desulfurization, the crude ligation mixture containing modified H4(A76C) was adjusted with the addition of TCEP to a minimum of 300 mM. The resulting mixture was sparged with argon for 30 minutes to remove oxygen, in order to prevent oxidation of methionine under reaction conditions. The desulfurization reaction was initiated with the addition of VA-044US and incubated at 42 °C. Reaction progress was monitored by MALDI-TOF MS and allowed to proceed until complete conversion of cysteine to alanine was the desulfurized product was observed, typically 6 hours (Fig. 10).
Figure 10. MALDI-TOF MS analysis of modified H4(A76C) desulfurization
Expected and observed [M+H]$^+$ for each peptide as described.
Purification of Modified H4 Proteins

Crude desulfurization mixtures were purified by reverse-phase or ion-exchange chromatography by HPLC. The variable physical properties of the prepared modified H4 proteins required the optimization of purification protocols for each species. Acetylated histones, H4-K77ac, K79ac and H4-K91ac, were typically purified by RP-HPLC. The H4-pT80 protein does not resolve from the intein-CBD fusion protein under reverse phase conditions and was therefore purified by ion-exchange chromatography with a NaCl gradient.

Crude H4-K77ac, K79ac and H4-K91ac desulfurization mixtures were brought to starting conditions with the addition of acetonitrile / 0.1% TFA and were purified by RP-HPLC over a Supelco C18 column following centrifugation to pellet insoluble material. To ensure maximum purification yield, pelleted material was resuspended and assessed for the presence of modified H4. If present, the resuspended pellet was subsequently purified under the same conditions. Collected RP-HPLC fractions were assayed for protein content, pooled, and lyophilized. The final purified product pool was assayed by RP-HPLC and MALDI-TOF MS to confirm identity and purity (Fig. 11). Typical synthetic yields were 0.6 mg H4-K77ac, K79ac from an initial 2.15 mg of thioester (21% yield) and 0.65 mg H4-K91ac from 1.25 mg thioester (38% yield).
Figure 11. RP-HPLC characterization of modified H4 proteins

RP-HPLC analysis of pure modified H4 protein pools. Proteins were resolved by RP-HPLC on a acetonitrile / H₂O + 0.1% TFA gradient as follows: H4-K77ac,K79ac (32-63%), H4-K91ac (32-59%), and H4-pT80 (36-54%). Purified H4-K77ac,K79ac (top) and H4-K91ac (middle) demonstrates minimal H4(1-75) truncation product which is abundant in the purified H4-pT80 sample (bottom).
The crude H4-pT80 desulfurization mixture was dialyzed extensively against water and 5 mM BME at 4 °C and lyophilized. The lyophilized mixture was resuspended with buffer and purified by ion-exchange chromatography. As we had previously observed in our study of semi-synthetic H3 proteins,\textsuperscript{12} full-length H4-pT80 did not fully resolve from the H4(1-75) truncated species. Individual fractions were analyzed by SDS-PAGE and fractions determined to contain > 50% H4-pT80 over H4(1-75) were pooled and extensively dialyzed versus water. Following dialysis, the final pool was lyophilized and analyzed by RP-HPLC (Fig. 11) and SDS-PAGE to quantify the ratio of full-length to truncated H4. Product identity was verified by MALDI-TOF MS (Fig. 12).

![Figure 12. Characterization of partially purified H4-pT80](image)

SDS-PAGE (left) and MALDI-TOF MS (right) analysis of partially purified H4-pT80 protein demonstrating the presence of the H4(1-75) truncation product. Expected and observed [M+H]\(^+\) for each species as described.
Refolding Histone Octamer with Semi-Synthetic Modified H4

Histone octamers were prepared from the semi-synthetic H4-K77ac,K79ac and H4-pT80 proteins as previously described. The lyophilized semi-synthetic H4-K77ac,K79ac was resuspended with equimolar amounts of recombinant H2A, H2B and H3. However, as the H4-pT80 protein does not completely resolve from truncated H4(1-75) during purification, H2A, H2B, and H3 recombinant proteins were added in equimolar amounts based on the total protein content of the partially purified sample not just the full length species.

Octamers were purified by size exclusion chromatography with compatible high salt buffer. It was determined that the octamer was stable at 25 °C for the length of a purification run, although purifications were preferentially conducted at 4 °C. Fractions were assayed by SDS-PAGE, pooled, concentrated to > 1.3 mg/mL and stored at 4°C on ice. The purity of concentrated octamer was confirmed by SDS-PAGE (Fig. 13) and identity was confirmed by MALDI-TOF MS (Fig. 14).

As discussed, the H4-pT80 protein does not completely resolve from truncated H4(1-75) during purification and it was necessary to closely monitor refoldings of this protein. We have previously reported that the C-terminal residues missing from the H3(1-109) truncated product form an important structural interface in the histone octamer and is
selectively excluded during octamer formation and purification.\textsuperscript{12} Similarly, the H4 C-terminal helix is located at the dimer-tetramer interface and is likely to be essential for octamer formation. However, it was necessary to verify that residual truncated H4(1-75) was not incorporated into refolded octamer. Although roughly 20\% of the crude protein mixture was truncated protein, following refolding, no truncation product was detectable in the octamer as determined by SDS-PAGE (Fig. 13) and verified by MALDI-TOF MS.
Although the H4(1-75) truncation product does not incorporate into fully reconstituted histone octamers, the αC helix does not form an important interface in the H3-H4 tetramer and is unlikely to be fully excluded from this structure. We therefore purify to homogeneity any protein preparation intended for refolding into H3-H4 tetramer.

![MALDI-TOF MS analysis of histone octamers reconstituted with H4-K77ac,K79ac (left) and H4-pT80 (right) demonstrating the absence of the H4(1-75) truncation species.](image)

**Figure 14. MALDI-TOF MS of semi-synthetic modified histone octamers**

MALDI-TOF MS analysis of histone octamers reconstituted with H4-K77ac,K79ac (left) and H4-pT80 (right) demonstrating the absence of the H4(1-75) truncation species.

**Preparation of Modified Semi-Synthetic Nucleosomes**

Refolded histone octamers incorporating recombinant wild-type H4 or semi-synthetic H4-K77ac,K79ac or H4-pT80 were combined with fluorescently labeled mp2-147 DNA and reconstituted into nucleosomes by double dialysis versus 0.5x TE buffer to gradually
remove salt. The resultant nucleosomes were analyzed by native PAGE (Fig. 15). The presence of residual DNA indicates incomplete nucleosome formation. Furthermore, two distinct nucleosome bands are observed due to variable positioning of the histone octamer on the DNA. \(^{81}\)

Figure 15. Crude H4-K77ac,K79ac and H4-pT80 reconstitutions

Native PAGE of reconstituted semi-synthetic nucleosomes with fluorescent imaging of all Cy5 containing species.
To generate a homogeneous sample of singly positioned DNA that would be amenable to biophysical characterization, the crude reconstitution mixture was fractionated on a sucrose gradient by ultracentrifugation. Individual fractions were assayed by native PAGE (Fig. 16). In this manner, the individual nucleosomes positions were resolved for the WT and H4-K77ac,K79ac reconstitutions, but not for the H4-pT80 nucleosomes (Fig. 17).

We have previously studied histone phosphorylation at the nucleosome dyad by characterizing semi-synthetic nucleosomes which contain phosphorylated threonine at position 118 of histone H3.\textsuperscript{54} In these studies, we observed various anomalous nucleosome structures, and classical nucleosomes that were formed were destabilized and required excess magnesium salt to maintain stability during prolonged storage and purification. It is important to note that no significant anomalous structures were observed in the H4-pT80 reconstitutions nucleosomes and no excess magnesium salt was necessary to maintain stability.
Figure 16. Sucrose gradient purification of semi-synthetic nucleosomes

Native PAGE analysis of individual fractions obtained by sucrose gradient-fractionation of crude H4-K77ac,K79ac (top) and H4-pT80 (bottom) nucleosome reconstitutions. Gels imaged by fluorescence detection of labeled DNA.
Figure 17. Sucrose gradient-purified modified semi-synthetic nucleosomes

Native PAGE analysis of sucrose gradient-purified nucleosomes demonstrating the isolation of primarily a single mononucleosome position as well as the exclusion of free DNA. Gel imaged by fluorescence detection of labeled DNA.

Study of H4-K77ac,K79ac Nucleosome Arrays

Following the successful preparation of mononucleosomes reconstituted with semi-synthetic H4-K77ac,K79ac, Marek Simon in the laboratory of Dr. Michael (OSU, Department of Physics) prepared nucleosome arrays. Arrays were reconstituted by combining DNA consisting of 17 individual 601 positioning sites separated by linker
DNA with unmodified wild-type or H4-K77ac,K79ac histone octamers and gently dialyzing the mixture to remove salt and allow array formation as previously reported. The crude array mixture was purified on a sucrose gradient to remove buffering DNA and partially assembled arrays. The prepared crude and purified arrays were assessed by native PAGE to determine the quality of reconstitution and purification (Fig. 18).

Sucrose gradient-purified arrays were immobilized on a fresh mica surface and assayed by AFM imaging. A total of 66 unmodified and 102 H4-K77ac,K79ac arrays were assessed representing a total of 1122 and 1734 nucleosome positioning sequences respectively (Fig. 19). Nucleosome saturation was calculated as the percentage of the total number of positioning sequences occupied as 98 ± 3% for unmodified and 95 ± 6% for H4-K77ac,K79ac arrays. Additionally, AFM demonstrated no gross morphological differences between the unmodified and H4-K77ac,K79ac containing arrays validating our observation on the mononucleosome level that H4-K77ac,K79ac does not demonstrably alter nucleosome structure.
Figure 18. Sucrose gradient-based purification of 17-mer nucleosome arrays

Native PAGE analysis of crude and sucrose gradient-purified nucleosome arrays for unmodified (left) and H4-K77ac.K79ac (right) containing nucleosomes. Gel imaged by fluorescence detection of labeled DNA. This work was completed by and the figure is courtesy of Marek Simon from the laboratory of Dr. Michael Poirier (OSU).56
To probe the impact of LRS acetylation on the stability of protein-DNA interactions within the nucleosome, semi-synthetic H4-K77ac,K79ac nucleosome arrays were immobilized and assessed by single molecule force extension (Fig. 20).[^56] Along with the
LRS modified arrays, we also probed the impact of acetylation at the dyad, by characterizing arrays reconstituted with semi-synthetic H3-K115ac,K122ac, and at the entry-exit region with H3-K56ac modified arrays (see Chapter 3).

Figure 20. Force extension of modified semi-synthetic nucleosome arrays

Schematic (top) of single molecule force extension experiment conducted on prepared semi-synthetic nucleosome arrays. This work was completed by and the figure is courtesy of Marek Simon from the laboratory of Dr. Michael Poirier (OSU).
Under the conditions tested, acetylation in either the entry-exit or LRS region does not have a significant impact on nucleosome stability as evidenced by the similar levels of nucleosome loss during force extension when compared to wild-type. In fact, nucleosomes acetylated at lysines 77 and 79 within the LRS region appeared to correspond with a mild reduction in nucleosome loss during force extension.

Although no significant impact of H4-K77ac,K79ac on nucleosome stability was observed, further characterization of semi-synthetic nucleosomes demonstrated a twofold increase in DNA site-accessibility as determined by assaying the binding of LexA protein to nucleosome DNA.\textsuperscript{56} This experiment was performed as detailed in Chapter 3 and yielded an observed $K_{eq}$ of 1.7 ± 0.4 for H4-K77ac,K79ac relative to unmodified nucleosomes.

Preparation of Highly Acetylated Mononucleosomes

Having successfully reconstituted nucleosomes with H4-K77ac,K79ac and H4-pT80 semi-synthetic proteins and verifying that modification within the LRS region does not lead to gross destabilization of the nucleosome structure, we wanted to probe the impact of simultaneously modifying multiple regions of the nucleosome. To this end, we reconstituted nucleosomes as previously described with semi-synthetic H3-K115ac,122ac and H4-K77ac,K79ac simultaneously introducing acetylation within the nucleosome LRS
and dyad regions. Nucleosome reconstitution was assayed by native PAGE to assess mononucleosome formation (Fig. 21). Nucleosomes reconstituted simultaneously with H3-K115ac,K122ac and H4-K77ac,K79ac proteins include 8 modified lysine residues, due to the symmetry of the nucleosome, with each acetylation located at important contact sites in the histone-DNA interface.

The observed similarities between the H3-K115ac,K122ac and H4-K77ac,K79ac octo-acetylated and wild-type demonstrates that the nucleosome is remarkably resistant to modification at the protein-DNA interface. This result had been predicted on a thermodynamic level, but this is the first concrete demonstration of this finding. Since dyad and LRS region modifications appear to define different dynamic events within the nucleosome, it will be particularly interesting to further characterize these nucleosomes simultaneously modified across multiple surfaces to precisely determine the effects on nucleosome stability and dynamics.
Figure 21. Reconstitution of highly acetylated nucleosomes

Structure (left) of H3-K115ac,K122ac (tan and cyan) and H4-K77ac,K79ac (red and blue) containing nucleosome with sites of lysine acetylation presented in green. Structure derived from the original PDB: 1KX3. Native PAGE analysis (right) of crude nucleosome reconstitutions with unmodified (WT), H3-K115ac,K122ac (Nuc H3-2ac), H4-K77ac,K79ac (Nuc H4-2ac), or H3-K115ac,K122ac and H4-K77ac,K79ac (Nuc 4ac) proteins. Gel imaged by fluorescence detection of labeled DNA.
Conclusions

Dynamic post-translational modification throughout the sequence of histone proteins regulates nucleosome stability and dynamics. The characterization of these various modifications requires the preparation of homogenous samples of nucleosomes incorporating proteins with a defined subset of modifications suitable for characterization. Herein, we introduce a novel ligation-desulfurization strategy for the preparation of semi-synthetic histone H4 proteins site-specifically modified within the LRS or αC helix regions of the proteins C-terminus. Furthermore, we demonstrate the success of our EPL strategy by preparing phosphorylated and acetylated semi-synthetic histone H4 proteins which we have begun to characterize.

Semi-synthetic H4-K77ac,79ac and H4-pT80 proteins prepared by the introduced ligation-desulfurization have been successfully reconstituted into homogeneous octamer and nucleosome samples suitable for characterization. Comparison of our ligation-desulfurization approach for histone H4 with our published EPL strategy for H3 to place modifications into the nucleosome dyad suggests certain features in common between synthetic approaches to the structured nucleosome core. In each EPL study, the synthetic peptide defines an important structural interface within the nucleosome core. Our studies with semi-synthetic H3 demonstrated that mixtures of H3(1-109) and full length H3 with greater than 50% full-length protein could successfully refold pure histone octamer, to the exclusion of truncated H3 lacking the αC helix. Similarly, semi-synthetic histone H4
co-purified with truncated H4(1-75) was successfully refolded into pure histone octamer without detectable incorporation of the truncated protein. Importantly, this applies only to the full histone octamer in which the H4 αC helix forms an important interface in the tetramer structure.

We have validated the suitability of generated nucleosomes for biophysical characterization by comparing H4-K77ac,K79ac nucleosomes with nucleosomes acetylated at H3-K56ac (Chapter 3) and H3-K115ac,K122ac. Completed characterization suggests that H4-K77ac,K79ac facilitates DNA unwrapping similarly to H3-K56ac modification within the nucleosome entry-exit region but has minimal impact on nucleosome array stability which is highly perturbed by H3-K115ac,K122ac modification within the nucleosome dyad.\(^{56}\)

There are multiple lines of evidence for interactions between dynamic events in the nucleosome, for example in DNA unwrapping and protein binding,\(^ {83}\) in nucleosome disassembly,\(^ {84}\) and in chromatin remodeling.\(^ {85}\) The development of multiple synthetic strategies is necessary to allow preparation of homogenous samples of nucleosomes with modifications across multiple functional regions of the histone-DNA interface. We have combined our semi-synthetic EPL strategies for the preparation of modified histones H3 and H4 to prepare the first semi-synthetic nucleosomes to contain eight site-specific modifications and demonstrated that despite the presence of these acetylations at the histone-DNA interface, the nucleosome still maintains its overall structure. The
simultaneous incorporation of eight acetylated lysine residues across four semi-synthetic proteins within the folded core of the nucleosomes represents the most highly modified core reported to date.
Acknowledgements

Several researchers in the Ottesen Laboratory contributed to elements of the H4 EPL-desulfurization project. Although I was personally involved in the design and/or completion of all aspects of the presented work, it could not have been completed without the effort and assistance of many members of the laboratory. Danny Waite participated in the initial cloning of the H4-Intein-CBD fusion protein and assisted with the initial isolation of H4 thioester. Michelle Ferdinand, Chris Smith, Tiffiny Rye, Chris Barna, Chris Thomas, and Eric Fredrickson contributed throughout the project by supporting the preparation of individual peptide segments, ligation to generated modified proteins, and purification of the various components necessary for EPL. Thanks to Justin North and Dr. Michael Poirier for assistance with troubleshooting in the refolding and purification of histone octamers and guidance in nucleosome reconstitution and sucrose gradient-purification. Thanks to Marek Simon for preparing and characterizing the nucleosome arrays, and developing and characterizing the magnetic tweezers force extension experiments. A special thank you to Dr. Dennis Bong for the use of his peptide synthesizer, and Drs. Justin Wu and Zucai Suo for the use of their FPLCs for the purification of semi-synthetic histone octamers.
Chapter 3: Total Synthesis of Histone H3 Acetylated at Lysine 56

Introduction

The EPL-desulfurization strategy described in Chapter 2 is appropriate for the introduction of histone modifications clustered near the terminus of a histone protein. However, biologically functional histone proteins often bear combinations of histone modifications throughout the protein sequence. Total chemical synthesis by sequential native chemical ligation (NCL) is the directed chemoselective combination of multiple peptide fragments to generate a larger protein. A sequential series of peptide condensations allows for the generation of proteins too large to obtain from the combination of two synthetic segments. Since peptide synthesis allows full control over each prepared segment, total chemical synthesis should be the most flexible method for the preparation of heavily modified histone proteins.

Several strategies have been developed to allow for multiple rounds of NCL to generate proteins by total chemical synthesis. Peptides prepared with different thioesters may be sequentially ligated in a kinetically-controlled manner determined by the reactivity of the
individual thioesters. Alternatively, reversible protection of the N-terminal Cys of one segment allows for the directed ligation of two peptide segments in a chemically controlled order. The ligation of a peptide bearing an N-terminal Cys and a second peptide which includes a protected N-terminal Cys and a C-terminal thioester may be initiated, and following ligation, the protected N-terminal Cys of the product may be deprotected to allow for subsequent ligation to a new thioester (Fig. 22).

Figure 22. Scheme for total synthesis by sequential NCL
The inherent flexibility of total chemical synthesis when compared to alternative techniques such as cysteine alkylation and codon suppression makes it the ideal method for the preparation of homogeneous histone proteins which are site-specifically modified within the folded protein core distal to the termini. Importantly, total chemical synthesis by sequential NCL would allow for the introduction of various modifications throughout the sequence of a histone and is limited only by the chemistry used to generate the individual peptide segments. We sought to confirm the value of sequential NCL for generating modified histone proteins by completing the total synthesis of histone H3 site-specifically acetylated at lysine 56.

Lysine 56 of histone H3 (H3-K56) lies within the αN extension of the folded protein core within the entry-exit region of the nucleosome. H3-K56 is oriented towards the DNA major groove and is within electrostatic binding distance of the DNA backbone. Acetylation of histone H3-K56 was first reported in Saccharomyces cerevisiae with ~28% of H3 molecules containing the modification. Acetylation of lysine introduces steric bulk and a neutralization of charge attenuating electrostatic interactions of the unmodified amino acid.

H3-K56ac is enriched at actively transcribed genes including core histone genes and is cell cycle regulated with peak expression during S phase. H3-K56 has been shown to be acetylated in newly synthesized H3 and associated with the CAF-1 chaperone complex.
prior to deposition into nucleosomes. In yeast, histone H3-K56 is acetylated by the acetyltransferase Rtt109 in the presence of the histone chaperone Asf1. Furthermore, H3-K56ac has also been implicated to play a role in DNA repair, genomic stability, and transcriptional regulation.

Due to its location within the αN helix of H3, distal from either the N- or C-termini, it is impractical to generate H3-K56ac protein by a simple one-step ligation reaction due to the inherent size limitations of peptide synthesis. We therefore decided to employ a sequential NCL strategy for the total synthesis of histone H3-K56ac.

Histone H3 is a highly conserved protein and in most species contains a single native cysteine at position 110, which we have previously employed for the generation of modified histones by EPL. Due to its proximity to the terminus, C110 is not amenable to a total synthesis strategy. We initially identified possible ligation sites by identifying cysteine residues by sequence alignment and design a two-step ligation strategy.

We successfully synthesized the H3-K56ac(R40C,S96C,C110A) protein and incorporated it into histone octamers and nucleosomes. In preliminary studies, nucleosomes containing the synthetic protein demonstrated perturbed DNA site-accessibility. Control experiments with recombinant H3(R40C,S96C,C110A) protein also displayed a similar perturbation suggesting that a significant proportion of the
observed effect was attributable to the cysteine substitutions incorporated through our ligation strategy.

Due to the unexpected result of cysteine substitution altering nucleosome dynamics we adapted a ligation-desulfurization strategy that would result in traceless introduction of the desired acetylation of lysine 56. We again used the H3(C110A) sequence and chose native alanine residues as the sites of ligation. Following preparation by sequential NCL, the synthetic protein was desulfurized to yield the native H3-K56ac(C110A) protein.

Synthetic H3-K56ac(C110A) was subsequently incorporated into histone octamers and nucleosomes for characterization. Acetylation of lysine 56 of histone H3 was found to impart increased DNA site-accessibility and binding of the LexA protein to prepared semi-synthetic nucleosomes. Additionally, single molecule force-extension experiments demonstrated that H3-K56ac dependent DNA site-accessibility does not significantly alter the stability of a semi-synthetic nucleosome array.
Experimental Methods

Peptide synthesis

All peptides (Table 2) were synthesized manually using standard Boc-N-α protection strategies and *in situ* neutralization protocols\(^{57}\) utilizing HBTU activation. C-terminal peptides C1 and C2 were synthesized on pre-loaded Boc-Ala-PAM resin (Novabiochem). Thioester peptides N1, N2, M1 and M2 were synthesized on MBHA resin with a mercaptopropionamide linker to generate the C-terminal thioester moiety necessary for subsequent ligation.\(^{94, 95}\) Acetylated lysine was incorporated as the commercially available Boc-protected derivative (N-α-t.-Boc-acetyllysine), and the protected N-terminal Cys was incorporated as thiazolidine (Thz; Boc-L-thiazolidine-4-carboxylic acid). Peptides were cleaved from the solid support with standard anhydrous hydrogen fluoride (HF) cleavage conditions utilizing p-cresol as scavenger. Following synthesis and purification, all peptide purities were assessed by RP-HPLC as >95% with the exception of peptide M1, which contained a mixture of the desired methionine as well as oxidized methionine [Met(O)].

Synthesis of H3-K56ac(R40C,S96C,C110A)

Synthetic H3-K56ac(R40C,S96C,C110A) proteins were generated by sequential native chemical ligation. In the first step of the ligation, peptide M1 propionamide thioester was
resuspended with 5-10 fold molar excess of peptide C1 in 100 mM HEPES pH 7.5, 1 M NaCl, 50 mM MESNA and 6 M guanidine HCl and reacted for 2 days at 25°C. Upon completion, direct addition of 500 mM methoxylamine HCl to the ligation mixture generated the free N-terminal Cys by unmasking the Thz\textsuperscript{96}. Complete conversion to the desired terminal Cys was observed within 6 hours. The M1C1 ligation product was purified to >95% by RP-HPLC over a gradient of 23-50% isopropanol / 0.1%TFA with a Vydac C4 column at 45°C and the product identity was confirmed by MALDI-TOF MS.

Purified ligation product M1C1 was resuspended with a 20-fold molar excess of peptide N1 in 100 mM HEPES pH 7.5, 1 M NaCl, 50 mM MESNA, 10 mM TCEP and 6 M Gdn-HCl. The ligation mixture was nutated for 24 hours at 25°C to generate the site specifically modified H3-K56ac(R40C,S96C,C110A). The final product was purified by RP-HPLC with a step gradient of acetonitrile / 0.1% TFA on a Supelco Widebore C18 column at 25°C as follows: 11-16% over 5 minutes, 16-43% over 5 minutes and 43-66% over 30 minutes. Fractions identified by MALDI-TOF MS to contain full length H3-K56ac(R40C, S96C, C110A) with and without the Met(O) species were pooled and lyophilized.

Reduction of Met(O) to Met in H3-K56ac(R40C,S96C,C110A) was accomplished by resuspending the lyophilized protein in 200µL TFA with 25µL dimethylsulfide and 0.045 M sodium iodide (NaI).\textsuperscript{97} Reduction was allowed to proceed for 1 hour on ice until complete as monitored by MALDI-TOF MS. Reduced protein was precipitated and
washed with cold anhydrous diethyl ether, then purified by RP-HPLC on a Supelco Widebore C18 column with a gradient of 43-66% acetonitrile / 0.1% TFA (Fig. X). MALDI-TOF MS confirmed that the Met(O) species was absent in the collected fraction. The total synthesis of H3-K56ac(R40C,S96C,C110A) as described provided an overall ligation yield of 2% (measured by UV quantification).

Synthesis of H3-K56ac(A47C,A91C,C110A)

Sequential NCL was employed to generate synthetic H3-K56ac(A47C,A91C,C110A). Thioester peptide M2 (typically 0.5 mg) was resuspended with 2.5 molar excess of C2 in 100 mM phosphate pH 7.5, 1 M NaCl, 60 mM 4-mercaptophenylacetic acid (MPAA), 20 mM TCEP and 50% trifluoroethanol (TFE) and reacted for 48 hours at 25 °C. We determined that if ring opening was carried out prior to the first purification step, we observed back-formation of Thz-M2C2 over the course of the second ligation reaction. We attribute this to co-purification of trace amounts of formaldehyde generated in cleavage of the His(Bom) side chain protecting group in peptide N2. This ring closure was minimized by the presence of methoxylamine in the ligation reaction. We therefore purified Thz-M2C2 prior to the ring-opening step. While the ligation product partially precipitated under these buffer conditions, reaction in a phosphate/Gdn buffer prevented this problem and was used in future ligation iterations. Product was resusupended in 100 mM phosphate pH 7.5, 1 M NaCl, 6 M Gdn HCl with 60 mM TCEP, and purified by RP-
HPLC on a Supelco Widebore C18 column with a 32-59% acetonitrile / 0.1% TFA gradient.

Lyophilized ligation product Thz-M2C2 was resuspended with 100 mM phosphate pH 7.5, 1 M NaCl, 6 M Gdn-HCl, 20 mM TCEP. 350 mM methoxylamine hydrochloride was added to convert the N-terminal Thz to Cys. Deprotection was allowed to proceed for 6 hrs at 25°C. The mixture was adjusted to pH 7.5 with NaOH and then MESNA was added to the mixture to a final concentration of 100 mM. The ligation was initiated with the addition of a 5-fold molar excess of peptide N2. Ligations were monitored by RP-HPLC and SDS-PAGE for 4 to 6 days at 25°C until no additional product formation was observed.

Desulfurization of H3-K56ac(A47C,A91C,C110A) to yield H3-K56ac(C110A)

The crude H3-K56ac(A47C,A91C,C110A) ligation mixture was directly desulfurized prior to purification using free-radical-desulfurization conditions. The mixture was adjusted to final concentrations of 50 mM Phosphate pH 7.5, 500 mM NaCl, 0.3 M TCEP, 100 mM MESNA, 5 M Gdn-HCl and the sample was sparged with argon for 30 minutes. Desulfurization was initiated with the addition of VA-044US (Wako Chemical) to a final concentration of 10 mM at 42°C, and the reaction was allowed to proceed until complete as monitored by MALDI-TOF analysis and was typically complete within 3
hours. The final desulfurized product H3-K56ac(C110A) was purified by RP-HPLC with a gradient of 41-59% acetonitrile / 0.1% TFA on a Supelco Widebore C18 column.

A typical ligation began with 0.5 mg of limiting peptide M2. The ligation and desulfurization procedures described yielded 93 µg of final product H3-K56ac(C110A) as determined by UV quantification on a NanoDrop 1000 (Thermo Scientific), for an overall synthetic yield of 7%. If Met(O) species was observed in the full-length native H3-K56ac(C110A), reduction was carried out on the purified product without the need for further purification.

Histone Octamer Refolding

Recombinant histones were expressed and purified as previously described. Plasmids encoding histones H2A(K119C), H2B, H3, and H4 were generous gifts from Dr. Karolin Luger (Colorado State University) and Dr. Jonathan Widom (Northwestern University). Mutations H3(R40C), H3(S96C), H3(C110A), H3(K56Q), and H4(S47C) were introduced by site-directed mutagenesis (Stratagene). H2A(K119C) was labeled before or after histone octamer refolding with Cy5-maleamide (GE Healthcare). We achieved labeling efficiency of 75-90% as determined by mass spectrometry and UV absorption. Histones: H2A(K119C), H2B, H4 and either H3, H3(C110A)$_{rec}$, H3(C110A)$_{syn}$, H3(K56Q), H3(K56Q,C110A) or H3-K56ac(C110A) were combined at equal molar
ratios and refolded as previously described.99 The purity of each octamer was confirmed by SDS-PAGE and MALDI-TOF MS.

Mononucleosome Reconstitutions

Nucleosomes were reconstituted by salt double dialysis77 with 7µg of DNA and 5µg of histone octamer (HO). The DNA and HO were mixed in 50 µl of 0.5x TE pH 8.0, 2 M NaCl and 1 mM BZA (benzamidine). The sample was loaded into an engineered 50 µl dialysis chamber which was placed in a large dialysis tube with 80 ml of 0.5x TE pH 8.0, 2 M NaCl and 1 mM BZA. The large tube was extensively dialyzed against 0.5x TE pH 8.0 with 1 mM BZA. The 50 µl sample was extracted from the dialysis button and purified by sucrose gradient-centrifugation.

Determination of Equilibrium Constant for DNA Site Exposure

All FRET and LexA binding measurements presented in this document were completed by Justin North in the laboratory of Dr. Michael Poirier as detailed in the literature.11 The equilibrium constant for site exposure of nucleosomes (K_{eq}) was determined by a LexA binding assay as previously described.11 Briefly, the S_{0.5-DNA} (the LexA concentration at which 50% of naked DNA is bound to protein) and S_{0.5-nuc} (the LexA concentration at
which 50% of nucleosome is bound to protein) terms are determined experimentally. The equilibrium constant may be determined from the equation $K_{eq} = S_{0.5\text{-DNA}} / S_{0.5\text{-nuc}}$. 
Results and Discussion

Histone H3 is a 135-residue protein that can be easily refolded and incorporated in vitro into nucleosomes suitable for biophysical characterization. Histone H3 may be heavily modified with biologically important PTMs reported to be present throughout the protein sequence. These properties make it an excellent candidate for total synthesis by sequential NCL. To confirm the suitability of histone H3 for total synthesis, we initially developed a sequential NCL scheme for the preparation histone H3 site-specifically acetylated at lysine 56 (H3-K56ac). As the basis for our ligation strategy, we selected the modified *Xenopus laevis* H3(C110A) sequence, which is commonly used in biophysical studies. The C110A substitution occurs in yeast and has not previously been reported to affect nucleosome structure, positioning, and DNA unwrapping. Additionally, the H3(C110A) substitution eliminates the presence of extraneous cysteine residues which may perturb desired ligation events through disulfide formation and non-productive transthioesterification.

We identified Arg40 and Ser96 based on homology alignment as possible sites for Cys substitution that could be introduced into the H3(C110A) sequence and would allow for a three-segment, two-step sequential NCL strategy (Fig. 23). The H3(R40C) substitution was identified in *Cairina moschata* and H3(S96C) is present in the H3.1 *Homo sapiens*, *Mus musculus*, and *Caenorhabditis elegans* variants.
In the proposed sequential NCL strategy, the longest synthetic segment would be a central 56-residue peptide containing an acetylated lysine that would eventually become Lys56 (Fig. 23 and Table 2). A key feature of the proposed scheme is the introduction of the N-terminal Cys of peptide M1 as thiazolidine. Thiazolidine serves as a masked cysteine which is compatible with peptide synthesis, stable under common ligation conditions, and prevents spurious ligation of the M1 peptide. Although several common
cysteine protection strategies are compatible with sequential NCL, thiazolidine is commonly employed.\textsuperscript{96, 105} When treated with methoxylamine at a pH of \( \sim 4 \), the thiazolidine ring will convert to an N-terminal cysteine available for subsequent ligation.\textsuperscript{105}

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Unmodified H3(1-39) Thioester</td>
<td>ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPH-COSR</td>
</tr>
<tr>
<td>M1</td>
<td>H3(40-95) R40Thz,K56ac Thioester</td>
<td>(Thz)YRPGTVALREIRRYQ(Kac)STELLIRKLPQRLVREIAQDFKT DLRFQSSAVMALQEA-COSR</td>
</tr>
<tr>
<td>C1</td>
<td>H3(96-135) S96C,C110A</td>
<td>CEAYLVALFEDTNLAIIHAKRVTIMPKDIQLARRIRGERA-COOH</td>
</tr>
<tr>
<td>N2</td>
<td>Unmodified H3(1-46) Thioester</td>
<td>ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRP GTV-COSR</td>
</tr>
<tr>
<td>M2</td>
<td>H3(47-90) A47Thz,K56ac Thioester</td>
<td>(Thz)LREIRRYQ(Kac)STELLIRKLFQRLVREIAQDFKTDLRFQSS AVM-COSR</td>
</tr>
<tr>
<td>C2</td>
<td>H3(91-135) A91C,C110A</td>
<td>CLQEASEAYLVALFEDTNLAIIHAKRVTIMPKDIQLARRIRGERA-COOH</td>
</tr>
</tbody>
</table>

Table 2. Peptides for the total synthesis of H3-K56ac

Sequential NCL for the Total Synthesis of H3-K56ac(R40C,S96C,C110A)

Peptides N1, M1 and C1 were prepared by manual SPPS with Boc \textit{in situ} neutralization protocols (Table 2). Peptide C1 was synthesized directly on substituted Boc-Ala-PAM
resin and Peptides N1 and M1 were synthesized on mercaptopropionamide derived MBHA resin which was prepared according to literature protocols. The compatibility of Boc chemistry with the direct preparation of base-labile peptide thioesters was a primary determinant in our choice of synthetic chemistry. Following synthesis, peptides were deprotected and cleaved from resin with anhydrous HF in the presence of p-cresol as a scavenger molecule. Crude peptides were precipitated in cold diethyl ether, resuspended with acetonitrile / H₂O + 0.1% TFA, and lyophilized.

Following synthesis, crude lyophilized N1 and C1 peptides were resuspended and purified by RP-HPLC. The purity and identity of the lyophilized pool was verified by RP-HPLC and MALDI-TOF MS (Fig. 24). Overall, 17 mg of peptide N1 and 21 mg of peptide C1 were obtained.

Manual synthesis of the M1 thioester peptide segment was initially completed as described for the N1 peptide. The synthesis of M1 was particularly challenging due in part to its large size and amino acid composition. Additionally, the requirement for the M1 peptide to be synthesized on mercaptopropionamide substituted resin further exacerbated the challenging synthesis as deprotection with TFA is known to lead to incremental non-productive cleavage of peptide from the resin which over a long synthesis will significantly reduce yield.
Figure 24. Characterization of purified peptides N1 and C1

RP-HPLC (left) and MALDI-TOF MS (right) analysis of pure peptide pools. Peptides were resolved by RP-HPLC on an acetonitrile / H₂O + 0.1% TFA gradient as follows: N1 (9-18%) and C2 (29-36%). Expected and observed [M+H]+ for each peptide as described.

To minimize the formation of deletion products and maximize overall yield of the peptide, the ninhydrin test for free amines was employed throughout the synthesis of peptide M1 to ensure maximum coupling efficiency. Once complete coupling of an amino acid was verified, the next deprotection and coupling cycle was performed. Following the coupling of residue 41 which corresponds to tyrosine 54 in the full sequence of H3, the resin was deprotected in preparation for the next coupling cycle and
a ninhydrin test was performed to verify the presence of free amines on the resin. Unexpectedly, the conducted test was negative for the presence of free amines. We therefore performed a test cleavage of the base resin to determine whether the negative result was due to complete chain termination, or whether it was a result of occluded termini resulting in a false negative result.

Figure 25. Comparison of RP-HPLC solvent systems for partial M1 synthesis

RP-HPLC of crude partial synthesis of peptide M1 corresponding to H3-K56ac residues 54-95. Peptide was resolved by RP-HPLC on a (left) 0-65% acetonitrile / H₂O + 0.1% TFA or (right) 18-36% isopropanol / H₂O + 0.1% TFA gradient at 40 °C.

Analysis of the crude test cleavage mixture by RP-HPLC with a gradient of the commonly used acetonitrile / H₂O + 0.1% TFA co-solvent system over a Grace Vydac C18 column was complex and difficult to interpret (Fig. 25, left). To better resolve the
species present in the cleavage mixture, we assessed various RP-HPLC conditions including column hydrophobicity (C4 versus C18), alternative co-solvents (methanol and isopropanol), and at temperatures ranging from 25-40 °C. We found that a gradient of 18-36% isopropanol / H2O + 0.1% TFA yielded the best resolution at 40 °C on a Grace Vydac C4 column (Fig. 25, right). Under these conditions, single peak resolution was obtained and the full length product was observed as the major species as verified by MALDI-TOF MS (Fig. 26).

![Figure 26](image)

**Figure 26.** MALDI-TOF MS of partial M1 peptide (H3-K56ac residues 52-95)

Expected and observed [M+H]⁺ for peptide M1 as described.

Since the desired 41-residue truncated M1 peptide was the major species present, we continued the manual synthesis of peptide M1 on the remaining resin. The remaining
amino acids were double coupled with HBTU activation. As previously described, thiazolidine was introduced as the N-terminal amino acid. The crude peptide mixture was deprotected and cleaved from the resin by standard HF protocols, precipitated, and lyophilized as previously described.

The resultant crude peptide mixture was purified by RP-HPLC with on a Grace Vydac C4 column with a gradient of 24-38% isopropanol / 0.1% TFA at 40 °C. Individual reactions were analyzed by RP-HPLC and MALDI-TOF MS for purity and identity. In addition to the desired species, MALDI-TOF MS analysis also detected the presence of a +16 Da species over the parent ion. We attributed this ion to the oxidation of methionine [Met(O)] which is known to occur during peptide cleavage as well as during prolonged storage of methionine containing resins. We identified literature methods for the reduction of Met(O) to yield the desired Met species that would be compatible with the full length H3-K56ac(R40C,S96C,C110A) protein. To maximize synthetic yield, all fractions determined to contain more than 95% of the Met and Met(O) species were pooled and lyophilized. Due to the low melting temperature of the isopropanol / H2O mixture, lyophilization was closely monitored to ensure minimal product loss due. The pure peptide pool was analyzed by RP-HPLC and MALDI-TOF MS to verify purity and identity (Fig. 27). Following lyophilization, 2 mg of peptide M1 was obtained from an initial 80 mg of crude peptide for a purification yield of 2.5%.
Figure 27. Characterization of purified peptide M1

RP-HPLC of purified peptide M1. Peptide was resolved by RP-HPLC on a 24-38% isopropanol / H₂O + 0.1% TFA gradient at 45 °C.

Total Chemical Synthesis of H3-K56ac(R40C,S96C,C110A)

Lyophilized M1 propionamide thioester peptide was resuspended with a 5-10 fold molar excess of peptide C1 in the presence of MESNA and 6 M guanidine HCl. Conversion of the propionamide thioester to the more active MESNA thioester may be monitored by RP-HPLC and MALDI-TOF MS and is typically complete within an hour. The ligation mixture was incubated at 25 °C with nutation. Ligation progress was monitored by RP-HPLC and was determined to be complete when no further product formation was observed, typically 2 days. Direct addition of methoxylamine to the ligation mixture generated the free N-terminal Cys by unmasking the Thz⁹⁶ and was typically complete within 6 hours. The resultant product (M1C1) was purified to >95% by RP-HPLC over a
isopropanol / 0.1%TFA gradient with a Grace Vydac C4 column at 45°C and product identity was confirmed by MALDI-TOF MS (Fig. 28).

Figure 28. Characterization of ligation product M1C1

RP-HPLC (left) and MALDI-TOF MS (right) analysis of purified ligation product M1C1. Ligation product was resolved by RP-HPLC on a 22-50% isopropanol / H₂O + 0.1% TFA gradient at 45 °C. Expected and observed [M+H]⁺ for M1C1 ligation product as described. Expansion in the region of the [M+H]⁺ peaks (inset) demonstrates the Met and Met(O) species.

Purified ligation product M1C1 was resuspended with a 20-fold molar excess of peptide N1 in ligation buffer and the mixture was nutated for 24 hours at 25°C to generate the site-specifically modified H3-K56ac(R40C,S96C,C110A). The final product was purified by RP-HPLC with an acetonitrile / 0.1% TFA gradient on a Supelco Widebore C18 column at 25°C. Fractions identified by MALDI-TOF MS to contain full length H3-
K56ac(R40C, S96C,C110A) with and without the Met(O) species were pooled and lyophilized.

Figure 29. Reduction of oxidized Met in H3-K56ac(R40C,S96C,C110A)

MALDI-TOF MS analysis of H3-K56ac(R40C,S96C,C110A) protein before (top) and after (bottom) reduction of Met(O) with NaI / DMS. Expected and observed [M+H]⁺ for protein as described.

Reduction of Oxidized Methionine in H3-K56ac(R40C,S96C,C110A)

Reduction of oxidized methionine in H3-K56ac(R40C,S96C,C110A) was accomplished by suspending the lyophilized purified protein in 200µL TFA with DMS and NaI. Reduction was allowed to proceed for 1 hr on ice until complete as monitored by
MALDI-TOF MS (Fig. 29). Reduced protein was precipitated and washed with cold anhydrous diethyl ether, then purified by RP-HPLC on a Supelco Widebore C18 column with a gradient of acetonitrile / 0.1% TFA. Fractions were analyzed by MALDI-TOF MS and confirmed that quantitative reduction of Met(O) to Met had occurred. Fractions determined to be > 95% pure were pooled and lyophilized. Product purity and identity was verified by RP-HPLC and MALDI-TOF MS (Fig. 30). The total synthesis of H3-K56ac(R40C,S96C,C110A) as described provided an overall ligation yield of 2% (measured by UV absorbance) based on the limiting peptide segment M1. While these yields were low, they were sufficient for preliminary characterization studies.

Figure 30. Analysis of purified H3-K56ac(R40C,S96C,C110A)

RP-HPLC (left) and MALDI-TOF MS (right) analysis of purified H3-K56ac (R40C,S96C,C110A) synthetic protein. Protein was resolved by RP-HPLC on a 27-54% isopropanol / H₂O + 0.1% TFA gradient at 45 °C. Expected and observed [M+H]^+ for protein as described.
Characterization of Synthetic H3-K56ac(R40C,S96C,C110A)

To allow for characterization of the synthetic H3-K56ac9R40C,S96C,C110A), the prepared protein was provided to Justin North in Dr. Michael Poirier’s laboratory. They reconstituted the synthetic H3-K56ac(R40C,S96C,C110A) into histone octamers with recombinant H2B, H4, and Cy5 labeled H2A-K119C proteins by double dialysis in an engineered chamber as previously described. Refolded octamers were purified by FPLC size-exclusion chromatography. Individual factions were assayed by SDS-PAGE, pooled, and concentrated to > 1 mg/mL.

Histone octamers were similarly prepared with recombinant unmodified H3, H3(R40C,S96C,C110A), or H3-K56Q(R40C,S96C,C110A) proteins as controls. The H3-K56Q(R40C,S96C,C110A) octamer incorporates the lysine to glutamine substitution which is a commonly used genetic manipulation to determine the effects of constitutively acetylated lysines in vivo. The H3-K56Q(R40C,S96C,C110A) octamer serves as an important means of determining to what extent the glutamine genetic substitution is able to reproduce the biophysical properties of the cognate acetyllysine modification.

Nucleosomes were reconstituted with a 5'-Cy3-labeled 147 bp 601 nucleosome positioning sequence containing a LexA binding site located between 8 bp and 27 bp, and the purified octamers containing either unmodified H3, H3(R40C,S96C,C110A), H3-K56Q(R40C,S96C,C110A), or synthetic H3-K56ac(R40C,S96C,C110A) by double
dialysis as previously described. The extent of nucleosome reconstitution was assayed by native PAGE (Fig. 31).

Figure 31. Native PAGE analysis of crude nucleosome reconstitutions

Native PAGE (left) of crude nucleosome reconstitution with either unmodified (WT) or modified H3 proteins as described. Nucleosome structure derived from PDB (1KX5) and labeled as described. This work was completed by and the figure is courtesy of Justin North from the laboratory of Dr. Michael Poirier (OSU).
In the prepared nucleosomes, when the DNA is tightly wrapped forming the classical structure, the Cy3 and Cy5 FRET pair is positioned near the entry-exit region of the nucleosome. The placement of one fluorophore on the DNA and the other on the histone octamer allows the detection of DNA movement relative to the octamer and FRET occurs between them. The mobility of the DNA in the entry-exit region may be measured by quantitating the overall FRET efficiency of the system. A decrease in FRET efficiency is indicative of the DNA spending an increased time dissociated from the entry-exit region and thereby limiting the cross-talk between the two fluorophores.

In preliminary studies, it was observed that a significant decrease in FRET efficiency from 0.62 ± 0.02 to 0.43 ± 0.01 occurred when wild-type H3-containing nucleosomes were compared to the control H3(R40C,S96C,C110A) nucleosomes (Fig. 32). FRET efficiency in nucleosomes containing H3-K56Q(R40C,S96C,C110A) or the synthetic H3-K56ac(R40C,S96C,C110A) was further decreased to 0.38 ± 0.02 and 0.35 ± 0.04, respectively. However, this additional reduction in FRET efficiency is significantly less than that induced with the Cys substitutions alone.

The observed effect of the introduced cysteine substitutions was greater than that of the acetylation mimic, glutamine, or the cognate modification introduced by synthetic H3-K56ac(R40C,S96C,C110A). This observation suggests that the H3(R40C,S96C,C110A) containing nucleosomes display altered structure and/or dynamics. This result
demonstrates the potential impact the introduction of non-native amino acid substitutions in the histone sequence may have on nucleosome structure and/or dynamics.

Figure 32. FRET efficiency of H3-K56ac(R40C,S96C,C110A)

This work was completed by and the figure is courtesy of Justin North from the laboratory of Dr. Michael Poirier (OSU).11

Total Synthesis of H3-K56ac(C110A)

The unexpected impact of the H3-K56ac(R40C,S96C,C110A) cysteine substitutions on nucleosome structure and/or dynamics necessitated the development of an alternative scheme for synthesis of histone H3 acetylated at lysine 56. We developed a second
generation sequential native ligation scheme that would yield the H3-K56ac(C110A) protein which yields the native protein sequence with the exception of the C110A substitution (Fig. 33).

Figure 33. Total synthesis scheme for the preparation of H3-K56ac(C110A)
Early desulfurization techniques required mixed-phase catalysts, such as Raney nickel, which we anticipated would be poorly compatible with marginally soluble proteins such as histones. However, during our development of the initial histone total synthesis scheme, a solution-based free-radical approach for desulfurization was reported. This introduced method requires a source of radicals as well as common ligation components including TCEP and a thiol proton source. We previously discussed this technique in the context of chapter 2. The presented ligation-desulfurization strategy allows for traceless ligation in proteins which do not include native cysteine residues by targeting native alanine residues for substitution. The introduced cysteine residues are used for ligation and post-ligation desulfurization regenerates the native alanine residues. We selected alanines 47 and 91 as potential junction sites, with the additional advantage of decreasing the size of the middle peptide segment which was synthetically limiting in our initial syntheses.

Peptides N2, M2, and C2 (Table 2) were prepared by manual peptide synthesis utilizing Boc in situ neutralization protocols. The C2 peptide was synthesized directly on substituted Boc-Ala-PAM resin and peptides N2 and M2 were synthesized on mercaptopropionamide derived MBHA resin which was prepared according to literature protocols. All peptides were purified by RP-HPLC with a gradient of acetonitrile / H$_2$O + 0.1% TFA on a Supelco Widebore C18 column. The purity and identity of the peptide pools was verified by RP-HPLC and MALDI (Fig. 34). A total of 70 mg of pure N2 peptide was obtained from an initial 320 mg of crude for a purification yield of 22%.
For peptide C2, 19.6 mg of pure peptide was isolated from 215 mg of crude for a purification yield to 9%.

Figure 34. Analysis of purified peptides N2, M2 and C2

RP-HPLC (left) and MALDI-TOF MS (right) analysis of pure peptide pools. Peptides were resolved by RP-HPLC on an acetonitrile / H₂O + 0.1% TFA gradient as follows: N2 (14-23%), M2 (36-45%) and C2 (30-39%). Expected and observed [M+H]⁺ for peptides as described.
Peptide M2 was also purified on a Supelco Widebore C18 column with a gradient of acetonitrile / H$_2$O + 0.1% TFA, and 3.7 mg of pure peptide was isolated from 71 mg crude for a purification yield of 5%. Unlike peptide M1, the purification of M2 did not require the use of isopropanol as a co-solvent at elevated temperatures. This is highly desirable as the use of heat in purification may lead to undesirable side reactions including hydrolysis of glutamine to glutamic acid or hydrolysis of the thioester yielding a peptide that is inactive for ligation. Although the overall yield was still low, multi-milligram quantities of peptide M2 could be obtained from each synthesis which is sufficient for the preparation of H3-K56ac(C110A) protein.

Minimal methionine oxidation was observed during the synthesis of peptides M2 and C2. We tested reduction conditions on small aliquots of peptide M2 and determined that resuspension of lyophilized peptide with TFA in the presence of DMS and NaI lead to quantitative reduction of the Met(O) species within one hour as expected. Unexpectedly, a side product of this reaction was the unmasking of thiazolidine to yield a free N-terminal Cys. Consequently, if necessary, Met(O) reduction was conducted on the final ligation product rather than on the individual peptide segments to eliminate the possibility of non-productive peptide cyclization.

Sequential NCL was employed to generate synthetic H3-K56ac(A47C,A91C,C110A) protein. Thioester peptide M2 (typically 0.5 mg) was resuspended with 2.5 molar excess of C2 in modified ligation buffer (100 mM phosphate pH 7.5, 1 M NaCl, 20 mM TCEP
and 50% trifluoroethanol) with 4-mercaptophenylacetic acid (MPAA) as the co-thiol. MPAA thioesters have been reported to demonstrate accelerated kinetics over the commonly used MESNA thiol. Due to the acidic nature of MPAA, the pH of the ligation mixture was closely monitored throughout the reaction and adjusted as needed to maintain an optimal pH of ~7.5. Following resuspension, the ligation mixture was reacted for at 25°C with nutation. Reaction progress was monitored by RP-HPLC and allowed to proceed until no further product formation was observed, typically 48 hours (Fig. 35, Top).

Initially, the N-terminal Thz of the M2C2 ligation product was unmasked to yield Cys prior to purification by RP-HPLC. Individual fractions were assessed for content and an initial pool was assayed for purity and identity by RP-HPLC and MALDI-TOF MS (Fig. 35). The resultant pool was lyophilized in preparation for further use.

During preliminary test ligations of the purified M2C2 product with the N-terminal N2 peptide, we observed progressive back-formation of Thz-M2C2 over the course of the second ligation reaction. We attribute this to co-purification of trace amounts of formaldehyde generated in the cleavage of the His(Bom) side chain protecting group in peptide N2. This non-productive ring closure could be minimized by the presence of methoxylamine in the ligation reaction. Consequently, in subsequent preparations, we purified the Thz-M2C2 species prior to the ring-opening step as described.
Figure 35. Characterization of M2C2 ligation product

Annotated RP-HPLC chromatogram of crude ligation mixture (top) on a 27-66% acetonitrile / H₂O + 0.1% TFA gradient. RP-HPLC of pure M2C2 (middle) on a 39-66% acetonitrile / H₂O + 0.1% TFA gradient. MALDI-TOF MS (bottom) of pure M2C2 product with expected and observed [M+H]⁺ values as described.

Exp m/z: 10491
Obs m/z: 10492
Lyophilized ligation product Thz-M2C2 was resuspended in ligation buffer (100 mM phosphate pH 7.5, 1 M NaCl, 6 M Gdn-HCl, 20 mM TCEP) which was optimized to allow for maximum solubility of the Thz-M2C2. Conversion of the N-terminal Thz to Cys was initiated with the addition of 350 mM methoxylamine hydrochloride. Deprotection was allowed to proceed for 6 hrs at 25°C. The pH of the mixture was adjusted to 7.5 with the addition of NaOH and MESNA was added to a final concentration of 100 mM. Ligation was initiated with the addition of a 5-fold molar excess of peptide N2. Ligations were monitored by RP-HPLC and SDS-PAGE for 4 to 6 days at 25°C until no additional product formation was observed (Fig. 36). We attribute the slow rate of ligation product formation to the C-terminal valine residue of peptide N2. The valine thioester is known to result in slow ligation kinetics. The crude ligation mixture was carried forward for desulfurization without further purification to reveal the final H3-K56ac(C110A) protein.

Crude ligation mixture which contained synthetic H3-K56ac(A47C,A91C,C110A) was brought to desulfurization conditions with the addition of an equal volume of prepared 0.6 M TCEP solution in 4 M guanidine, and the overall MESNA concentration was maintained at a minimum of 75 mM. The resulting mixture was sparged with Ar and desulfurization was initiated with the addition of VA-044US to a final concentration of 10 mM. The sample was incubated at 42 °C and reaction progress was monitored by MALDI-TOF MS (Fig. 37). Quantitative desulfurization to yield the desired H3-K56ac(C110A) was typically complete within 4 hours.

96
Figure 36. Crude ligation mixture for H3-K56ac(A47C,A91C,C110A)

RP-HPLC chromatogram of H3-K56ac(A47C,A91C,C110A) crude ligation mixture on a 41-59% acetonitrile / H₂O + 0.1% TFA gradient.

Figure 37. Desulfurization to yield H3-K56ac

MALDI-TOF MS analysis of H3-K56ac protein before (top) and after (bottom) desulfurization. Expected and observed [M+H]⁺ for proteins as described.
The crude desulfurization mixture was brought to purification conditions with the addition of acetonitrile + 0.1% TFA and purified by RP-HPLC on a Supelco Widebore C18 column. Collected fractions were analyzed by MALDI-TOF MS and RP-HPLC. Fractions determined to be >95% pure were pooled and lyophilized. The purified H3-K56ac(C110A) product purity and identity was verified by RP-HPLC and MALDI-TOF MS (Fig. 38). A typical synthesis began with 0.5 mg of limiting peptide M2 and yielded 93 μg of the final product H3-K56ac(C110A) as determined by UV quantitation for an overall synthetic yield of 7%. We attribute the observed increase in synthetic yield to the use of only two chromatography steps through our ligation pathway. More recently we have identified that a significant portion of the full-length H3-K56ac(C110A) protein elutes in the solvent front during RP-HPLC under the conditions described. The loss of uncollected full-length protein in the solvent front may have also been a significant factor in the low observed yields for H3-K56ac(C110A).

Methionine oxidation during desulfurization was minimized by adequate sparging. In the cases where significant amounts of Met(O) were detected in purified H3-K56ac(C110A), the lyophilized protein could be reduced with NaI / DMS as previously described. Reduced protein was ether precipitated, lyophilized and directly used in octamer refolding studies without further purification.
Figure 38. Characterization of H3-K56ac(C110A)

RP-HPLC (left) and MALDI-TOF MS (right) analysis of purified synthetic H3-K56ac(C110A). Pure H3-K56ac(C110A) was resolved on a 43-61% acetonitrile / H$_2$O + 0.1% TFA gradient. Expected and observed [M+H]$^+$ for protein as described.

Characterization of Synthetic H3-K56ac(C110A)

Synthetic H3-K56ac(C110A) was characterized by Justin North and Dr. Michael Poirier to determine the impact of site-specifically introducing lysine acetylation within the entry-exit region of the nucleosome. Briefly, the synthetic H3-K56ac(C110A) was refolded with recombinant histone proteins into semi-synthetic histone octamers which were purified by size-exclusion chromatography as described for the H3-K56ac(R40C,S96C,C110A) protein. Histone octamers were similarly prepared with recombinant H3(C110A) or H3-K56Q(C110A) proteins. Purified octamers were
reconstituted with 5’-Cy3-end-labeled 147 bp 601 nucleosome positioning sequence containing a LexA binding site located between 8 bp and 27 bp to yield semi-synthetic nucleosomes suitable for biophysical characterization. Reclaimed reconstituted nucleosome was purified by ultracentrifugation over a sucrose gradient and fractionated. Mononucleosome containing fractions were pooled and the resultant pool was assayed by native PAGE (Fig. 39)

Figure 39. Native PAGE of sucrose gradient purified nucleosomes

Native PAGE analysis of nucleosome reconstitutions incorporating unmodified (WT), recombinant H3-K56Q(C110A) (K56Q), or H3-K56ac(C110A) (K56ac) proteins. This work was completed by and the figure is courtesy of Justin North from the laboratory of Dr. Michael Poirier (OSU).11
The Cy3 and Cy5 FRET pair is positioned within the purified mononucleosomes adjacent to the entry-exit region and allows the detection of DNA movement relative to the octamer by measuring the FRET efficiency of the system. A decrease in FRET efficiency is indicative of the DNA spending an increased time dissociated from the entry-exit region and thereby limiting the cross-talk between the two fluorophores. FRET efficiency was determined to be $0.73 \pm 0.01$ for H3(C110A), $0.53 \pm 0.01$ for H3-K56Q(C110A), and $0.52 \pm 0.01$ for the H3-K56ac(C110A) reconstituted nucleosomes (Fig. 40). The significant observed decrease in FRET efficiency in the H3-K56ac and H3-K56Q nucleosomes is indicative of increased dynamics of the DNA in relation to the histone octamer within the nucleosome entry-exit region.

Figure 40. FRET efficiency of purified nucleosomes

This work was completed by and the figure is courtesy of Justin North from the laboratory of Dr. Michael Poirier (OSU).

101
In order to determine whether altered DNA dynamics within the entry-exit region of H3-K56ac(C110A) and H3-K56Q(C110A) bearing nucleosomes resulted in a change of DNA site accessibility within this region, nucleosomes were reconstituted with DNA which contained the consensus binding sequence for the DNA binding protein LexA. If accessible, the interaction of the LexA protein with its binding site would sequester the DNA resulting in a further decrease in FRET efficiency (Fig. 41, top). Titration of LexA protein into nucleosome sample leads to an observable decrease in FRET efficiency (Fig. 41). Significantly, altered binding kinetics for LexA to the H3-K56ac(C110A) and H3-K56Q(C110A) bearing nucleosomes are observed compared to the unmodified species (Fig. 41).

The equilibrium constant for site exposure of nucleosomes ($K_{eq}$) is a quantification of DNA accessibility and was calculated from the LexA binding study to be 0.0055 ± 0.0009 for H3(C110A), 0.010 ± 0.002 for H3-K56Q(C110A), and 0.010 ± 0.002 for H3-K56ac(C110A) nucleosomes. The observed 1.8 ± 0.4 fold increase of $K_{eq}$ for the H3-K56ac(C110A) and H3-K56Q(C110A) bearing nucleosomes demonstrates a significant increase in DNA site accessibility in the entry-exit region of the nucleosome resulting from the introduced modifications. Interestingly, the lysine to glutamine genetic substitution which is routinely employed to mimic lysine acetylation in vivo quantitatively mimics the effects of acetylation under the conditions tested.
Figure 41. LexA binding to H3-K56ac nucleosomes

Schematic (top) for measuring DNA site-accessibility by FRET analysis of LexA binding to nucleosomal DNA. This work was completed by and the figure is courtesy of Justin North from the laboratory of Dr. Michael Poirier (OSU).¹¹

We further probed the effects of H3-K56 acetylation in the entry-exit region, by assembling an array consisting of 17 nucleosomes assembled on a single DNA molecule. The array was immobilized and force-extension experiments were conducted as previously described.⁵⁶ The observed nucleosome dissociation pattern for arrays incorporating recombinant H3(C110A) and those containing H3-K56ac(C110A) were remarkably similar (Fig. 42). This result suggests that acetylation of K56 has a negligible effect on array stability despite attenuating DNA site accessibility in mononucleosomes.
In contrast, the acetylation of histone H3 lysines 115 and 122 in the nucleosome dyad which do not alter DNA site accessibility significantly destabilize the nucleosome array.

Figure 42. Force-extension of modified semi-synthetic nucleosome arrays

Schematic (left) of single molecule force extension experiment conducted on prepared semi-synthetic nucleosome arrays. This work was completed by and the figure is courtesy of Marek Simon from the laboratory of Dr. Michael Poirier (OSU).
Conclusions

Total synthesis is a powerful tool for the generation of homogeneous samples for biophysical characterization. Herein, we have presented a sequential native chemical ligation (NCL) strategy for the total chemical synthesis of histone H3.

The initial preparation of the H3-K56ac(R40C,S96C,C110A) protein validated the use of total synthesis for the preparation of homogeneous samples suitable for biophysical characterization. Despite the successful reconstitution of H3-K56ac(R40C,S96C,C110A) into histone octamers and nucleosomes, we found that the introduced cysteine substitutions at the sites of ligation had a larger impact on nucleosome dynamics than the introduced lysine 56 acetylation.

We therefore adapted our initial approach to incorporate post-ligation desulfurization which could convert ligation junction cysteines to native alanine residues. With this approach, we additionally see a reproducible 3-fold increase in overall synthetic yield, and can reliably generate sufficient fully synthetic protein for biochemical and biophysical characterization.

Acetylation within the entry-exit region of the nucleosome introduced by H3-K56ac attenuates DNA-protein interactions resulting in a decrease in DNA association with the histone octamer in this region. Acetylation within the entry-exit region further allows for
increased DNA site-accessibility and facilitates recognition and site-specific binding of DNA by the LexA protein. Additionally, we find that the lysine to glutamine genetic substitution commonly employed for in vivo characterization of acetylation qualitatively reproduces the effect of acetylation in the case of H3-K56. Despite significantly altering DNA-histone interactions within the entry-exit region, characterization of H3-K56ac(C110A) nucleosome arrays does not demonstrate a significant change in stability when compared to unmodified arrays. This result suggests that the primary impact of lysine 56 acetylation is localized within the entry-exit region and is not propagated throughout the DNA-octamer interface.

The described characterization of nucleosomes reconstituted with synthetically prepared H3-K56ac(C110A) protein strongly correlates with the findings of a recent study of H3-K56ac bearing nucleosomes which were prepared with modified protein generated with codon suppression. The agreement of our work with the study of recombinantly prepared H3-K56ac further validates that the observed perturbation of DNA-histone binding interactions in the entry-exit region is a direct result of lysine 56 acetylation and is not an artifact of our synthetic route to protein preparation.

Herein, we have demonstrated the successful preparation of homogeneous site-specifically modified histone H3 by a sequential NCL based total chemical synthesis strategy. Additionally, as histone H3 is the largest of the core histones, it is our belief that the presented total synthesis strategy may be adapted to the other core histones to
allow for the preparation of proteins which may be similarly modified throughout their sequence.
Acknowledgements

Thank you to Tiffany Velvet Summers for her initial synthesis of peptides N1 and C1. Thank you to Justin North for his tireless efforts in reconstituting and characterizing the H3-K56ac semi-synthetic nucleosomes. Thank you to Dr. Michael Poirier for his sage advice throughout this project and to Marek Simon for preparing and characterizing the nucleosome arrays and developing and characterizing the magnetic tweezers force extension experiments. A special acknowledgment to Dr. Jen Ottesen who would not allow me to scrap any strategy before its time and Michelle Ferdinand for just barely keeping me sane.
Chapter 4: Reversible Protection and Substitution of Dbz

Introduction

Two distinct sets of chemical protocols have been developed for solid phase peptide synthesis (SPPS) and are defined by the N-terminal protection strategy employed. The t-butoxycarbonyl (Boc) and 9-fluoronylmethoxycarbonyl (Fmoc) chemistries share a common set of activating agents and are primarily differentiated by their requirements for deprotection and cleavage. The Boc protecting group is removed with acid treatment, typically neat TFA, and following synthesis, the peptide is deprotected and cleaved from the solid support in anhydrous HF. The need for routine deprotection of resin with TFA is detrimental to automated systems and renders Boc synthesis incompatible with the introduction of acid labile amino acids including phosphorylated serine and threonine as well as tri-methylated lysine. In contrast, the Fmoc group is removed by treatment with dilute organic base, typically piperidine, and is compatible with most automated peptide synthesizers. The differences between Boc and Fmoc chemical protocols are significant enough that careful consideration is necessary before selecting which chemistry is best suited for the preparation of a peptide of interest.
In the preceding chapter, we introduced a sequential NCL scheme for the preparation of synthetic modified histone H3 proteins by the chemoselective condensation of three peptide segments prepared by Boc in situ neutralization protocols.\textsuperscript{57} We used Boc chemistry due to its compatibility with the direct synthesis of peptide thioesters necessary for NCL, but this chemistry is not suitable for the preparation of peptides incorporating amino acid phosphorylation and tri-methylation which are commonly observed and biologically important modifications in histones. Boc chemistry is incompatible with most automated peptide synthesizers and is therefore not commonly used in most peptide synthesis labs. We therefore sought to develop a strategy for the preparation of ligation capable peptide segments by Fmoc chemistry which would increase the general applicability of our total synthesis scheme.

As stated, Fmoc chemistry is not compatible with base labile species including the $\alpha$-thioester moiety which is required for peptide condensation by native chemical ligation. Rather, several groups have developed a variety of indirect methods which allow for the preparation of a peptide thioester compatible with Fmoc chemistry, including post-cleavage carboxylic acid terminus conversion,\textsuperscript{110} use of the alkanesulfonamide safety-catch linker,\textsuperscript{111} among others.\textsuperscript{112-122} Although a plethora of methods have been reported, no one technique has been widely adopted.\textsuperscript{123}

The Dawson lab has recently introduced an elegant strategy for the generation of peptide thioesters by Fmoc synthesis.\textsuperscript{58} In the Dawson method, peptides are synthesized by
Fmoc SPPS on resin substituted with the 3,4-diaminobenzoic acid (Dbz) linker. After synthesis, the protected peptide is treated with p-nitrophenylchloroformate (p-NPC), washed extensively, and addition of DIEA converts the original Dbz species into an N-acyl-benzimidazolinone (Nbz) derivative, also termed an N-acylurea. The prepared Nbz terminus is susceptible to post-cleavage thiolysis to generate the desired C-terminal α-thioester moiety (Fig. 43). The Dawson method has been reported to allow for the preparation of peptide thioesters in high yield, maintain stereochemistry of the C-terminal amino acid, eliminates the need for complex resin handling, and is compatible with common automated synthesis protocols.

Figure 43. Scheme for Dbz as a thioester auxiliary
During our initial attempts to prepare peptides derived from the C-terminal sequence of histone H4 on Dbz substituted resin, we identified that the free amine of Dbz was not sufficiently deactivated as evidenced by the significant formation of branched peptide products, and this formation was exacerbated in the synthesis of long and challenging peptides. Branched products resulting from acylation of both Dbz amines renders the peptide incapable of undergoing subsequent conversion to the acylurea form necessary for thioester formation.

We therefore developed a reversible allyloxycarbonyl (Alloc) protection strategy for the unacylated amine of Dbz which allowed us to employ more reactive activating agents and capping steps that would minimize the formation of deletion products. Importantly, incorporation and deprotection of the Alloc protecting group is orthogonal to most commonly employed protecting groups.

After successfully validating the use of Alloc as a reversible protecting strategy, we next sought to exploit the observed reactivity of the free amine of Dbz to allow for the site-specific incorporation of affinity labels and biophysical probes, such as biotin and fluorophores. Our results suggest that this could be a widely applied technique for the site-specific incorporation of biophysical probes distinct from the commonly employed lysine and cysteine modification strategies. Notably, peptides which were not modified at the Dbz amine during coupling of a label, could be converted to the Nbz form, and following purification the species could be resolved and the individual purified species utilized for multiple purposes.
Experimental Methods

General Automated Peptide Synthesis Protocols

All peptides were synthesized on an Apex 396 automated peptide synthesizer (Aapptec) except as otherwise noted. Deprotection was carried out by 2 x 5 minute treatment with 20% piperidine in NMP and the resin was washed extensively with DMF. Amino acid coupling was routinely conducted with 6 equivalents amino acid over resin loading with 5.4 equivalents activating agent (HBTU, HCTU, or HATU) in NMP with DIEA. The resin was washed and if acetyl capping was employed, the resin was mixed with capping solution for 3 minutes, washed, and the next cycle was initiated.

Synthesis of H4N and H4C peptides on unprotected Dbz resin

Peptides H4N-Dbz and H4C-Dbz were synthesized on Fmoc-Rink amide MBHA LL resin. Where noted, C-terminal Fmoc-(Dmb)Gly-OH (2.2 eq.) was manually coupled onto deprotected Dbz with HBTU activation; all other amino acids were coupled 2 x 45 minutes (6 eq. of 1 AA / 0.9 HBTU / 1.8 DIEA). All non-terminal glycine residues were introduced as the Fmoc-Gly-OH derivative. For peptide H4C, N-terminal Gly was introduced as Boc-Gly-OH. For peptide H4N, N-terminal Ser was introduced as Fmoc-
Ser(tBu)-OH to allow potential acetylation of the peptide N-terminus. Product identities were confirmed by RP-HPLC and MALDI-TOF MS.

Synthesis of LYRAGA and LYRAGF Peptides

To simplify analysis by generating only one Dbz isomer, the first amino acid was coupled to protected 4’-Alloc-Dbz-Arg MBHA LL resin for 2 x 1 hour (15 eq. of 1 AA / 0.9 HATU / 1.8 DIEA). The Alloc group was removed, revealing the 4’ amine throughout the remainder of the synthesis. Subsequent amino acids were coupled for 30 minutes (6 eq. of 1 AA / 0.9 HCTU / 1.5 DIEA). The N-terminal leucine was introduced as Boc-Leu-OH. After synthesis was complete, a single 5 minute capping cycle was carried out with an excess of capping solution (15% acetic anhydride / 15% DIEA / 70% DMF).95

Synthesis of H4N and H4C Peptides on Fmoc-Dbz(Alloc) Resin

The C-terminal Fmoc-Gly-OH residue (15 eq.) was activated with HATU (1 AA: 0.9 HATU : 1.8 DIEA) and coupled to Dbz(Alloc) derived Rink amide MBHA LL resin with 2 x 1 hour couplings. Subsequent residues were introduced with HCTU (H4C) or HBTU (H4N) activation. For peptide H4C, N-terminal Gly was introduced as Boc-Gly-OH.
Fmoc-Dbz(Alloc)-Arg derived Rink amide MBHA LL resin was prepared as described. Identity and purity of the desired Fmoc-Dbz(Alloc)-Arg species was determined to be >95% by RP-HPLC and MALDI-TOF analysis ([M+H]^+ Observed: 614.1 m/z; Expected 614.2 m/z). Following Fmoc deprotection with 20% piperidine / NMP, amino acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Glu(tBoc)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(tBoc)-OH, Fmoc-Nle-OH, Fmoc-Phe-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Trp(Trt)-OH, Fmoc-Tyr(tBu)-OH and Fmoc-Val-OH) were introduced with 2 x 1hr coupling using amino acid (15 eq.) activated with HATU / DIEA. Peptides were cleaved from resin and products were resolved by RP-HPLC on a Grace Vydac C18 column (0-45 % acetonitrile / H₂O / 0.1% TFA over 30 minutes). Product identity was confirmed by MALDI-TOF MS. Integration of the obtained RP-HPLC chromatogram at 218 nm yielded area values for the coupled and uncoupled species. Coupling efficiency was defined as a ratio of determined area values as follows:

Coupling Efficiency = (Coupled / Coupled + Uncoupled) x 100 %.
Dbz containing resins prepared on Rink amide MBHA LL base resin were Alloc protected following addition of the first amino acid to the Dbz. Reaction conditions were chosen to minimize double addition of amino acids to Dbz. Resins to be protected were swollen in DMF. Amines were activated by nutating the resin in 10% DIEA / DMF for 5 minutes. Following base activation, the resin was washed with DMF and converted first to “wet” DCM and then anhydrous DCM under a N₂ atmosphere. Resin was drained and 500 mM allyl chloroformate in anhydrous DCM was added to the resin under nitrogen. The reaction was nutated overnight at RT. The level of protection was monitored by analytical cleavage of ~ 5 mg of resin. If incomplete, Alloc protection was repeated until <5% unprotected species remained.

Alloc protection may also be completed at reduced concentrations of allyl chloroformate in the presence of one resin loading equivalent of DIEA. Complete protection (>95%) has been successfully achieved after a single 24 hour protection with concentrations as low as 300 mM allyl chloroformate under these conditions.
Synthesis of 3′-Fmoc(Dbz)

The mono-protected 3′-Fmoc(Dbz) was prepared by modifying previously reported protocols. Briefly, 3,4-diaminobenzoic acid (1 g, 6.5 mmol) was resuspended in 125 mL of a 1:1 mixture of CH$_3$CN / NaHCO$_3$ and the reaction was initiated with the dropwise addition of 15 mL of Fmoc-OSu (2.4 g, 7.1 mmol) resuspended in the same mixture. The protection reaction proceeded for 2 hours and was quenched with the addition of HCl to a final pH of 1.0 to selectively precipitate the mono- and di-Fmoc protected species. The mixture was filtered and the filtrate was resuspended in DMSO (4 mL). The 3′-Fmoc-Dbz species was selectively precipitated from the DMSO with the addition of acidified reaction buffer, washed extensively, and dried under vacuum to yield a light gray product. Product identity and purity was validated by NMR. Product was used without further purification. In a typical synthesis, 1.0 g of 3′-Fmoc(Dbz) was obtained for an overall synthetic yield of 41%.

Preparation of 3′-Fmoc-Dbz(Alloc) Resin

The small molecule, 3′-Fmoc-Dbz (1.5 eq.), was coupled to base resin with HCTU (1.4 eq.) activation in the presence of DIEA. Reaction progress was monitored by the ninhydrin test for free amines and was typically complete within 1 hour. Resultant resin
was washed with DMF and converted to DCM. Alloc protection of the resultant resin was conducted as previously described.

Alloc Deprotection

Alloc deprotection of the Dbz amine was conducted upon completion of synthesis. Fully protected resin was swollen in DCM and sparged briefly with Ar to remove oxygen and minimize evaporation. Deprotection was initiated with the addition of 0.35 molar equivalents of Pd(PPh₃)$_4$ and 20 molar equivalents of PhSiH₃ and the reaction proceeded at 25 °C with nutation for 30 minutes. Resin was washed with DCM and DMF. Complete deprotection was typically observed after a single deprotection cycle as verified by test cleavage of peptide. If deprotection was determined to be incomplete, a second cycle was employed.

Nbz conversion

Protected peptides synthesized on Dbz resin were converted to the Nbz acylurea form prior to cleavage from resin by literature methods.$^{58}$ Briefly, resins to be converted were swollen in DCM. Following the addition of 5 molar equivalents of p-nitrophenyl-chloroformate (p-NPC) to the swollen resin, the reaction mixture was nutated at 25 °C for 45 minutes. Resin was washed with DCM and converted to DMF. Resin was nutated for
15 minutes with 10% DIEA in DMF, washed with DMF, converted to DCM and lyophilized.

Automated Nbz Conversion

Automated conversion of Dbz derivatized resin to the Nbz form was completed on an Apex 396 synthesizer. The Nbz conversion was carried out on a 0.05 mmol scale. Two separate solutions of p-NPC (0.25 M in DCM) and DIEA (0.5 M in DMF) were prepared. Previously swollen resin was washed with DCM and then the p-NPC solution (1 mL, 5 eq.) was added and mixed for 45 minutes. The resin was first washed with DCM and then DMF. DIEA solution (1 mL) was added and the resin was mixed for 15 minutes. Converted resin was subsequently washed with DMF, DCM, and lyophilized.

Peptide Cleavage

Except where noted, all peptides were cleaved with a standard cleavage solution (95% TFA, 2.5% H₂O, 2.5% EDT) for one hour with nutation. All phosphothreonine containing peptides were cleaved with Reagent K (82.5% TFA, 5% H₂O, 5% phenol, 5% thioanisole, 2.5% EDT) for three hours. All crude cleavage solutions were concentrated with nitrogen, peptide was precipitated with cold anhydrous diethyl ether, and the
resultant pellets were subsequently washed 3 x with cold diethyl ether. The final precipitate was resuspended with 50% acetonitrile / H₂O + 0.1% TFA and lyophilized.

**Thioester Preparation**

Lyophilized Nbz peptides were resuspended in 6 M Gdn-HCl, 500 mM NaCl, 100 mM phosphate buffer pH 7.5 with 50 mM MESNA. Conversion to thioester was monitored by RP-HPLC and MALDI-TOF MS until complete.

**Determining Rate of Fmoc-Gly Coupling to Boc-LYRGA-Dbz**

Fmoc-Gly (0.13 M) was coupled to Boc-LYRGA-Dbz resin (10 mg) with activation (0.125 M of HATU, HBTU, HCTU, or COMU) in the presence of DIEA (0.23 M) for 2 hours to determine the optimal coupling efficiency. Coupling was repeated as above with HATU and variable DIEA concentration (0.33 or 0.54 M) to determine the effect of base concentration on coupling efficiency. Coupling of the Fmoc-Gly symmetric anhydride was also completed (conditions here). N-terminal Fmoc protection was removed with piperidine (20% v/v in NMP). Deprotected resin was washed, cleaved and lyophilized as previously described. Lyophilized crude peptide mixture was analyzed by RP-HPLC with a 0-65% acetonitrile / 0.1% TFA gradient. Integration of the obtained RP-HPLC
chromatogram at 218 nm yielded area values for the coupled and uncoupled species. Coupling efficiency was determined as previously described.

Subsequent couplings of Fmoc-Gly were conducted as described, but with more concentrated solutions. Briefly, Fmoc-Gly (0.42 M) was coupled to Boc-LYRGA-Dbz resin with HATU activation (0.4 M, 1.2 M DIEA). Resin aliquots were removed at 1, 2, and 3 hours. An additional time point was obtained at 12 hours under slightly altered conditions (0.45 M Fmoc-Gly, 0.43 M HATU, 0.82 M DIEA). Resin aliquots were deprotected, cleaved and analyzed as previously described.

**Determining Rate of Biotin Coupling to Boc-LYRGA-Dbz**

Biotin (0.4 M) was coupled to Boc-LYRGA-Dbz resin with HATU activation (0.37 M, 0.72 M DIEA). Resin aliquots were removed at 2, 4, 6 and 12 hours and the reaction was allowed to proceed. Collected aliquots were cleaved and analyzed as previously described.
Synthesis of H3T-Dbz(Gly) and H3T-Dbz(Biotin) Peptides

Biotin or Fmoc-Gly (0.36M) was coupled to peptide H3T-Dbz resin with HATU (0.34 M) activation in the presence of DIEA (0.72 M) and allowed to react for a total of 12 hours. For each coupling, aliquots of resin were removed at 2, 4, and 6 hours after reaction start. All Fmoc-Gly aliquots were deprotected with piperidine prior to cleavage. All resin aliquots were cleaved with Reagent K and peptide product was analyzed by RP-HPLC to determine the amount of coupling as previously described.

Antibody Pulldown

Assay was completed with 4 different peptides: H3K4-Dbz, H3K4-Dbz(biotin), H3K9, acH3T-Dbz(biotin) and an antibody alone control. Dynabeads MyOne Streptavidin T1 beads (10 µL) were washed 3 x 5 minutes with 400µL of pulldown buffer (10 mM Tris, pH 7.5, 200 mM NaCl). An 80 ng aliquot of each peptide was pre-incubated with 2 µg of histone H3 trimethylated Lys4 antibody (Active Motif) in 10 mM Tris pH 7.5, 200 mM NaCl with 2% BSA. Peptide and antibody where mixed on a nutator for 45 minutes and the resultant mixture was added to the washed beads and nutated for 45 minutes. Supernatant was removed and the beads were first washed with 50 µL of a 2 % BSA solution in pulldown buffer and then 3 x with 400 µL pulldown buffer.
After washing, 20 µL of 2 x SDS dye was added to the beads and the samples were boiled at 100°C for 15 minutes. A 5 µL aliquot of supernatant was resolved by SDS-PAGE for 2 hours at 180V. The gel was then transferred to a nitrocellulose membrane using the Mini Trans-Blot electrophoretic transfer cell (Bio-Rad) for 1.5 hours at 100V. The membrane was blocked with 25mM Tris, pH 7.5, 155 mM NaCl, 2% milk for 30 minutes. Goat anti-rabbit IgG secondary antibody was added at a 1:2000 dilution in buffer for 1 hour. The membrane was washed 5 x 5 minutes with 50 mM Tris pH 7.5, 155 mM NaCl, and 0.5% Tween 20. Membrane was visualized by chemiluminescence with Visualizer™ western detection reagent (Millipore) on a Typhoon Tri imager (GE Healthcare).

Synthesis of TAMRA-coupled H3T Peptides

TAMRA (5-(and 6)-carboxytetramethylrhodamine) was coupled to H3T-Dbz(Alloc) resin following deprotection of the N-terminal Fmoc with 9 eq of TAMRA-OSu and 9 eq of DIEA added to the resin. Coupling was allowed to proceed for 24 hours. After coupling, resin was washed and cleaved analytically using Reagent K. TAMRA was also coupled to acH3T-Dbz resin following deprotection of the Alloc group under identical conditions.
Synthesis of TAMRA-H3T-Dbz(biotin) Peptide

Resin that had the TAMRA added to the N-terminus was alloc deprotected according to methods. This resin was then coupled with 0.36M biotin using HATU/DIEA activation (1 : 0.9 : 2) and allowed to react for 12 hours. The resin was then cleaved analytically using Reagent K.
Results and Discussion

The direct synthesis of peptide thioesters suitable for use in native chemical ligation reactions has classically been accomplished with Boc \textit{in situ} techniques\textsuperscript{57} due primarily to the base labile nature of the thioester. However, the requirement for strong acid deprotection in Boc does not allow for synthesis automation and the introduction of important acid sensitive post-translational modifications such as amino acid phosphorylation and tri-methylation. We therefore sought to explore the use of automated Fmoc based solid phase peptide synthesis (SPPS) protocols for the preparation of ligation capable peptides suitable for the preparation of modified histones.

The thioester auxiliary 3,4-diaminobenzoic acid (Dbz), first introduced by the Dawson lab,\textsuperscript{58} has been reported to allow for the indirect preparation of peptide thioesters by automated Fmoc SPPS. The two amines of Dbz are chemically distinct due to the electronics of the aromatic ring system, and are both capable of forming an amide bond with an incoming activated amino acid during SPPS. The 3-amino group is markedly more reactive under peptide synthesis conditions and after a single round of SPPS, a 4:1 ratio of coupling is observed at the 3’ over 4’ position.\textsuperscript{58} Importantly, it has been reported that initiation of peptide chain extension from an amine on Dbz leads to sufficient inactivation of the other due to steric and electronic considerations. The remaining amine is sufficiently deactivated that extraneous acylation during peptide synthesis has been reported to be negligible. Although partially deactivated, the
unacylated Dbz amine necessarily retains enough activity that following synthesis it will react quantitatively with p-NPC which is the first step in Dbz conversion to the thiol sensitive Nbz form.

Peptide H4C (Table 3) was synthesized on Dbz substituted resin by standard automated Fmoc SPPS protocols on an Apex 396 peptide synthesizer with HBTU activation. Peptide H4C is a particularly glycine rich peptide derived from the C-terminus of histone H4. Following synthesis, the crude ligation mixture was assayed by RP-HPLC (Fig. 44).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4C</td>
<td>GRTLYGFGG</td>
</tr>
<tr>
<td>H3K4</td>
<td>ART(Km3)QTARKSTM-Dbz-R</td>
</tr>
<tr>
<td>H3K9</td>
<td>Biotin-ARTKQTAR(Kmet2)-NH2</td>
</tr>
<tr>
<td>H3T</td>
<td>RPG(pT)VALREIRRYQ(Kac)STELLIRK</td>
</tr>
<tr>
<td>acH3T</td>
<td>Ac-RPG(pT)VALREIRRYQ(Kac)STELLIRK</td>
</tr>
</tbody>
</table>

Table 3. Sequence of prepared peptides

Unexpectedly, our initial synthesis of H4C resulted in the formation of multiple products. RP-HPLC peak fractions were collected and assayed by MALDI-TOF MS to establish identity. We found a series of branched peptides larger than the expected product resulting from acylation of both amines on the Dbz. Analysis revealed that glycine has a high propensity to couple to the free amine of peptide-Dbz leading to the observed
formation of branched species. Branching also initiated at phenylalanine as demonstrated by product H4C-C3, but was significantly more likely at glycine.

<table>
<thead>
<tr>
<th>Peak</th>
<th>Product</th>
<th>Expected [M+H]^+ m/z</th>
<th>Observed [M+H]^+ m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4C-C0</td>
<td>GRTLYFGGG-Dbz-CONH₂[a]</td>
<td>1060.5</td>
<td>1060.4 ± 0.3</td>
</tr>
<tr>
<td>H4C-C1</td>
<td>GRTLYFGGG-Dbz(GRTLYFGGG)-CONH₂</td>
<td>1969.0</td>
<td>1968.6 ± 0.5</td>
</tr>
<tr>
<td>H4C-C2</td>
<td>GRTLYFGGG-Dbz(GRTLYFG)-CONH₂</td>
<td>1912.0</td>
<td>1911.5 ± 0.3</td>
</tr>
<tr>
<td>H4C-C3</td>
<td>GRTLYFGGG-Dbz(GRTLYGF)-CONH₂</td>
<td>1854.9</td>
<td>1854.7 ± 0.2</td>
</tr>
<tr>
<td>H4C-C4</td>
<td>GRTLYFGGG-Dbz(GRTLYG)-CONH₂</td>
<td>1707.9</td>
<td>1707.7 ± 0.6</td>
</tr>
<tr>
<td>H4C-C9</td>
<td>GRTLYFGGG-Dbz(G)-CONH₃</td>
<td>1117.6</td>
<td>1117.5 ± 0.1</td>
</tr>
<tr>
<td>H4C-C9</td>
<td>GRTLYFGG(Dmb)G-Dbz-CONH₂</td>
<td>1210.6</td>
<td>1210.2 ± 0.1</td>
</tr>
</tbody>
</table>

[a] Desired product

Figure 44. RP-HPLC and MALDI analysis of crude H4C-Dbz

RP-HPLC (top) analysis of H4C-Dbz synthesis with a gradient of 14-50% acetonitrile / H₂O + 0.1% TFA. Table of products identified by MALDI-TOF MS (bottom).
The small steric profile of glycine is likely to play a contributing factor in its apparent propensity to misacylate Dbz. In order to test this hypothesis, the H4C peptide was resynthesized and the C-terminal glycine residue was introduced as the commercially available Fmoc-(Dmb)Gly-OH species. Protection of the glycine terminal amine with Dmb increases the steric profile of the amino acid when compared to unprotected glycine. Following the manual coupling of Fmoc-(Dmb)Gly-OH to Dbz substituted resin with HBTU activation, subsequent rounds of peptide synthesis were completed by standard automated protocols as previously described. The presence of the bulky Dmb moiety did minimize double acylation during the initial coupling as evidenced by the disappearance of the H4C-C₁ species (Fig. 45). However, Dmb protection was not sufficient to eliminate mis-acylation of subsequent glycine residues. Fmoc-(Dmb)Gly-OH is an expensive derivative and may be incompletely removed during peptide cleavage, which renders its routine use as a substitute for Fmoc-Gly-OH impractical.

The significant reduction in the formation of branched products observed during the coupling of Fmoc-(Dmb)Gly-OH strongly supports the hypothesis that steric effects are a key discriminator for extraneous acylation of Dbz. To test this and to determine whether branched product formation only occurs when glycine is the C-terminal residue, we synthesized the test peptides Boc-LYRAGA-Dbz and Boc-LYRAGF-Dbz with HCTU activation. Following synthesis of the peptides, the resins were treated with a single round of acetyl capping solution commonly employed in manual peptide synthesis.⁹⁵
RP-HPLC analysis of H4C-Dbz synthesis on Fmoc-(Dmb)Gly-Dbz resin with a gradient of 14-50% acetonitrile / H₂O + 0.1% TFA. Individual species were identified by MALDI-TOF MS as described in Fig. 44.

Three primary products were observed for each synthesis (Fig. 46) and identified by MALDI-TOF MS (Appendix A). In addition to the desired product, acylation of a second chain during Fmoc-Gly-OH coupling (A₁: 9%; F₁: 6% by integration of RP-HPLC chromatogram) and formation of the acetylated species (A₂: 12%; F₂: 7%) were observed. It should be noted that optimized capping conditions have been reported to minimize acetylation of the unprotected Dbz amine, but any accumulation of non-productive acetylation product at each step of a long peptide synthesis would necessarily result in reduced yield. Further, the over-acylation observed during the glycine coupling cycles demonstrates that while the flexible Gly-Gly terminus likely exacerbated the extent of over-acylation in the context of peptide H4C, this side reaction does occur on more hindered termini.
Figure 46. RP-HPLC analysis of extraneous acylation on LYRAGX-Dbz resin

RP-HPLC analysis of Boc-LYRAGA-Dbz (left) and Boc-LYRAGF-Dbz (right) syntheses with a gradient of 0-65% acetonitrile / H₂O + 0.1% TFA.

The extraneous acylation of Dbz during routine amino acid coupling would be exacerbated over the course of a long peptide synthesis. This would be a particular problem in syntheses conducted under highly activating conditions and incorporating repetitive acetylation to minimize deletion product formation. We therefore sought to develop a reversible protection strategy for the unreacted amine of Dbz which would allow us to maximize the purity and yield of the desired product by minimizing non-productive Dbz acylation.

We decided to assess the feasibility of Alloc protection with allyl chloroformate based on its chemical similarity to p-nitrophenylchloroformate (p-NPC) which is integral to the
conversion of Dbz to Nbz. The Alloc group is stable to the basic and acidic conditions commonly employed in peptide synthesis by Fmoc chemistry, and the metal-catalyzed deprotection of Alloc is orthogonal to most commonly employed side chain protecting strategies, including the masked cysteine derivative thiazolidine, making it an ideal candidate for the reversible protection of Dbz.

We prepared Dbz substituted base resin and introduced the first amino acid as Fmoc-Gly-OH in minimal excess (1.1 eq.) with mild activation by HBTU to minimize formation of the branched product. The resultant resin was thoroughly washed with DMF, “wet” DCM and finally anhydrous DCM. Resin was incubated with 500 mM allyl chloroformate in anhydrous DCM and the protection reaction was allowed to proceed for 24 hours. The addition of one equivalent DIEA to initial resin loading is required to maintain the pH of the solution. A single round of protection was typically sufficient to achieve quantitative protection as determined by test cleavage, but the protection may be repeated as warranted.

We assessed the impact of Alloc protection on the quality of peptide synthesis by generating the H4C-Dbz(Alloc) peptide on the prepared resin. Following synthesis, the crude H4C-Dbz(Alloc) cleavage mixture was analyzed by RP-HPLC (Fig. 47, top). A single primary product was observed and product identity was verified by MALDI-TOF MS to be the desired H4C-Dbz(Alloc) species.
Figure 47. Alloc removal and Nbz conversion of H4C-Dbz(Alloc) peptides

RP-HPLC (left) and MALDI-TOF MS (right) analyses of peptides with a gradient of 14-50% acetonitrile / H₂O + 0.1% TFA. Top: H4C-Dbz(Alloc)-NH₂. Middle: H4C-Dbz-NH₂. Bottom: H4C-Nbz-NH₂. Expected and observed [M+H]+ for protein as described.

We next verified that Alloc deprotection was compatible with standard peptide protecting groups and quantitatively yielded the desired Dbz linker. Resin was treated with palladium and phenylsilane to remove the Alloc group. Peptides were deprotected and cleaved from a small aliquot of the resultant resin. The post-cleavage mixture was
analyzed by RP-HPLC and MALDI-TOF MS to verify quantitative Alloc deprotection (Fig 47, middle). The absence of additional side products verified that Alloc deprotection is orthogonal to the amino acid protecting groups present in the peptides tested.

After establishing that Alloc deprotection was complete, the bulk of the remaining resin was converted to the Nbz form by literature methods. Briefly, resin was treated with p-NPC in DCM, washed thoroughly and converted to DMF. The resin was next treated with DIEA in DMF to produce the desired Nbz species. Peptide was deprotected and cleaved from the resin. RP-HPLC and MALDI-TOF MS analysis of the crude product demonstrated complete conversion to the Nbz species (Fig. 47, bottom).

Automated Alloc Deprotection and Nbz Conversion

The compatibility of Fmoc chemistry with automated peptide synthesis was an important factor in our initial decision to evaluate the use of Dbz as a thioester precursor. It has been previously reported that in addition to peptide synthesis, the conversion of Dbz to Nbz may be automated. We wanted to determine if Alloc deprotection was similarly amenable to automation.

H4C-Dbz(Alloc) resin was deprotected and subsequently converted to its Nbz form on an Apex 396 peptide synthesizer. Briefly, palladium and triphenylsilane solutions were
prepared in DCM which had been gently sparged to minimize oxygen content. The individual solutions were then added to previously swollen resin under a nitrogen atmosphere and mixed with shaking on the synthesizer for 30 minutes. Following mixing, the resin was washed with DCM in preparation for Nbz conversion.

Automated conversion to Nbz of the deprotected resin was completed. A p-NPC solution prepared in DCM was added to the previously swollen resin and mixed for 45 minutes. The resin was washed extensively with DCM and then DMF. Nbz formation was initiated with the addition of DIEA in DMF and the resin was mixed for 15 minutes. Subsequent analysis of the resultant peptide following conversion demonstrated two main products (Fig. 48, top). The products were verified by MALDI-TOF MS to be the desired H4C-Nbz peptide as well as the same peptide exhibiting formylation of the Nbz. Nbz formylation has previously been reported to occur during automated conversion due to trace amounts of DMF in the reaction. Importantly, Nbz formylation does not affect subsequent thioester formation and incubation of the resultant crude peptide-Nbz mixture with MESNA resulted in quantitative conversion to the desired H4C-MESNA thioester (Fig. 48, bottom).
Figure 48. Automated Nbz conversion of H4C-Dbz and thioester formation

RP-HPLC analysis of automated H4C-Nbz formation (top). Two products are observed: H4C-Nbz(formyl), [M+H]+ observed m/z 1114.8, expected m/z 1114.5; H4C-Nbz, [M+H]+ observed 1086.2, m/z, expected 1086.5 m/z. RP-HPLC (bottom) analysis after thiolyis to generate the H4C-MESNA thioester: [M+H]+ observed 1051.0 m/z, expected 1051.4 m/z.

A significant complication to the presented reversible Alloc protection strategy is the need to couple the first amino acid to Dbz prior to protection. Although coupling conditions can be tightly regulated to minimize extraneous acylation of Dbz, different batches of resin would need to be prepared for each C-terminal amino acid and optimal
coupling conditions would similarly needed to be determined. It would be preferable to generate a single Dbz(Alloc) base resin that could subsequently be loaded with the desired C-terminal amino acid.

The first step to preparing a general use Dbz(Alloc) base resin was to prepare the singly protected 3'-Fmoc-Dbz species by modified literature protocols. NMR analysis of the resultant product demonstrated that the protecting reaction was highly selective for the 3'-Fmoc-Dbz species with negligible 4'-Fmoc-Dbz detectable. The prepared 3'-Fmoc-Dbz was directly coupled to amine substituted resin with standard SPPS protocols, and the resultant base resin was treated with allyl chloroformate in the presence of DIEA to generate the desired 3'-Fmoc-Dbz(Alloc) base resin. A significant advantage of generating the 3'-Fmoc-Dbz(Alloc) species is that peptide chain extension is limited to a single position on Dbz; thereby, eliminating the mixture of 3’ and 4’ products observed when peptide synthesis is conducted directly on unprotected Dbz.

We next determined the coupling efficiency of an array of Fmoc protected amino acids to the prepared base resin. Aliquots of the 3’-Fmoc-Dbz(Alloc) substituted resin were deprotected by standard automated protocols. To ensure maximum coupling efficiency, each amino acid tested was coupled to the resin in 15-fold molar excess with HATU activation for 1 hour and the coupling cycle was repeated.
Following cleavage, coupling efficiency for each amino acid was determined by integration of RP-HPLC chromatograms (Table 4 and Appendix B). For all amino acids tested, coupling efficiency was determined to be > 95% with the exception of the β-branched amino acids valine and isoleucine. The large steric bulk immediately adjacent to the Dbz ring is likely to be the source of this incomplete coupling. It is notable that valine and isoleucine thioester termini are not preferred in ligation reactions due to poor kinetics of the condensation reaction at these sites. However, due to sequence restrictions, a Val ligation site is used in our sequential NCL scheme to generate synthetic modified histone H3.

The difficulty of incorporating valine is not particularly surprising as we have previously found that coupling of valine to unprotected Dbz does not go to completion when HBTU is used as the coupling agent (Chapter 5). Additionally, coupling of Fmoc-Val-OH to unprotected Dbz with HATU activation does not lead to detectable formation of branched products observed in glycine coupling under less activating conditions. Therefore, for the synthesis of peptides with a C-terminal valine or isoleucine residue, addition of the terminal amino acid to unprotected Dbz followed by subsequent Alloc protection would be appropriate.
<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>% Loading (^{[a]})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Arginine (Pbf)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Glutamic Acid (OtBu)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Glutamine (Trt)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Glycine</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Histidine (Trt)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>33 %</td>
</tr>
<tr>
<td>Leucine</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Lysine (Boc)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Norleucine</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Serine (tBu)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Threonine (tBu)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Tryptophan (Boc)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Tyrosine (tBu)</td>
<td>&gt;95 %</td>
</tr>
<tr>
<td>Valine</td>
<td>47 %</td>
</tr>
<tr>
<td>Valine (^{[b]})</td>
<td>&gt;95 %</td>
</tr>
</tbody>
</table>

\(^{[a]}\) % Loading determined by integration of RP-HPLC chromatograms. Coupling was carried out with 15-fold excess amino acid and HATU activation for 2 x 1 hour except where noted.

\(^{[b]}\) Fmoc-Val-Dbz-Arg resin was subsequently protected to generate Fmoc-Val-Dbz(Alloc)-Arg.

Table 4. Percent loading of amino acids on Dbz(Alloc) resin

Exploiting the Activity of Dbz for Peptide Modification

The residual activity of the unsubstituted amine on Dbz during peptide synthesis, which allows for Alloc protection, Nbz conversion or extraneous acylation, suggests that the opportunity exists to exploit the Dbz linker to introduce other functional groups site-
specifically into peptides. We sought to identify the specific conditions that would allow for the optimized addition of biophysical probes, including biotin and fluorophores, onto a peptide-Dbz support following synthesis.

As a test system, we used peptide H3K4, a short peptide derived from the 11 N-terminal residues of histone H3 tri-methlyated on lysine 4 which had been prepared to use as a thioester for other purposes. The peptide was synthesized on Fmoc-Met-Dbz(Alloc) base resin with standard automated SPPS protocols. Following synthesis, the Alloc protecting group was removed to reveal a free amine on Dbz. We have previously shown that coupling to the free Dbz amine on peptide-Dbz derived resin is strongly affected by steric factors adjacent to the site of incorporation. In the synthesis of H3K4-Dbz(Biotin), the C-terminal residue of the peptide was methionine which would be expected to render the free Dbz amine less reactive than if glycine were present at the terminus.

Biotin was coupled to the resin with HATU activation for 12 hours following pre-activation, and the reaction was repeated twice to maximize yield. Cleavage of peptide from the resin and subsequent RP-HPLC analysis of the crude product demonstrated a complex mixture of species (Fig. 49, left). Identification of individual peaks was determined by MALDI-TOF MS. Although a significant amount of unsubstituted H3K4-Dbz peptide remain, four fractions were identified as H3K4-Dbz peptides substituted with biotin. The presence of four resolved species is due to a mixture of the oxidized and non-oxidized methionine forms of the 3’-biotin-4’-H3K4-Dbz and 3’-H3K4-4’-biotin-
Dbz species. Biotin substituted H3K4-Dbz peptides were purified out of the crude mixture and the four distinct forms were pooled together (Fig. 49, right). A total of 5.4 mg of the purified H3K4-Dbz(biotin) peptide pool was obtained.

We conducted reduction of the purified peptide mixture with NaI / DMS\textsuperscript{107} and resolved a doublet of peaks corresponding to the 3’-biotin-4’-H3K4-Dbz and 3’-H3K4-4’biotin-Dbz peptides containing methionine in its non-oxidized form (Fig. 50).

Figure 49. Biotin may be introduced onto crude H3K4-Dbz peptide

<table>
<thead>
<tr>
<th>Peak</th>
<th>Sequence</th>
<th>Expected m/z</th>
<th>Observed m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>ART(Kmet\textsuperscript{O})QTARKST[Met(O)]-[Dbz(biotin)]-R</td>
<td>1952.0</td>
<td>1951.8 ± 0.5</td>
</tr>
<tr>
<td>3,4</td>
<td>ART(Kmet\textsuperscript{O})QTARKST[Met]-[Dbz(biotin)]-R</td>
<td>1937.1</td>
<td>1936.3 ± 0.7</td>
</tr>
</tbody>
</table>

RP-HPLC characterization of H3K4-Dbz(biotin) crude (left) and purified (right) peptide mixtures were resolved on a 0-36% acetonitrile / H\textsubscript{2}O + 0.1% TFA gradient. Collected fractions were assayed by MALDI-TOF MS as described (bottom).
Figure 50. RP-HPLC of reduced H3K4-Dbz(biotin)

RP-HPLC characterization of purified and reduced H3K4-Dbz(biotin) peptide mixture resolved on a 0-36% acetonitrile / H₂O + 0.1% TFA gradient. Peaks identified as described in Fig. 49.

Boc-LYRGA-Dbz substituted resin was prepared to assess the parameters of Fmoc-Gly-OH coupling to the free amine on Dbz. We performed a series of coupling reactions in which the identity of the activating agent was varied, but the concentrations of amino acid, coupling agent, and DIEA were kept constant. Following pre-activation of the coupling solution, it was added to fully protected Boc-LYRGA-Dbz resin and the coupling reaction was allowed to proceed for 1 hour. Following peptide cleavage, the crude mixture was resolved by RP-HPLC and coupling efficiency was determined by integration of the chromatogram (Fig. 51).
Figure 51. Effect of coupling agent on yield of LYRGA-Dbz(Gly)

Coupling of Fmoc-Gly-OH to Boc-LYRGA-Dbz resin was conducted with variable coupling agents as described in each panel. RP-HPLC characterization of crude cleavage product resolved on a 0-66% acetonitrile / H$_2$O + 0.1% TFA gradient and the glycine coupled product is identified (*). Table of % loading of Fmoc-Gly onto Dbz as determined by integration of presented chromatograms.

<table>
<thead>
<tr>
<th>Activating Agent</th>
<th>% Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>HATU</td>
<td>21.8</td>
</tr>
<tr>
<td>HBTU</td>
<td>2.6</td>
</tr>
<tr>
<td>HCTU</td>
<td>3.6</td>
</tr>
<tr>
<td>COMU</td>
<td>9.8</td>
</tr>
<tr>
<td>DIC / Symmetric Anhydride</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Coupling efficiency was determined to be greatest with HATU and diisopropyl-carbodiimide symmetric anhydride activation. The significant decrease in observed coupling when HCTU or COMU were used as activating agents was of particular note.
Although almost as efficient as HATU, symmetric anhydride formation requires twice as much material to maintain the same concentration of activated species, involves additional pre-activation workup, and is known to be susceptible to amino acid racemization. We therefore used HATU activation in all further studies.

To further establish the optimal conditions for directed acylation of the second Dbz amine, we tested role of base concentration in coupling efficiency. We monitored the effect of base by varying the equivalents of DIEA over amino acid while maintaining constant Fmoc-Gly-OH and HATU concentrations (Fig. 52). Interestingly, we found that the maximal incorporation of Fmoc-Gly-OH was attained at 1.8 eq. DIEA, the lowest excess tested, in contrast to previous reports that acylation of the second Dbz amine may be reduced by maintaining a minimal excess of DIEA over amino acid. It is possible that a lower fold excess of DIEA might limit coupling.

While determining the optimal coupling parameters, the concentrations of amino acid and activating agents were held lower than standard SPPS protocols to allow for an observable range of coupling efficiencies. Having identified that HATU in conjunction with 1.8 equivalents DIEA yielded maximal acylation, we next sought to maximize total coupling by using these agents at a higher concentration to increase yield of the desired product.
Figure 52. Effect of DIEA concentration on coupling efficiency

Coupling of Fmoc-Gly-OH to Boc-LYRGA-Dbz resin was conducted with variable amounts of DIEA as described. RP-HPLC characterization of crude cleavage products were resolved on a 0-27% acetonitrile / H2O + 0.1% TFA gradient and % loading of Fmoc-Gly onto Dbz as determined by integration of presented chromatograms.

Fmoc-Gly-OH was coupled to Boc-LYRGA-Dbz resin with HATU activation for a total of 3 hours, with aliquots removed at hourly time points. An additional time point was prepared by coupling Fmoc-Gly-OH to resin for 12 hours in a separate reaction. Following Fmoc deprotection and subsequent cleavage, the resultant crude peptide mixtures were analyzed by RP-HPLC and coupling efficiency was determined by integration of the resultant chromatograms (Fig. 53).
An initial burst of reactivity lead to 62% coupling after 1 hour and the rate of coupling was significantly reduced over the following two hours. At 12 hours, 94% addition of glycine to Dbz was detected. We anticipate that maximal yield may be achieved by employing 2 or more shorter coupling cycles of Fmoc-Gly-OH rather than a single 12 hour reaction.

Figure 53. Fmoc-Gly coupling to LYRGA-Dbz

Coupling of Fmoc-Gly-OH to Boc-LYRGA-Dbz was completed at a higher concentration of amino acid and HATU. RP-HPLC characterization of crude cleavage products resolved on a 0-66% acetonitrile / H2O + 0.1% TFA gradient was conducted at various time points. The (*) corresponds to the coupled product, LYRGA-Dbz(Gly).
While biotin couples through a carboxylic acid, reactivity is significantly different than a standard α-amino acid. Therefore, the coupling efficiency of biotin directly to the available amine on Boc-LYRGA-Dbz substituted resin was probed. Concentrations were necessarily reduced due to the limited solubility of biotin in DMF. Time points were collected and biotin incorporation was quantitated following cleavage of resin and subsequent RP-HPLC analysis as previously described. At 12 hours, approximately 65% biotin incorporation was observed (Fig. 54).

![Figure 54. Biotin coupling to LYRGA-Dbz](image)

Coupling of biotin to Boc-LYRGA-Dbz was completed with HATU activation as described. RP-HPLC characterization of crude cleavage products resolved on a 0-66% acetonitrile / H₂O + 0.1% TFA gradient was conducted at various time points. The (*) corresponds to the coupled product, LYRGA-Dbz(biotin).
We next tested the incorporation of Fmoc-Gly and biotin on acH3T-Dbz resin in which the H3T peptide was acetylated at the N-terminus. The C-terminal Lys(Boc) moiety adjacent to the Dbz is representative of a typical ligation terminus with increased steric bulk relative to a C-terminal alanine. Fmoc-Gly-OH and biotin couplings were completed as previously described and coupling efficiency was similarly determined from integration of RP-HPLC chromatograms. When compared to peptides which contain the Gly-Dbz and Ala-Dbz, coupling was reduced, but a maximal incorporation of 80% Fmoc-Gly and 52% biotin was observed (Fig. 55).

![Graph showing coupling efficiency over time](image)

Figure 55. Fmoc-Gly and biotin coupling to acH3T-Dbz

The coupling of Fmoc-Gly-OH (♦) and biotin (●) to acH3T-Dbz resin was completed with HATU activation and the rate of coupling was monitored by analytical cleavage followed by RP-HPLC analysis as described. The % loading for each time point was determined by integration of the RP-HPLC trace at A218 and plotted above.
The partially biotinylated resin was treated with Nbz conversion protocols. Cleavage of the crude peptide demonstrated that Nbz conversion was quantitative with no apparent effect on the biotinylated sample as determined by MALDI-TOF MS analysis. We subsequently purified the crude peptide mixture on a Supelco Discovery Widebore C18 column which allowed us to resolve the Nbz form of the peptide from the biotinylated in good purity and yield (Fig. 56). The modified amino acids typically used in the synthesis of histone N-terminal tails are often very costly. One practical application for this type of “one-pot” reaction would be functional recovery of pools of peptides for use in binding assays at the peptide level or for incorporation into histones or nucleosomes.

To verify that biotin conjugated to Dbz is accessible despite the steric restrictions of the linker, four independently purified peptide samples were assessed for interaction with a rabbit IgG antibody specific for histone H3 lysine 4 tri-methylation (Active Motiff) in the presence of BSA: H3K4-Dbz(biotin), H3K4-Dbz, biotin-H3K9, and acH3T-Dbz(biotin). Bound proteins were resolved by SDS-PAGE gel, transferred to a nitrocellulose membrane, and probed with goat anti-rabbit IgG antibody. Only the lane which contained the H3K4-Dbz(biotin) peptide demonstrated a positive signal for antibody demonstrating the successful pull-down (Fig. 57). Importantly, the H3K4-Dbz peptide which lacked biotin was not retained in the pull-down assay and no cross reactivity was observed between the antibody and the di-methylated biotin-H3K9 peptide.
Figure 56. RP-HPLC of one-pot acH3T-Nbz and acH3T-Dbz(biotin) preparation

RP-HPLC analysis of one-pot preparation of acH3T-Nbz (left) and acH3T-Dbz(biotin) (right) peptides. Pure peptides were resolved on a 27-54% acetonitrile / H₂O + 0.1% TFA gradient. The crude peptide mixture (bottom) was resolved on an 18-45% acetonitrile / H₂O + 0.1% TFA gradient.
The reported substitution of Dbz with glycine and biotin demonstrates the utility of the Dbz moiety as a site-specific site for the introduction of biophysical probes. We next sought to determine if the orthogonal protection strategies of the Fmoc protected N-terminus and the Alloc protected amine of Dbz would allow for the combinatorial assembly of multiply labeled peptides in a highly regulated manner (Fig. 58).
The double substitution of Fmoc-H3T-Dbz(Alloc) substituted resin was initially completed by coupling TAMRA-OSu (9 eq.) to the deprotected N-terminal amine in the presence of DIEA (9 eq.) for 24 hours. Quantitative addition of TAMRA was observed and following Alloc deprotection, biotin was coupled to the free amine of Dbz with HATU activation for 12 hours. Resultant peptide was deprotected and cleaved from the resin, and the crude peptide mixture was resolved by RP-HPLC (Fig. 59, top). Peak identity was confirmed by MALDI-TOF MS. The desired product, TAMRA-H3T-Dbz(biotin), resolves as two independent peaks due to the incorporation of two distinct
TAMRA isoforms. The doubly-labeled peptide was purified by RP-HPLC and the purified peptide pool was assessed by RP-HPLC and MALDI-TOF MS to confirm purity and identity (Fig. 59, bottom).

Figure 59. RP-HPLC characterization of TAMRA-H3T-Dbz(biotin)

RP-HPLC analysis of TAMRA-H3T-Dbz(biotin). The two isoforms of the desired product are identified (1,2) and were resolved on a gradient of acetonitrile / H₂O + 0.1% TFA: (top) crude mixture 0-66%; (bottom) RP-HPLC purified peptide 27-54%. MALDI-TOF MS analysis of pure sample: [M+H]⁺ observed 3678 m/z, expected 3680 m/z.

To test the feasibility of reversing the order of labeling, TAMRA was coupled directly onto acH3T-Dbz substituted resin as TAMRA-OSu as previously described. Subsequent cleavage and RP-HPLC analysis of the crude peptide mixture demonstrated no detectable
incorporation of TAMRA onto Dbz (Fig. 60). Although glycine and biotin molecules have been successfully coupled to Dbz, this result was not surprising due to the significant steric bulk of the TAMRA fluorophore.

Figure 60. TAMRA does not couple efficiently to free amine of acH3T-Dbz

RP-HPLC analysis of TAMRA coupling to acH3T-Dbz resin. The crude peptide mixture was resolved on a gradient of 18-45% acetonitrile / H₂O + 0.1% TFA. The major species was identified by MALDI-TOF MS to be unreacted acH3T-Dbz.

If the observed lack of TAMRA incorporation on Dbz was due largely to steric interference, an alternative synthetic route would be the introduction of a less hindered linker that could be derivatized for the subsequent incorporation of bulkier groups. To test this hypothesis, Fmoc-Gly-OH was independently coupled to ac-H3T-Dbz and biotin-H3T-Dbz substituted resins with HATU activation for 3x 1.5 hours. Fmoc deprotection yields a new free primary amine less occluded by the Dbz structure which was targeted for introduction of TAMRA-OSu with coupling for 24 hours as previously
described. The crude peptide mixture for each reaction was assayed by RP-HPLC and collected peaks were identified by MALDI-TOF MS (Fig. 61). The desired substituted products were identified and resolved as two isoforms as expected.

![Figure 61](image_url)

Figure 61. Characterization of doubly-labeled peptides

RP-HPLC (top) analysis of TAMRA coupling to (left) ac-H3T-Dbz(Gly) and (right) biotin-H3T-Dbz(Gly) resin. The crude peptide mixtures were resolved on a gradient of 0-66% acetonitrile / H2O + 0.1% TFA. Products were identified by MALDI-TOF MS (bottom) as described.
Conclusions

The use of Alloc protection of Dbz greatly increases the synthetic purity and yield of peptides synthesized with Dbz on the terminus. This lays the groundwork for the synthesis of peptide thioesters suitable for sequential NCL of modified histone proteins while providing the benefits of automated peptide synthesis.

Not only is the second amine of Dbz insufficiently deactivated to prevent acylation during chain extension, we have demonstrated it is sufficiently activated to allow controlled substitution. We have exploited the residual activity of Dbz as a synthetic route to introduce biotin, an affinity label, and TAMRA, a fluorophore, as functional labels. Partially reacted Dbz resin may additionally be converted to Nbz and the two species may be resolved thereby allowing the one-pot generation of the Nbz and derivatized species. It is therefore practical to utilize peptides synthesized on Dbz for multiple purposes including but not limited to ligation. We further demonstrate this in the context of the H3K4 peptide, a resin synthesized for ligation purposes but rescued from the freezer for introduction of biotin. Not described in this document is the use of this peptide to study specific interaction partners of histone H3 tri-methylated lysine 4. The N-terminal tails of histone proteins are typically heavily modified. These modification sites often serve as binding sites to recruit active factors to chromatin, and these interactions are often assessed by peptide binding assays. Our approach allows the limited use of costly modified amino acids to simultaneously generate labeled peptides
suitable for peptide assays and for ligation to generate the corresponding full-length histone protein.
Acknowledgements

The experiments presented herein were designed, performed, and characterized either directly by myself or in close collaboration with several members of the Ottesen laboratory. Dr. Santosh Mahto determined the optimal conditions for Alloc protection and deprotection, and synthesized the H4C peptides characterized herein. C.J. Howard developed the synthetic protocols for the small molecule Dbz derivatives and for loading of the Fmoc-Dbz(Alloc) resin. The functional labeling of peptide-Dbz resins was developed along with C.J. Howard and Sarah Dreher, with assistance from Christina Knight. Furthermore, Sarah Dreher assisted in troubleshooting the functional antibody interaction assay.
Chapter 5: Preparation of Synthetic Modified Histones by Fmoc SPPS

Introduction

In Chapter 3, the total chemical synthesis of histone H3-K56ac(C110A) by sequential native chemical ligation (NCL) was described. In this synthesis, all peptides were generated by solid phase peptide synthesis (SPPS) with Boc in situ neutralization protocols.\textsuperscript{57} Boc chemistry was used primarily due to its compatibility with the direct synthesis of peptide α-thioesters. However, the need to manually synthesize each peptide segment as well as the inability to generate phosphorylated or tri-methylated peptides, which are common histone post-translational modifications (PTMs), significantly limits the general applicability of a total synthesis strategy employing Boc chemistry. To increase the types of modifications which may be incorporated into synthetic histones by sequential NCL, as well as make the technique more accessible to a larger number of researchers, we sought to develop a practical total synthesis strategy utilizing peptides prepared by the more commonly employed Fmoc synthetic protocols.

The N-terminal Fmoc group is removed by treatment with organic base, typically piperidine diluted in synthesis solvent. As dilute base is less corrosive than neat TFA,
Fmoc based synthesis is compatible with most commercially available automated peptide synthesizers which is a key determinant in its general use. Additionally, Fmoc chemistry is amenable to the incorporation of acid labile moieties including phosphorylation and trimethylation. Despite these strengths, Fmoc synthesis is not compatible with the direct synthesis of peptides bearing a C-terminal α-thioester due to its base labile nature.

In chapter 4, we described the use of the 3,4-diaminobenzoic acid (Dbz) linker as a masked thioester, pioneered by the Dawson laboratory (Fig. 62). We demonstrated that extraneous acylation of the free amine of Dbz may occur during SPPS which leads to the accumulation of non-productive branched peptides, and introduced Alloc protection of the extraneous amine as an improved synthetic route to peptide thioesters by Fmoc-SPPS.

Figure 62. Scheme for Dbz as a thioester auxiliary
In this chapter, we describe the development of synthetic methods for the preparation of peptide thioesters suitable for sequential NCL to generate modified histone H3 using automated Fmoc SPPS protocols. The presented approaches were developed simultaneously with the development of reversible Alloc protection for Dbz presented in chapter 4.

We first describe our initial attempts to generate the challenging peptide thioesters segments necessary for the total synthesis of histone H3 by automated Fmoc SPPS protocols on unprotected Dbz substituted resin. These initial syntheses, though successful, resulted in low yields of the desired product due to the presence of significant deletion products which were difficult to resolve during RP-HPLC purification. Syntheses were initially optimized through the coupling of select pseudoproline dipeptides in regions known to be synthetically challenging in an effort to minimize deletion product formation resulting from secondary structure formation during synthesis.

In subsequent syntheses, reversible Alloc protection of the free amine of Dbz allowed for the use of multiple coupling cycles, stronger coupling agents, and the introduction of post-coupling acetylation steps. These chemically optimized syntheses resulted in a significant reduction in observed side products. Overall, the presented optimized synthesis strategies generated a reproducible increase in pure peptide yield over the initial syntheses on unprotected Dbz.
Finally, the optimized automated synthesis protocols were used for the preparation of a library of variably modified peptides derived from the H3 sequence. The prepared peptides are suitable for use in sequential NCL, and a subset of the Fmoc synthesized peptides were used to prepare fully synthetic histones H3(C110A) and H3-K56ac(C110A). Combinatorial ligation using the full library of prepared peptides could yield up to 24 distinct variably modified histone proteins bearing biologically important modifications (Fig. 63).

Figure 63. Scheme for combinatorial library of modified histone H3
Experimental Methods

Synthesis of H3 C-terminal Peptides (Residues 91-135)

All peptides were generated with automation on an Aapptec Apex 396 synthesizer. Syntheses were conducted on a 0.05 mmole scale on Fmoc-Ala-Wang LL resin with 6 molar equivalents of amino acid and 5.4 molar equivalents of activating agent in the presence of 9 molar equivalents of DIEA in NMP. Completed syntheses employed HBTU / 6-Cl HOBt or HCTU as activating agents. A typical synthetic cycle was as follows:

3 x 5 minute deprotection with 20% piperidine / NMP
6 x 1 minute resin wash with DMF
AA coupling for 30 min
3 x 1 minute resin wash with DMF
AA coupling for 30 min
3 x 1 minute resin wash with DMF
Capping of resin for 3 min with acetic anhydride, 6-Cl HOBt in DMF / DCM
3 x 1 minute resin wash with DMF

Residue 120 of H3(C110A) was introduced as either the native methionine with Fmoc-Met-OH or as the hydrocarbon Met mimic norleucine (Nle) with Fmoc-Nle-OH (Novabiochem). Acetylated lysine was introduced as Fmoc-Lys(ac)-OH (Aapptec) and the N-terminal residue was introduced as Boc-thiazolidine-OH (Bachem).
Alloc Protection of Amino Acid Substituted Resin

Dbz containing resins prepared on Rink Amide LL base resin were Alloc protected following addition of the first amino acid to the Dbz. Reaction conditions were chosen to minimize double addition of amino acids to Dbz. Resins to be protected were swollen in DMF. Amines were activated by nutating the resin in 10% DIEA / DMF for 5 minutes. Following base activation, the resin was washed with DMF and converted first to “wet” DCM and then anhydrous DCM under a N₂ atmosphere. Resin was drained and 500 mM allyl chloroformate in anhydrous DCM was added to the resin under nitrogen. The reaction was nutated overnight at RT. The level of protection was monitored by analytical cleavage and if necessary, Alloc protection was repeated until < 5% unprotected species remained.

Synthesis of H3 Middle Peptide Segments (Residues 47-90)

All peptides were generated with automation on an Aapptec Apex 396 synthesizer. Syntheses were conducted on a 0.05 mmole scale on prepared di-Fmoc-Dbz, Fmoc-Met-Dbz(Alloc), or Fmoc-Nle-Dbz(Alloc) derived Rink amide MBHA LL resins with 6 molar equivalents of amino acid and 5.4 molar equivalents of HCTU in the presence of 9 molar equivalents of DIEA in NMP. A typical synthetic cycle was as follows:
3 x 5 minute deprotection with 20% piperidine / NMP
6 x 1 minute resin wash with DMF
AA coupling for 30 min
3 x 1 minute resin wash with DMF
AA coupling for 30 min
3 x 1 minute resin wash with DMF
Capping of resin for 3 min with acetic anhydride, 6-Cl HOBut in DMF / DCM
3 x 1 minute resin wash with DMF

Acetylated lysine was introduced as Fmoc-Lys(ac)-OH. Tri-methylated lysine was introduced as Fmoc-Lys(me³)-OH (Novabiochem). Serine 57 and threonine 58 were introduced as the pseudoproline dipeptide Fmoc-Ser(tBu)-Thr(ψMe,Me pro)-OH (2.2 eq. over resin loading), and the N-terminal residue was introduced as Boc-thiazolidine-OH (Bachem) as described. Peptide identity and purity were confirmed using RP-HPLC and MALDI-TOF MS.

Synthesis of H3 N-terminal Peptide Segments (Residues 1-46)

All peptides were generated with automation on an Aapptec Apex 396 synthesizer. Syntheses were conducted on a 0.05 mmole scale on prepared Fmoc-Val-Dbz or Fmoc-Val-Dbz(Alloc) derived Rink amide MBHA LL resins with 6 molar equivalents of amino acid and 5.4 molar equivalents of HBTU or HCTU in the presence of 9 molar equivalents of DIEA in NMP. A typical synthetic cycle was as follows:
3 x 5 minute deprotection with 20% piperidine / NMP
6 x 1 minute resin wash with DMF
AA coupling for 30 min
3 x 1 minute resin wash with DMF
AA coupling for 30 min
3 x 1 minute resin wash with DMF
Capping of resin for 3 min with acetic anhydride, 6-Cl HOBt in DMF / DCM
3 x 1 minute resin wash with DMF

Acetylated lysine was introduced as Fmoc-Lys(ac)-OH. Tri-methylated lysine was introduced as Fmoc-Lys(me³)-OH (Novabiochem). Peptide identity and purity were confirmed using RP-HPLC and MALDI-TOF MS.

Removal of the Alloc Protecting Group

Alloc deprotection of the Dbz amine was conducted upon completion of synthesis. Fully protected resin is swollen in DCM and sparged briefly with Ar to minimize evaporation. Reaction is initiated with the addition of 0.35 molar equivalents of Pd(PPh₃)₄ and 20 molar equivalents of PhSiH₃. Deprotection proceeds at 25 °C with nutation for 30 minutes under an argon atmosphere. Resin is washed with DCM and DMF. Complete deprotection was typically observed after a single deprotection cycle as verified by test cleavage of peptide. If deprotection was determined to be incomplete, a second cycle was employed.
Conversion of Peptide-Dbz to the Nbz Form

Protected peptides synthesized on Dbz resin were converted to the Nbz acylurea form prior to cleavage from resin. Resins to be converted were swollen in DCM. Following the addition of 5 molar equivalents of p-nitrophenylchloroformate (p-NPC) to the swollen resin, the reaction mixture was nutated at 25 °C for 45 minutes. Resin was washed with DCM and converted to DMF. Resin was nutated for 15 minutes with 10% DIEA in DMF, washed extensively with DMF, converted to DCM, and lyophilized.

Side Chain Deprotection and Cleavage from Resin

Peptides were deprotected and cleaved from resin under acidic conditions. All peptides were cleaved with 95% TFA, 2.5% H₂O, 2.5% TIS solution for a minimum of two hours except where noted. Peptides which contained Met or Thp were cleaved with 94% TFA, 2.5% H₂O, 2.5% EDT, 1% TIS. The pT45 containing N-terminal peptide was cleaved with Reagent K (82.5% TFA, 5% thioanisole, 5% H₂O, 5% phenol and 2.5% ethane dithiol).

Following cleavage, resin was filtered away and washed with excess TFA to maximize peptide yield. TFA was evaporated with nitrogen and peptide was precipitated with the addition of 10-fold volume excess of cold anhydrous diethylether. Precipitated peptide
was pelleted with centrifugation and the pellet was washed a minimum of 4 times with fresh cold diethyl ether. The final washed pellet was air dried, resuspended with 50% ACN / 50% H$_2$O w/ 0.1% TFA and lyophilized. Crude yield was determined by weight following lyophilization.

Peptide Purification

Peptides were purified by RP-HPLC on Supelco Widebore C18 preparative or semi-preparative columns with an acetonitrile / H$_2$O gradient in the presence of 0.1% TFA. Purification gradients applied were adapted for optimal purification of each peptide. Collected fractions were assayed by MALDI-TOF and fractions deemed to be >95% pure were pooled and subsequently lyophilized. Identity and purity of peptide pools was determined by MALDI-TOF and RP-HPLC. Synthetic yields were calculated.

Ligation to Prepare Synthetic Truncated H3(A47C,S96C) Residues 47-135

Ligation to generate histone H3 residue 47 to 135 products was accomplished by combining desired purified middle peptide containing activated Nbz with 2 molar equivalents of desired C-terminal peptide. Peptides were resuspended in 6 M Guanidine, 100 mM phosphate pH 7.5, 750 mM NaCl, 75 mM MESNA, 10 mM TCEP at ~ 30 mg /
mL total protein concentration. Combined mixture was nutated at 25 °C for a minimum of 2 days. Reaction progress was monitored by RP-HPLC and SDS-PAGE. Ligation reactions were allowed to proceed until no further product formation was observed.

The N-terminal thiazolidine residue was unmasked to yield a terminal cysteine with the addition of methoxylamine and the resultant reduction of the reaction pH to ~4. Cysteine conversion was allowed to proceed with nutation at 25°C for a minimum of 6 hours and reaction progress was monitored by RP-HPLC and MALDI-TOF MS. The resultant ligation product corresponding to synthetic histone H3(A47C,S96C) residues 47-135 was purified by RP-HPLC, fractions identified to be >95% pure for the ligation product were pooled, and the final purified protein pool was lyophilized.

Ligation to Prepare Full Length H3(A47C,A91C,C110A)

Purified ring open first ligation product was resuspended in 6 M Guanidine, 500 mM NaCl, 100 mM phosphate, pH 7.5 at > 10 mg/mL concentration along with a minimum 5 fold molar excess of N-terminal peptide. Ligation was initiated with the introduction of 50 mM MESNA or MPAA co-thiol. Reaction progress was monitored by RP-HPLC and SDS-PAGE. Ligation reactions were allowed to proceed until no further product formation was observed, typically 5 days. The reaction pH was closely monitored and
maintained with the addition of NaOH as needed. To minimize oxidation of sulfur and the formation of disulfide bonds, aliquots of 10 mM TCEP were added daily.

Unexpected conversion of the N-terminal Cys to its protected thiazolidine form was observed over the course of several ligations. Reversion to the desired Cys was initiated with the addition of 200 mM methoxylamine HCl. Reaction progress was monitored by RP-HPLC and allowed to proceed until no thiazolidine species was observed, typically three hours. Adjusting the pH of the mixture to 7.5 with NaOH reinitiated the ligation reaction which was allowed to proceed.

Desulfurization of Synthetic H3 Proteins

Crude ligation mixtures were desulfurized to yield the native C110A product. If the ligation was completed with MPAA as the co-thiol, it was necessary to remove this species by dialysis into ligation buffer prepared without MPAA or MESNA due to the known incompatibility of MPAA with desulfurization. A 0.5 volume equivalent of 1.2 M TCEP was added to the crude ligation mixture to a final concentration of 400 mM TCEP and a minimal MESNA concentration of 75 mM was maintained. The resultant mixture was sparged with argon for a minimum of 30 minutes and the reaction was initiated with the addition of VA-044US (Wako Chemicals) to a final concentration of 10 mM. The
resulting mixture was incubated at 44°C for a minimum of four hours. Reaction progress was monitored by MALDI-TOF.
Results and Discussion

The success of the sequential NCL strategy for the total synthesis of histone H3-K56ac(C110A) presented in Chapter 3 was tempered by the low observed yields of the middle peptide segment and the full-length product. Additionally, peptide synthesis with Boc in situ neutralization limits the classes of modifications which may be introduced into synthetic histones by this method and is not compatible with standard peptide synthesis automation. To establish a more generally applicable scheme for the preparation of synthetic modified histone proteins, it was necessary to adapt the initial strategy to allow for the synthesis of ligation compatible peptide segments by automated Fmoc SPPS.

All synthetic peptides segments (Table 5) are derived from the commonly studied modified X. laevis sequence which contains the alanine for cysteine substitution at position 110 (C110A). This substitution was necessitated by the use of desulfurization to convert ligation junction cysteine residues to their native alanine state. As this substitution is present in all peptide and protein products, it will not be explicitly stated.
<table>
<thead>
<tr>
<th>Peptide</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N3</td>
<td>Unmodified H3(1-46)</td>
<td>ARTKTARKSTGGKAPRKLATKAARKSAPATGGVKKPHRYRPGTV</td>
</tr>
<tr>
<td>N4</td>
<td>H3(1-46) K4m³</td>
<td>ART(Km³)QTARKSTGGKAPRKLATKAARKSAPATGGVKKPHRYRPGTV</td>
</tr>
<tr>
<td>N5</td>
<td>H3(1-46) K9ac,K14ac,K18ac,K23ac</td>
<td>ARTKQTAR(Kac)STGG(Kac)APR(Kac)QLAT(Kac)AARKSAPA TGGVKKPHRYRPGTV</td>
</tr>
<tr>
<td>N6</td>
<td>H3(1-46) pT45</td>
<td>ARTKQTARKSTGGKAPRKLATKAARKSAPATGGVKKPHRYRPG(pT)V</td>
</tr>
<tr>
<td>M4M</td>
<td>H3(47-90) A47Thz,K56ac</td>
<td>(Thz)LREIRRYQ(Kac)STELLIRKLPFQRVLREIAQDFKTDLRFQSS AVM</td>
</tr>
<tr>
<td>M5M</td>
<td>H3(47-90) A47Thz</td>
<td>(Thz)LREIRRYQ(Kac)STELLIRKLPFQRVLREIAQDFKTDLRFQSS AVM</td>
</tr>
<tr>
<td>M4N</td>
<td>H3(47-90) A47Thz,K56ac, M90Nle</td>
<td>(Thz)LREIRRYQ(Kac)STELLIRKLPFQRVLREIAQDFKTDLRFQSS AVM(Nle)</td>
</tr>
<tr>
<td>M5N</td>
<td>H3(47-90) A47Thz,M90Nle</td>
<td>(Thz)LREIRRYQ(Kac)STELLIRKLPFQRVLREIAQDFKTDLRFQSS AVM(Nle)</td>
</tr>
<tr>
<td>M6</td>
<td>H3(47-90) A47Thz,K79m³, M90Nle</td>
<td>(Thz)LREIRRYQ(Kac)STELLIRKLPFQRVLREIAQDF(Kmet³)TDLR FQSSAV(Nle)</td>
</tr>
<tr>
<td>C3M</td>
<td>H3(91-135) A91Thz</td>
<td>(Thz)LQEASEAYLVALFEDTNLAAIHAKRVTIMPKDIQLARRIRGERA-COOH</td>
</tr>
<tr>
<td>C3N</td>
<td>H3(91-135) A91Thz,M120Nle</td>
<td>(Thz)LQEASEAYLVALFEDTNLAAIHAKRVTI(Nle)PKDIQLAR RIRGERA-COOH</td>
</tr>
<tr>
<td>C4</td>
<td>H3(91-135) A91Thz,K115ac, M120Nle,K122ac</td>
<td>(Thz)LQEASEAYLVALFEDTNLAAIHAKRVTI(Nle)PKDIQLAR RIRGERA-COOH</td>
</tr>
</tbody>
</table>

Table 5. Sequence of prepared synthetic peptide segments
Synthesis of C-terminal Peptide Segments

Peptides C3M, C3N and C4 (Table 5) were synthesized by automated SPPS with Fmoc chemistry directly on commercial substituted Fmoc-Ala-Wang LL resin (Novabiochem). In a typical synthesis, each amino acid to be incorporated was coupled with HBTU activation, and acetyl capping was performed in between cycles to insure that deletion product formation was minimized.

In the synthesis of the C-terminal peptide segments, the N-terminal cysteine residue was introduced as Boc-thiazolidine moiety. Thiazolidine eliminates the possibility of racemization which is known to occur during coupling of protected Fmoc-Cys species. Following deprotection and cleavage of the crude peptide mixture from resin, the resultant mixture was lyophilized and resuspended in 6 M guanidine with 100 mM phosphate buffer and 500 mM NaCl. Conversion of the terminal thiazolidine to cysteine was initiated with the addition of 300 mM methoxylamine HCl which reduced the pH of the solution to ~4. Progress of cysteine conversion was monitored by RP-HPLC and was typically complete within 5 hours.

The crude peptide mixtures were brought to RP-HPLC starting conditions and were centrifuged to remove any particulate. The resultant supernatants were purified by RP-HPLC over a Supelco column. Fractions determined to be >95% pure were pooled and lyophilized. The resultant peptide pools were assessed by RP-HPLC and MALDI-TOF.
MS to confirm purity and identity: peptide C3M (Fig. 64), C3N (Fig. 65) and C4 (Fig. 66). Purified peptide yields were 28.2 mg C3M (11% theoretical yield), 19.8 mg C3N (8%) and 20 mg C4 (8%) as determined by weight.

Figure 64. Characterization of peptide C3

RP-HPLC analysis of peptide C3 resolved with acetonitrile / H2O + 0.1% TFA on a gradient of 18-66% for the crude (top, left) and 27-45% for the pure (bottom, left). MALDI-TOF MS of pure peptide C3 with expected and observed [M+H]+ as described.
Figure 65. Characterization of peptide C3N

RP-HPLC analysis of peptide C3N resolved with acetonitrile / H₂O + 0.1% TFA on a gradient of 18-66% for the crude (top, left) and 27-45% for the pure (bottom, left). MALDI-TOF MS of pure peptide C3N with expected and observed [M+H]⁺ as described.

Significant methionine oxidation was detected in the crude cleavage mixture for peptide C3M which includes the native methionine residue at position 120. Peptide C3N was synthesized with the conservative norleucine (Nle) for methionine substitution (Fig. 65). Nle is a Met isostere which is commonly introduced in synthetic peptide segments to eliminate the possibility of peptide oxidation. This substitution allows for the preparation of homogeneous peptide and ligation product samples without the requirement for post-
synthesis reduction with NaI / DMS. Additionally, as the desired product resolves as a single species rather than two, purification is simplified allowing for an increase in product yield.

Figure 66. Characterization of peptide C4 purification

RP-HPLC analysis of peptide C4 resolved with acetonitrile / H₂O + 0.1% TFA on a gradient of 18-66% for the crude (top, left) and 32-50% for the pure (bottom, left). MALDI-TOF MS of pure peptide C4 with expected and observed [M+H]⁺ as described.
Peptide C4 incorporates the M120Nle substitution as well as acetylated lysines at positions 115 and 122 of the final H3 sequence. The preparation of this peptide when combined with modified N-terminal and/or middle peptide segments by sequential NCL will allow for the first reported preparation of a synthetic histone protein simultaneously modified in more than one structural and functional region.

Optimizing Synthesis on Dbz Substituted Resin

We initially synthesized the middle peptide segment acetylated at lysine 56 of the full H3 sequence (M4M, Table 5) on unsubstituted Rink amide MBHA LL resin to determine the quality of synthesis with Fmoc based automated protocols as described. The peptide was synthesized with HBTU activation and amino acids were introduced with 2x 30 minute couplings. Following synthesis, the peptide was deprotected and cleaved from the resin and the crude product was analyzed by RP-HPLC (Fig. 67, left). The crude peptide mixture contained a single major product which was identified by MALDI-TOF MS as the expected product, M4M-NH₂.

Identical conditions were used to synthesize peptide M4M on Dbz substituted Rink amide MBHA LL resin. First, Dbz was manually coupled to unsubstituted resin as Di-Fmoc-3,4-Dbz (Anaspec) with HATU activation and the remaining amino acids were coupled with automation and HBTU activation as described. Once synthesis was complete, the
peptide was deprotected, cleaved, and the resultant crude peptide mixture was analyzed by RP-HPLC and MALDI-TOF MS (Fig. 67, right). In contrast to the clean product prepared on unsubstituted resin, the crude peptide mixture arising from synthesis on Dbz resin contained a variety of products that did not resolve by RP-HPLC and only a small percentage of the total peptide mass corresponded to the desired product.

![Figure 67. Crude synthesis of M4M on Rink and Dbz resins](image)

RP-HPLC analysis with a gradient of 32-59% acetonitrile / H₂O + 0.1% TFA for crude M4M-NH₂ (left) and M4M-Dbz (right) peptides. MALDI-TOF MS analysis: M4M-NH₂, [M+H]⁺ observed 5441 m/z, expected 5438 m/z; M4M-Dbz, [M+H]⁺ observed 5568 m/z, expected 5572 m/z.

The synthesis of peptide M4M was repeated on Dbz substituted Rink amide resin using identical synthesis conditions. Simultaneously, the unacetylated middle segment (M5M, Table 5) was synthesized on Dbz-Rink substituted Chem Matrix resin (Matrix Innovation). Analysis of the crude peptide products obtained from an analytical scale
cleavage of each resin by RP-HPLC demonstrated a marked increase in synthesis quality, but the synthesis was still not as clean as that observed for M4M-NH₂ (Fig. 68). Importantly, there was no significant difference in the quality of the syntheses resulting from the use of different resins demonstrating that the poor synthesis is not due to the physical properties of the solid support employed.

Figure 68. RP-HPLC analysis of crude M4M-Dbz and M5M-Dbz

RP-HPLC analysis with a gradient of 32-59% acetonitrile / H₂O + 0.1% TFA for crude M4M-Dbz (left) and M5M-Dbz (right) peptides. MALDI-TOF MS analysis: M4M-Dbz, [M+H]⁺ observed 5572 m/z, expected 5572 m/z; M5M-Dbz, [M+H]⁺ observed 5531 m/z, expected 5529 m/z.

The crude M4M-Dbz and M5M-Dbz resins were treated with p-NPC followed by DIEA to convert the terminal Dbz to its Nbz form by standard methods. The resultant resin was treated with a TFA based cleavage solution and the deprotected crude peptide mixture was precipitated and washed with cold diethyl ether, resuspended in acetonitrile / H₂O + 0.1% TFA, and lyophilized. Peptides were purified by RP-HPLC over a Supelco
column, individual fractions were assessed, and fractions determined to be >95% pure were pooled. The final peptide pools were assessed by RP-HPLC and MALDI-TOF MS to confirm purity and identity (Fig. 69), and then lyophilized.

![Figure 69. Characterization of pure M4M-Nbz and M5M-Nbz peptides](image)

RP-HPLC analysis of purified M4M-Nbz (top, left) and M5M-Nbz (top, right) peptides on a gradient of 32-59% acetonitrile / H₂O + 0.1% TFA. MALDI-TOF MS of pure peptide M4M-Nbz (bottom, left) and M5M-Nbz (bottom, right) with expected and observed [M+H]⁺ as described.

A total of 12.4 mg M4M-Nbz and 17.5 mg M5M-Nbz purified peptides were obtained following lyophilization as determined by weight difference, for an overall synthetic yield of ~4% for each peptide as determined by the initial scale of the syntheses.
Although the amount of peptide generated and overall synthetic yields were low, there was a significant increase over the 3.7 mg of corresponding peptide M2 isolated from the previously discussed manual Boc *in situ* neutralization syntheses (Chapter 3).

The observed poor quality of the M4M and M5M syntheses on Dbz substituted resin may in part be attributed to residual activity of the free amine on Dbz which has been previously identified and comprehensively discussed in Chapter 4. In an effort to increase the quality of peptide synthesis on Dbz substituted resins, all future syntheses of the full middle peptide segment were completed on reversibly protected Dbz(Alloc) substituted resin.

Prior to completing a full synthesis of the middle peptide segment on Dbz(Alloc) substituted base resin, a small peptide [FQ(Kac)STELLIR] which corresponds to residues 54-63 of the H3-K56ac sequence was synthesized (Fig. 70, left). This partial synthesis revealed that the region directly adjacent to Gln 55 and Lys 56 is synthetically challenging as evidenced by a significant amount of observed truncated product. The formation of secondary structure in a growing peptide chain is a possible cause of truncation, which arises when the terminal amine is oriented such that it is not accessible during a coupling cycle. The introduction of a secondary amine such as proline is capable of breaking up the secondary structure of a particularly challenging region.
Figure 70. Pseudoproline dipeptide decreases formation of truncated product

RP-HPLC analysis of crude cleavage mixtures from synthesis of peptide FQ(Kac)STELLIR synthesized without (left) or with (right) Fmoc-Ser(tBu)-Thr($\psi^{Me,Me}$pro)-OH pseudoproline dipeptide. Desired peptide product identified as (*).

The small peptide was resynthesized under identical conditions however the residues corresponding to serine 57 and threonine 58 were introduced as the pseudoproline dipeptide Fmoc-Ser(tBu)-Thr($\psi^{Me,Me}$pro)-OH. As described, the incorporated pseudoproline may disrupt structure formation during synthesis and is unmasked to yield the native S57-T58 sequence upon treatment with TFA during peptide deprotection and cleavage. Analysis of the crude peptide synthesized with the pseudoproline dipeptide demonstrated a single product corresponding to the desired species confirming the synthetic benefit that may be obtained by selective use of pseudoprolines (Fig. 70, right).
Preparation of Amino Acid Substituted Dbz(Alloc) Resins

Over the course of this study, two distinct methods were employed to prepare amino acid substituted Dbz(Alloc) base resins. (Fig. 71) Fmoc-Met-Dbz(Alloc) and Fmoc-Nle-Dbz(Alloc) base resins suitable for the preparation of peptides derived from residues 47-90 of histone H3 were initially prepared by directly coupling Fmoc-Met-OH or Fmoc-Nle-OH to Dbz substituted resins with HBTU activation. The resin was protected with allyl chloroformate according to literature protocols. Alternatively, following coupling of the small molecule 3’-Fmoc-4’Alloc-Dbz to base resin, the resin was deprotected and Fmoc-Met-OH or Fmoc-Nle-OH was coupled in excess (10 eq.) with HATU activation. The addition of amino acid to the 3’-Fmoc-4’-Alloc-Dbz substituted resin was preferable as it yielded a single isoform with chain extension occurring only from the 3’ position, while also eliminating the possibility of double amino acid incorporation on the Dbz.

All ligation compatible peptides derived from residues 1-46 of histone H3 were synthesized on Fmoc-Val-Dbz(Alloc) substituted resin. The β-branched amino acid valine does not efficiently couple to previously substituted Dbz(Alloc) base resin due to steric occlusion at the reactive center. Interestingly, coupling of Fmoc-Val-OH to unprotected Dbz substituted base resin with HBTU activation does not yield detectable amounts of doubly substituted product and single loading of valine on Dbz is typically incomplete under these conditions resulting in the formation of valine deletion products. Valine was first coupled to Dbz substituted base resin as Fmoc-Val-OH with HATU
activation, and a single coupling cycle is typically sufficient to ensure complete acylation of the Dbz. Fmoc-Val-Dbz resin was subsequently Alloc protected. The direct coupling of valine to Dbz results in a mixture of the 3’ and 4’ substituted isoforms of the desired product but may resolve as independent species during RP-HPLC analysis.

Figure 71. Schemes for the preparation of Met-Dbz(Alloc) resin
Synthesis of Middle Peptide Segments

Variably modified peptide segments corresponding to residues 47-90 of histone H3 were synthesized on prepared Fmoc-Nle-Dbz(Alloc) resins (Table 5): M4N (w/ K56ac), M5N (unmodified), and M6 [tri-methylated lysine 79 (K79m3)]. Subsequent residues were coupled by standard automated peptide synthesis protocols with HCTU or HATU activation. Double coupling of amino acids was routinely employed along with acetyl capping prior to deprotection to maximize product formation and minimize the appearance of deletion products. Tri-methylated lysine was manually introduced as N-α-Fmoc-N-ε-trimethyl-N-ʟ-lysine chloride (1.5 eq.) with PyBop activation. Residues corresponding to serine 57 and threonine 58 were incorporated as the pseudoproline dipeptide with manual addition on the synthesizer in lower excess (2.2 eq.) than a standard amino acid to conserve the dipeptide, but coupling times were routinely doubled to one hour. The N-terminal residue was introduced as Boc-thiazolidine-OH.

Following synthesis, analytical cleavage of the M4N-Dbz(Alloc) and M5N-Dbz(Alloc) resins was completed to assess the quality of the syntheses. RP-HPLC of the resultant crude peptide mixtures demonstrated a significant increase in synthesis quality when compared to the M4M and M5M syntheses on unprotected Dbz resins (Fig. 68). Identity of the major observed product in each synthesis was confirmed to be desired full length product by MALDI-TOF MS. Additionally, the second major product in the synthesis of the M5N peptide was identified as a truncation product resulting from acetylation after
the introduction of the arginine residue corresponding to position 69 of the full sequence. This residue was not double coupled which likely contributed to the incomplete coupling observed in the M5N synthesis. In all future couplings, double coupling of this residue will be employed.

Figure 72. Characterization of M4N and M5N purification

RP-HPLC analysis of peptides M4N and M5N synthesized on Dbz(Alloc) resin on a gradient of acetonitrile / H$_2$O + 0.1% TFA: crude M4N-Dbz (top, left) and M5N-Dbz (top, right) (27-72%), purified M4N-Nbz (middle, left) (32-59%), and M5N-Nbz (middle, right) (32-50%) MALDI-TOF MS of pure peptide M4M-Nbz (bottom, left) and M5M-Nbz (bottom, right) with expected and observed [M+H]$^+$ as described.
The M4N-Dbz(Alloc) and M5N-Dbz(Alloc) resins were subjected to manual Alloc deprotection and Nbz according to standard methods. Following cleavage of the Nbz converted peptides from the solid support, the crude peptide mixtures were purified by RP-HPLC. Fractions determined to be >95% pure were pooled and the resultant peptide pools were analyzed by RP-HPLC and MALDI-TOF MS to confirm purity and identity (Fig. 72). The final pools were lyophilized and peptide yield was determined to be 20 mg of M4N-Nbz and 15.4 mg of M5N-Nbz by weight. The overall theoretical synthetic yield for the syntheses was determined to be 8% and 6%, respectively. The observed synthetic yields represent a reproducible 1.5 to 2-fold enhancement over syntheses completed on unprotected Dbz resin.

Peptide M6 incorporating lysine tri-methylation at the position corresponding to lysine 79 in the full H3 sequence was prepared as described. Analytical cleavage of the M6-Dbz(Alloc)-substituted resin was completed and the resultant crude peptide mixture was resolved by RP-HPLC on a Supelco C18 column (Fig. 73). The poor synthesis quality is likely a direct result of the incorporated unprotected tri-methyllysine. Analysis of the side products by MALDI-TOF MS suggested that the tertiary amine of the modified lysine was partially demethylated over the course of the synthesis which introduces an additional reactive center for peptide chain initiation.
Figure 73. Analysis of M6-Dbz(Alloc) crude peptide mixture

RP-HPLC analysis of crude M6-Dbz(Alloc) peptide mixture with a gradient of 27-66% acetonitrile / H$_2$O + 0.1% TFA. Desired product identified by (*) and confirmed by MALDI-TOF MS analysis: [M+H]$^+$ observed 5751 m/z, expected 5750 m/z.

Synthesis of N-terminal Peptide Segments

N-terminal peptide segments corresponding to residues 1-46 of histone H3 were synthesized on previously prepared Fmoc-Val-Dbz(Alloc) by automated SPPS with HCTU activation (Table 5): N3 (unmodified), N4 (K4m$^3$), N5 (K9ac,K14ac,K18ac, K23ac), and N6 [phosphorylated at threonine 45 (pT45)]. To maximize the quality of the syntheses, a standard synthetic cycle included two sequential coupling steps, followed by a single treatment with acetyl capping solution to minimize the formation of deletion products. Tri-methylated lysine was introduced into peptide N4 by manual coupling with PyBop activation as previously described. Phosphorylated threonine 45 was introduced
to peptide N6 with manual coupling of N-α-Fmoc-O-benzyl-l-phosphothreonine (2 eq.) with HBTU activation on freshly deprotected Val-Dbz(Alloc) resin for one hour.

Synthesis of the unmodified N-terminal peptide segment N3 was first completed on Dbz(Alloc) substituted resin as described. Following synthesis, the Alloc protecting group was removed from the bulk resin and an analytical cleavage was completed to yield the crude N3-Dbz peptide. The crude peptide mixture was assayed by RP-HPLC (Fig. 74, top). The primary product was confirmed as the desired N3-Dbz species by MALDI-TOF MS and the two isoforms partially resolved under the RP-HPLC conditions tested. The remaining resin bound peptide was converted to its Nbz form as described and cleaved. Purification of the resultant crude N3-Nbz peptide was performed by RP-HPLC over a Supelco C18 column and individual fractions were tested. Fractions determined to be >95% pure were pooled and the resultant peptide pool was analyzed by RP-HPLC and MALDI-TOF MS to confirm purity and identity (Fig. 74). The final pool was lyophilized and 18.8 mg of purified N3-Nbz was obtained for a theoretical synthetic yield of 7%. A previous synthesis of the N3-Nbz peptide on unprotected Dbz resin was found to have a 5% theoretical yield. The smaller observed increase in theoretical yield for peptide N3 compared to the middle peptides is likely a direct result of the C-terminal valine residue which occludes the free Dbz amine even in the absence of a protecting group.
Figure 74. Characterization of N3-Nbz purification

RP-HPLC analysis of peptide N3 resolved with acetonitrile / H$_2$O + 0.1% TFA on a gradient of 9-36% for the crude (top, left) and 9-27% for the pure (bottom, left). MALDI-TOF MS of pure peptide C4 with expected and observed [M+H]$^+$ as described.

The modified N-terminal peptides (N4, N5, and N6) were prepared simultaneously with unmodified N3 as described. Cleavage of the unmodified peptides was completed on an analytical scale from the protected Dbz(Alloc)-substituted resins. Crude peptide mixtures were resolved by RP-HPLC on a Supelco C18 column to assess quality and product identity was confirmed by MALDI-TOF MS (Fig. 75). RP-HPLC sufficiently resolves the individual species and subsequent purification should provide enough material for use in the preparation of synthetic modified histone H3 proteins.
Figure 75. Crude syntheses of N-terminal peptides on Val-Dbz(Alloc) resin

RP-HPLC (left) and MALDI-TOF MS (right) analysis of peptides on a gradient of acetonitrile / H₂O + 0.1% TFA: 9-50% for crude N4-Dbz(Alloc), 14-41% for N5-Dbz(Alloc), and 9-27% for N6-Dbz(Alloc). MALDI-TOF MS analysis as described.
Total Chemical Synthesis of Histone H3 Proteins

The purified peptide segments prepared by automated SPPS were chemoselectively condensed to prepare synthetic histone H3 proteins by sequential NCL as first presented in Chapter 3 (Fig. 76). Desulfurization of the final ligation product regenerates the native H3(C110A) protein sequence through conversion of ligation site cysteine residues to alanines. We initially prepared fully synthetic unmodified and site-specifically acetylated histones in this manner.

Figure 76. Total synthesis scheme for the preparation of histone H3
Unmodified histone H3, termed H3(C110A)_{syn}, was prepared by the ordered ligation of peptides N3-Nbz, M5M-Dbz, and C3M. Briefly, lyophilized M5M-Nbz was resuspended with 2 molar equivalents of the C3M peptide in 6 M guanidine ligation buffer and ligation was initiated with the addition of 75 mM MESNA. Ligation proceeded at 25 °C with nutation and progress was monitored by RP-HPLC until no further ligation product formation was observed, typically 2 days. The first ligation product was purified by RP-HPLC with the N-terminal thiazolidine (Thz) residue intact. Purification fractions were assessed and fractions determined to be >95% pure were pooled and lyophilized.

Lyophilized ligation product was resuspended in 6M guanidine ligation buffer and the Thz terminus was converted to cysteine with the addition of methoxylamine. Following Thz deprotection, the pH was adjusted to 7.5 and the ligation was initiated by addition of 75 mM MPAA and a 20-fold molar excess of peptide N2. Ligation progress was monitored for three days and the resulting reaction mixture was dialyzed to remove MPAA prior to desulfurization as described. H3(C110A)_{syn} was purified by RP-HPLC. A final yield of 160 µg was determined by UV quantitation, and protein identity was confirmed by MALDI-TOF MS (Fig 77, left). This fully synthetic, unmodified H3 was reconstituted into nucleosomes as described in chapter 3 and compared to nucleosomes reconstituted with unmodified, recombinant H3 protein, which demonstrated that the chemical steps required for sequential NCL did not introduce any changes that might not be detected by RP-HPLC or MALDI-TOF MS.\textsuperscript{11}
Fully synthetic histone H3 acetylated at lysine 56 (H3-K56ac) was prepared in an identical manner from the sequential NCL of peptides N3-Nbz, M4M-Nbz, and C3M. Following purification, protein identity was confirmed by MALDI-TOF MS (Fig. 77, right). The successful preparation of the H3-K56ac(C110A) protein first characterized in chapter 3 from peptide segments synthesized with automated Fmoc SPPS protocols validates the compatibility of our synthetic strategy with this chemistry.

![MALDI-TOF MS analysis of synthetic H3 proteins](image)

Figure 77. MALDI-TOF MS analysis of synthetic H3 proteins
MALDI-TOF MS analysis of H3(C110A)$_{syn}$ (left) and H3-K56ac(C110A) (right) proteins with expected and observed [M+H]$^+$ as described.
The total synthesis of histone H3 acetylated at lysines 56, 115, and 122 has recently been initiated. The preparation of synthetic H3-K56ac,K115ac,K122ac will allow biophysical characterization of changes in stability and dynamics arising from the simultaneous modification of the nucleosome entry-exit and dyad regions.

H3-K56ac,K115ac,K122ac(A47C,A91C) was prepared by sequential NCL as follows. Lyophilized M4N-Nbz (6.2 mg) was resuspended with 2 molar equivalents of C3N peptide in 6M guanidine ligation buffer and reaction was initiated with the addition of 50 mM MESNA. Ligation proceeded at 25 °C with nutation for 3 days and progress was monitored by RP-HPLC and SDS-PAGE. The N-terminal Thz residue was converted to cysteine with the addition of 400 mM methoxylamine and the conversion reaction proceeded for 6 hours. The first ligation product was purified by RP-HPLC. Purification fractions were assessed and fractions determined to be >95% pure were pooled and lyophilized. Characterization of the resultant pool to confirm purity and identity was completed by RP-HPLC and MALDI-TOF MS (Fig. 78).
Figure 78. Characterization of H3-K56ac,K115ac,K122ac first ligation product

RP-HPLC (left) and MALDI-TOF MS (right) analysis of purified first ligation product, residues 47-135 of H3-K56ac, K115ac,K122ac(A47C,S96C) on a 32-63% acetonitrile / H2O + 0.1% TFA gradient. MALDI-TOF MS analysis as described.

Purified first ligation product was resuspended with 5 molar equivalents of peptide N3-Nbz in 6M guanidine buffer and ligation was initiated with the addition of 50 mM MESNA. Ligation was allowed to proceed at 25 °C with nutation and reaction progress was monitored by RP-HPLC and SDS-PAGE. MESNA was initially chosen as the ligation co-thiol due to its compatibility with desulfurization, but the formed MESNA thioester exhibits reduced ligation kinetics when compared to MPAA. After 24 hours, minimal product formation was observed and MPAA was added to the crude ligation mixture to a final concentration of 50 mM. Following the addition of MPAA, a significant increase in product formation was observed. After 4 days, an aliquot of the full length product, H3-K56ac,115ac,122ac (A47C,A91C), was purified by RP-HPLC.
from the crude ligation mixture. Product purity and identity was confirmed by RP-HPLC and MALDI-TOF MS (Fig. 79).

Figure 79. Characterization of H3-K56ac,K115ac,K122ac(A47C, A91C)

RP-HPLC (left) and MALDI-TOF MS (right) analysis of purified H3-K56ac,K115ac, K122ac(A47C,S96C) protein on a 32-63% acetonitrile / H₂O + 0.1% TFA gradient. MALDI-TOF MS analysis as described.
Sequential NCL for the total chemical synthesis of histone H3 is amenable to the preparation of histones variably modified throughout the protein sequence and is primarily limited by the chemistry used to prepare the individual peptide segments. Herein we describe the development of protocols for the automated Fmoc SPPS of peptide segments suitable for use in sequential NCL. In addition to allowing automation, Fmoc SPPS is compatible with the introduction of phosphorylated and tri-methylated amino acids into prepared peptide segments.

Initial syntheses of peptides by automated Fmoc SPPS on unprotected Dbz resin were successful for the preparation of ligation compatible peptide thioesters. Synthetic yields of the generated peptides were low due largely to the presence of significant deletion products which were difficult to resolve during purification. Syntheses were initially optimized through the coupling of select pseudoproline dipeptides in regions known to be synthetically challenging in an effort to minimize deletion product formation.

In subsequent syntheses, reversible Alloc protection of the free amine of Dbz allowed for the use of multiple coupling cycles, stronger coupling agents, and the introduction of post-coupling acetylation steps. These chemically optimized syntheses resulted in a significant reduction in the formation of side products. Alloc protection reproducibly increased synthetic peptide yields by 50-100% over unprotected syntheses.
Optimized synthetic protocols were used to generate a library of modified peptide segments suitable for use in the total synthesis of histone H3. Peptides incorporating acetylated, phosphorylated, and methylated amino acids were prepared. Additionally, the norleucine for methionine substitution was employed to eliminate methionine oxidation which was commonly observed during the total synthesis of H3-K56ac(C110A). The prepared library of ligation compatible peptide segments allows for the combinatorial synthesis of a variety of modified histone H3 proteins.

Sequential NCL was completed with peptides prepared by automated Fmoc SPPS to yield synthetic unmodified histone H3 and H3-K56ac. The fully synthetic, unmodified H3 was reconstituted into nucleosomes and compared to nucleosomes reconstituted with unmodified, recombinant H3 protein. Importantly, no significant difference was observed between the nucleosomes prepared with synthetic or recombinant unmodified histone H3, demonstrating that the chemical steps required for sequential NCL do not alter nucleosome structure or dynamics.\(^{11}\)

The total synthesis of histone H3-K56ac,K115ac,122ac(A47C,A91C) has recently been completed. Following desulfurization, the prepared protein may be reconstituted into semi-synthetic nucleosomes which are modified within the entry-exit and dyad regions. Characterization of nucleosomes simultaneously modified in multiple regions of the
DNA-protein interface will increase our understanding of the functional role that histone modification cross-talk plays in nucleosome structure and dynamics.
Acknowledgements

Sarah Dreher has extensively assisted the purification of the modified peptide library. CJ Howard contributed to the use of pseudoprolines in the central segment. Additionally, Justin North and Dr. Michael Poirier performed the described characterizations of the synthetic H3(C110A) protein.
Histones are post-translationally modified throughout their sequence. These modifications have a significant impact on nucleosome structure and dynamics. Modifications may exist within the unstructured tail domains or within the folded core of the protein. The identification of post-translational modifications within the folded core of histone proteins distal from the unstructured N-terminal tail region, and the identification of cross-talk between these modifications, has made it necessary to develop novel strategies to allow for the preparation of homogeneous protein samples simultaneously modified in variable regions of the nucleosome. The work presented herein describes the development of a series of synthetic tools suitable for this purpose.

Chapter 2: We introduced an expressed protein ligation (EPL) desulfurization strategy for the semi-synthesis of histone H4 proteins modified within the functional LRS (loss of rDNA silencing) domain and αC helix. The presented strategy yields the native protein sequence by combining ligation with desulfurization of the ligation site cysteine to a native alanine residue. We successfully employed this strategy to prepare acetylated and phosphorylated histones: H4-K77ac,K79ac, H4-pT80, and H4-K91ac.
The semi-synthetic modified proteins were initially reconstituted into histone octamers. Notably, as previously demonstrated in refoldings with semi-synthetic histone H3, we determined that the recombinant H4(1-75) truncated product does not fold into histone octamers and need not be fully purified away from semi-synthetic proteins. Purified semi-synthetic histone octamers containing H4-K77ac,K79ac or H4-pT80 proteins were refolded with DNA to yield mononucleosomes which were confirmed by native PAGE to form wild-type structure.

The impact of acetylation within the LRS region of the nucleosome was probed by reconstituting an array of nucleosomes incorporating the semi-synthetic H4-K77ac,K79ac protein. Arrays prepared by our collaborators, Marek Simon and Dr. Michael Poirier, were assayed by single molecule force extension to assess stability by measuring octamer retention following force unwrapping. H4-K77ac,K79ac modified arrays demonstrated no significant impact on octamer retention when compared to wild-type arrays. However, these modifications did impact DNA unwrapping, and increased protein binding to a site 27 bp into the nucleosome.

We further reconstituted nucleosomes with semi-synthetic H3-K115ac,K122ac and H4-K77ac,K79ac proteins. The prepared nucleosomes included eight independent acetyllysines positioned within both the LRS and dyad regions of the histone-DNA interface, in three clusters. Despite the large neutralization of charge and introduction of steric bulk introduced by acetylation, no significant change in the mononucleosome
structure was observed by native PAGE. A more rigorous study of nucleosome stability and dynamics will be necessary to establish the impact of the introduction of acetylation simultaneously within the LRS and dyad regions.

Chapter 3: Although EPL enables the study of modifications adjacent to the protein termini, EPL is not suitable to the introduction of modifications more centrally located within a protein sequence. Furthermore, the site specific introduction of multiple modifications throughout a protein sequence requires a more robust synthetic technique. We developed a total chemical synthesis approach to expand the accessible regions for the introduction of modifications.

We used a sequential NCL strategy to prepare site-specifically modified histone proteins by the ordered condensation of peptide segments in independent ligation steps to yield a final full length modified product. In our initial strategy, three peptide segments prepared manually by Boc in situ neutralization\textsuperscript{57} were combined to yield synthetic histone H3-K56ac(R40C,S96C,C110A) which incorporated non-native cysteine substitutions at sites identified by sequence alignment\textsuperscript{102-104}. Unfortunately, comparison of the resultant nucleosomes with those reconstituted with recombinant H3(C110A) and H3(R40C,S96C,C110A) proteins demonstrated that the non-native ligation site cysteine substitutions had a larger observable effect on nucleosome DNA dynamics than the introduced acetylated lysine. This result demonstrates the importance of carefully characterizing changes introduced into the histone sequence.
Our second generation synthetic strategy incorporated post-ligation desulfurization to allow the use of alanine ligation sites, rather than the introduction of non-native cysteines. This strategy allowed for the synthesis of more accessible peptide segments, generated a reproducible 3-fold increase in overall synthetic yield, and reliably produced sufficient fully synthetic H3-K56ac(C110A) protein suitable for biochemical and biophysical characterization. In collaboration with the Poirier laboratory, we determined that acetylation within the entry-exit region of the nucleosome attenuates DNA-protein interactions, increases DNA site-accessibility, and facilitates recognition and site-specific binding of the LexA protein. The lysine to glutamine genetic substitution commonly employed for in vivo characterization of acetylation qualitatively reproduced the observed effects of H3-K56 acetylation. While acetylation at H3-K56ac increased DNA unwrapping in the entry-exit region, it did not significant increase histone octamer dissociation from DNA as measured by force extension unwrapping experiments. These results suggest that the primary impact of lysine 56 acetylation is localized within the entry-exit region and is not propagated throughout the DNA-octamer interface.

This chapter described the generation of the first fully synthetic histone protein, which theoretically enables the introduction of histone modifications through the histone sequence, a theme that was further explored in chapter 5. Further, since histone H3 is the largest and most synthetically challenging of the core histone proteins, it should be possible to extend these techniques to other histone proteins as well.
Chapter 4: The success of our total synthesis strategy for the preparation of histone H3-K56ac was tempered by low synthetic yields of the middle peptide segment and the incompatibility of Boc chemistry with automated peptide synthesis protocols. Further, the introduction of tri-methylated lysines and phosphorylated serines or threonines is incompatible with Boc-SPPS chemistry, which limits the modifications that may be incorporated by this approach. We therefore sought to further refine our synthetic approach through the use of Fmoc SPPS techniques for the preparation of peptide segments.

Fmoc SPPS is not compatible with the direct synthesis of peptide thioesters suitable for ligation. Initial attempts to utilize the Dbz masked thioester linker\(^{58}\) as a platform for the preparation of peptide thioesters were limited by low synthetic yields resulting in part from an accumulation of branched products. Additionally, it was demonstrated that the free amine on the Dbz auxiliary was susceptible to acetylation during the capping steps necessary to minimize the formation of side products during the synthesis of long, challenging peptide sequences such as these required histone-derived thioesters. We were able to exploit this reactivity as an avenue for controlled substitution with an affinity label (biotin) and a fluorophore (TAMRA). This technique allows the dual use of expensive peptides synthesized on a Dbz linker for use as a peptide thioester in ligation, or as a peptide ligand, for example in the study of histone binding proteins.
As a solution to the reactivity of Dbz, we developed a reversible protection strategy for the free amine of Dbz. Alloc protection of Dbz greatly increased the synthetic purity and yield of our peptide thioesters. Additionally, we introduced a general use resin substituted with 3’-Fmoc-Dbz(Alloc) that may be directly employed for peptide synthesis of most common C-terminal thioesters. This work lays the groundwork for the synthesis of peptide thioesters suitable for sequential NCL of modified histone proteins while providing the benefits of automated peptide synthesis.

Chapter 5: The thioester peptide segments required for total histone synthesis are synthetically challenging. While our refinement of automated Fmoc SPPS protocols for the synthesis of peptide thioesters laid the groundwork for the preparation of sequential native chemical ligation compatible peptide segments, initial syntheses yielded the desired peptide thioester product in low synthetic yields. Several rounds of synthesis optimization increased up to 20-fold the yields of purified synthetic peptide products compared to the initial manual Boc-SPPS syntheses. These yields were sufficient to allow us to proceed with total histone synthesis using an Fmoc-SPPS approach.

We have used this approach to generate fully synthetic histone H3, both unmodified and modified at H3-K56ac. The fully synthetic, unmodified H3 was reconstituted into nucleosomes and compared to nucleosomes reconstituted with unmodified, recombinant H3 protein. Importantly, no significant difference was observed between the
nucleosomes prepared with synthetic or recombinant unmodified histone H3, demonstrating that the chemical steps required for sequential NCL do not alter nucleosome structure or dynamics.\textsuperscript{11}

Our approach is sufficiently robust to allow the synthesis of a library of modified histone H3 peptide segments that will be used for the combinatorial preparation of a library of fully synthetic, heavily modified, histone H3 proteins. Specifically, we have prepared peptides suitable for the incorporation of histone H3 modifications (including phosphorylation, acetylation, and methylation) identified to have various biological roles: signals of active chromatin: lysine 4 tri-methylation (K4m\textsuperscript{3}),\textsuperscript{23, 127-129} N-terminal tail hyper-acetylation (K9ac,K14ac,K18ac,K23ac), and lysine 115 and 122 acetylation (K115ac,K122ac);\textsuperscript{63} DNA replication and apoptosis: phosphorylated threonine 45 (pT45);\textsuperscript{130, 131} newly synthesized histones and DNA repair: lysine 56 acetylation\textsuperscript{89-91, 132-136} and lysine 79 tri-methylation (K79m\textsuperscript{3})\textsuperscript{137} which has also been implicated in transcriptional silencing.\textsuperscript{138, 139} Further, prepared peptides incorporate both methionine and the norleucine substitution that eliminates methionine oxidation to simplify purification; direct comparison of these proteins will demonstrate the suitability of this substitution. We are now in the process of generating a combinatorial library of modified histone H3 proteins from these ligation-compatible peptide segments; this will demonstrate the utility of a reproducible, facile approach to histone proteins with the full complement of modifications representative of histone proteins in the cell.
In summary, this thesis introduces several approaches for the use of native chemical ligation to address the synthesis of modified histone proteins. The work presented demonstrates the systematic development of a generally accessible sequential native chemical ligation total synthesis strategy for the preparation of simultaneously site-specifically modified histone proteins. We are continuing our efforts to incorporate not only multiple modifications within a single protein, but multiple modified histones within a single nucleosome to allow for the study of cross-talk that between modifications in a single, heavily modified, nucleosome.
References


217


Appendix A: MALDI-TOF MS of LYRAGX-Dbz-R Products

A)

A_G: LYRAGA-Dbz-R

Exp: 939.5 m/z
Obs: 939.6 m/z

B)

A_K: LYRAGA-Dbz[Ac]-R

Exp: 981.5 m/z
Obs: 981.6 m/z

C)

A_J: LYRAGA-Dbz-R
LYRAG/

Exp: 1499.8 m/z
Obs: 1499.3 m/z

D)

F_G: LYRAGF-Dbz-R

Exp: 1015.5 m/z
Obs: 1015.3 m/z

E)

F_K: LYRAGF-Dbz[Ac]-R

Exp: 1057.5 m/z
Obs: 1057.5 m/z

F)

F_J: LYRAGF-Dbz-R
LYRAG/

Exp: 1575.8 m/z
Obs: 1575.2 m/z
Appendix B: RP-HPLC analysis of amino acid loading on Dbz(Alloc) resin