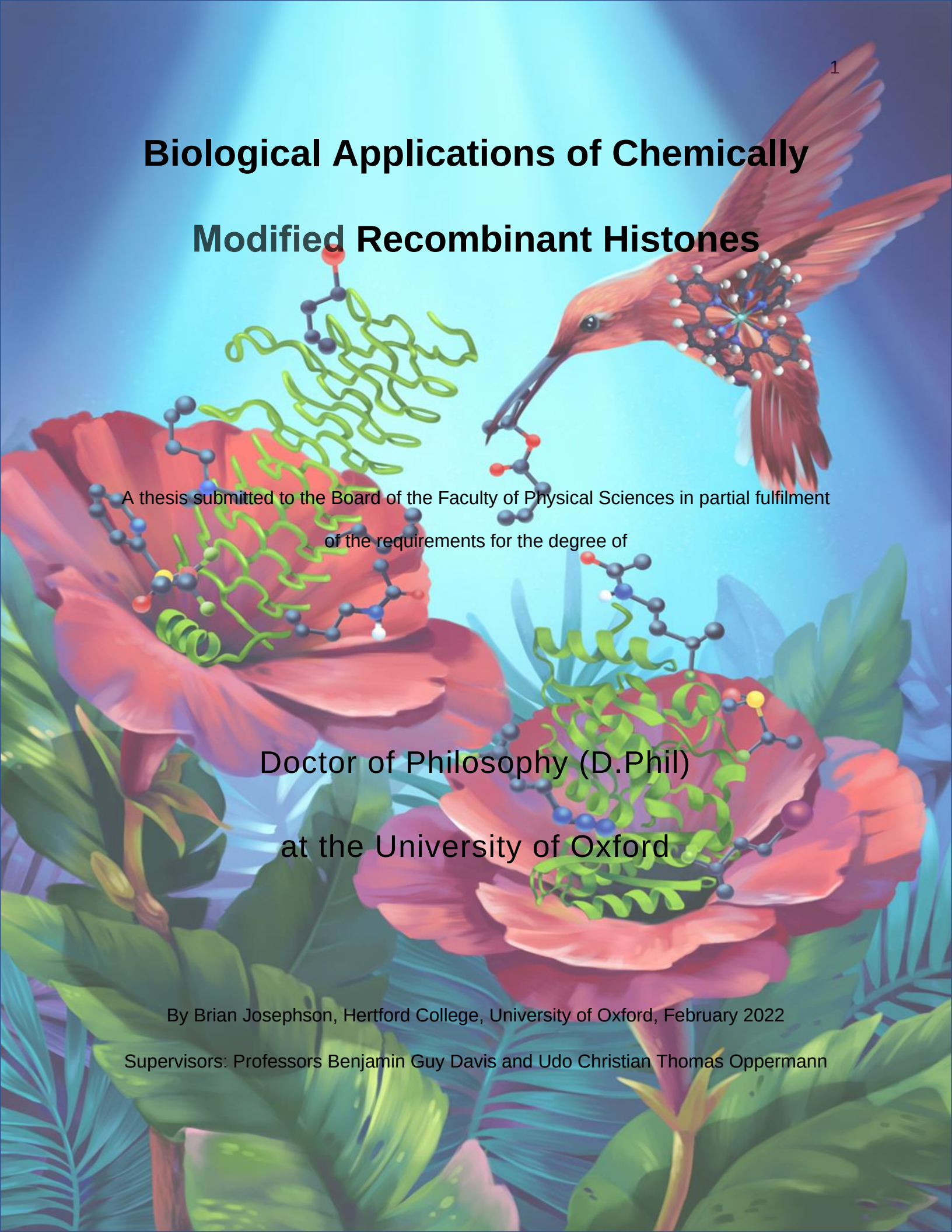
Biological Applications of Chemically Modified Recombinant Histones



A thesis submitted to the Board of the Faculty of Physical Sciences in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (D.Phil)

at the University of Oxford

By Brian Josephson, Hertford College, University of Oxford, February 2022

Supervisors: Professors Benjamin Guy Davis and Udo Christian Thomas Oppermann

Abstract

In eukaryotes, DNA is packaged into repeating units called nucleosomes, which consist of ~150 base pairs of DNA wrapped around eight proteins called histones. These histones are often laden with post-translational modifications (PTMs), and while many PTMs have been shown to have profound effects on gene expression and chromatin organization, most remain uncharacterized. This thesis strives to elucidate the functions of PTM containing histones. **Chapter 1** describes a light-driven reaction that site-specifically inserts sidechains of interest into histones via radical addition to previously installed dehydroalanine residues, with a scope encompassing many native and PTM containing sidechains, as well as several reactive functional groups. **Chapter 2** utilizes the reaction from **Chapter 1**, as well as others from our group, in several independent applications: The enzymatic removal of installed histone PTMs is tested, histone methylation and phosphorylation are assessed in the structural context of a nucleosome, PTM processing on fluorinated histone sidechains is tracked by ^{19}F-NMR, and interaction partners are crosslinked to histones bearing electrophilic sidechains. **Chapter 3** studies histones in more complex, living systems. The understudied ability of histones to penetrate cell membranes is investigated, revealing the incorporation of extracellular histones into the chromatin of nearby cells, provoking a strong inflammatory and immune response. Furthermore, several light-driven histone modification reactions are attempted in the translucent model organism, the zebrafish. In conclusion, this thesis describes powerful tools for modifying histones and applies them to relevant problems in epigenetics, revealing unknown roles and functions of histones and their PTMs.

Acknowledgments

Writing up my graduate work has been a constant exercise in reflection. Going over the data collected years ago makes it so obvious how research never exists in a vacuum, and each result is attached with memories of who was around in the lab, what was happening outside of the lab, and what was going on in the world. Numbers and graphs brought back memories of celebration, anger, boredom, excitement, confusion, and curiosity. Altogether, this thesis is so much more than just a collection of my research. It is also a reflection of over four years of support from colleagues, friends, and family.

First and foremost, I owe an enormous amount of thanks to Professor Ben Davis. I truly think there is no other group in the world where I could have investigated my interests as widely, freely, and with as much collaborative support as in the BGD group, and I am grateful for the opportunity to be part of it. Ben was a constant provider of inspiration, collaborative motivation, seemingly bottomless source of scientific knowledge, and an entertaining colleague and friend. I hope we continue working together.

Next, I must acknowledge my fellow soldiers in the trenches, my friends and colleagues over the years who worked in the ICL. Together, we made the very best out of an old and unideal workplace, under the consistently unhelpful, dismissive, and unsafe actions of the department. Among the ICL members over the years (Ben Bower, Naseer, Charlie, Mateusz, Patrick, Tim, Adeline, Andrew, Xuanxiao, Caitlin, Alice Bedwell, Alice Eddershaw, and honorary members Ryan and Lindsey) can be found my best and most supportive friends during my graduate studies. My favorite grad school memories, both inside and outside of the lab, took place in their company.

Other members of the BGD group deserve credit as well. The pandemic, coupled with the lab move to RFI and Pharmacology, made the last four years more difficult and disruptive than most. Yet many went above and beyond to help train, explain, provide material, or give advice on countless science problems, and were treasured friends outside of the lab (Alex, Simon, Tobi, Aziz, Gustaf, Satoshi, Sarah, Jack, George, Tom, Nan, Yibo, Lan, Xiaochao, Jitka, Xia-Ping, Dimitrios, Kim, Abirami, Jesse, Yue, Timothy, Moritz, Claudio, Sandrine, Takuya, Artem, Marielle, and Mylene).

Many scientists outside of our group, and department, were also extremely helpful, either as collaborators, or purely by offering advice or access and training on instruments. Special thanks goes to the groups of Professor Akane Kawamura (and member Grace), Professor Shabaz Mohammed, Professor Veronique Gouverneur (and members Patrick, Adeline, Claudio, Sandrine, Alex, and Jeroen), Professor Richard Compton (and members Christopher and Lifu), Professor Fumiko Esashi, Professor Mathilda Mommersteeg (and member Will), Professor Andrew Baldwin (and member Charlie), Professor Daniel Anthony (and member Abi) as well as various members of the nearby groups of Professor Hagan Bayley and Professor Christopher Schofield.

Outside of science, many friends, housemates, and teammates have been crucial pillars of support, providing much needed fulfillment and distraction from lab work. Many great memories were had with Hertford college MCR members, such as Jonny, Emma, Alex Fletcher, Alex C. S., Georgia, Antonis, and countless others. Both the Hertford college rowing team and soccer team provided much needed physical outlets, as did running partners Alice Eddershaw, Ryan McNamee, and Mateusz Imiolek. Housemates Jonny, Klem, Dan Mosseley and Dan M. R. were valuable company during the lockdown.

Looking further back, many people were crucial in fostering my love of science. My uncle Jim and cousin Lisa were valuable early influences, as was my undergraduate supervisor, Professor Michael Freitag, and many of the members of his group. Professor Ryan Mehl introduced me to chemical biology and chemical protein modification, along with Kari van Zee, and I have been hooked ever since.

By far, my biggest supporters throughout my whole life have been my family. My parents, Gregg and Joyce, have been my biggest cheerleaders, both in academics and in all my other endeavors. Studying abroad presented its own set of challenges, and they have been incredibly supportive. Traveling home during the holidays, while difficult and less frequent during the pandemic, always provided me with a much-needed refresh. My sister Rachel, initially my adversary as a kid, later transformed into one of my best friends and companions, and one of the people I look up to the most.

And Christine, there are no words that can describe the role you have filled in my life. My closest confidant and a pillar of support from afar, the only thing better than reflecting on these past years is looking forward to my future with you.

Thank you to Patrick, Andrew and Adeline for invaluable feedback on my thesis, and the Clarendon Scholarship for funding my studies. To my future thesis examiners, thank you for your time and consideration, good luck.

Table of Contents

Introduction 9
Overview of protein post-translational modifications (PTMs) 10
Background 10
Types and roles of PTMs 13
Overview of epigenetics and histone PTMs 17
Background 17
Types and roles of histone PTMs 18
Outstanding challenges in studying histone PTMs 21
Current approaches to creating desired histone PTMs 24
"Histonefection": an underappreciated ability of histones to penetrate cell membranes and localize to chromatin 31
Goals and scope of this Thesis 38
Goals and scope of Chapter 1 38
Goals and scope of Chapter 2 39
Goals and scope of Chapter 3 41
Chapter 1: Development of Novel Chemical Methods for Introducing PTMs into Histones 43
Introduction 44
Recent strategies in chemical protein modification 44
Optimizing a light-driven radical approach to dehydroalanine based PTM installation 50
Previous work within the group 50
Matching photocatalyst strength with sidechain precursor 56
Testing catechol variants 65
Protein scope, compatibility, and scalability 66
Substrate scope 68
Characterization of installed protein modifications by HPLC and MS 71
Substrate scope limitations 78
'On-protein' chemistry afforded by the installation of electrophile containing sidechains 79
Installation of various electrophile containing sidechains 79
Slow halogen exchange of alkyl halide sidechains 81
Reactivity of alkyl halide sidechains with "off-protein" nucleophiles 83
Conclusion 85
Summary and advantages brought about by this chapter 85

Prospects for future work....87
Statement on collaboration....89
Experimental Details....90
General Methods....90
Protein Production....96
Protein modification characterization details....99
Chapter 2: Investigating Novel Applications with Chemically Modified Histones....145
Introduction....146
Current applications of chemically modified histones....146
Properties of chemical tools needed to study histone PTMs....148
Goals of this Chapter....150
Protein-protein crosslinking with “on-protein” electrophiles....153
Overview of current protein-protein crosslinking methods....153
Crosslinking KDM4A to histone H3 via electrophilic sidechains....156
Validation of crosslinking via LC-MS/MS and Zn ejection assay....159
Potential for site-specific discovery of unknown interaction partners....163
Investigating histone H3 tail structure in response to phosphorylation using Boronoalanine (Bal) as an NMR Probe....165
Previous work within the group....165
Creation of Bal containing nucleosomes....168
Structure insights into the histone tail in response to phosphorylation revealed by Bal mediated DICE NMR....170
Investigating histone modifying enzyme function with installed native and unnatural histone PTMs....173
Investigating histone deacylation with Sirt2....173
Production of histone H3-Me_2/Me_3Lys36 containing nucleosomes....176
Tracking PTM status, enzyme processing, and complex formation with γCF_2-containing PTM sidechains....180
Previous work within the group....180
Rationale for incorporating fluorine into proteins for ^{19}F-NMR....182
Tracking histone PTM processing via ^{19}F-NMR....184
Significance of methionine as oxidation PTM site and relevant oncohistone mutation....186
Distinguishing Met oxidation PTMs via ^{19}F-NMR....188
Following protein complex (histone octamer) formation via ^{19}F-NMR....191
Conclusions....193
Summary and advantages brought about by this chapter....193
Prospects for future work....196
Statement on collaboration....198

Experimental Details 199
General Methods 199
Protein modification details 218
Chapter 3: Incorporation of Extracellular Histones into Chromatin and Exploring *In Vivo* Histone Modification Chemistry 239
Introduction 240
Background on extracellular histone incorporation 240
Initial investigations into histone uptake into HeLa cells 242
Choice of a “trackable” epitope tagged histone H3 242
Initial investigations into histone H3 uptake into Hela cells 244
Transcriptional response of HeLa cells to extracellular histone uptake assessed by RNA-seq 247
Histone incorporation, uptake validation, and RNA prep 247
RNA-seq results and analysis 250
ChIP-seq and future plans 253
Tracking native and radiolabelled histones injected intravenously or intracranially into rodents 255
Mouse tail vein injections of WT and ^{18}F-radiolabelled histones 255
Stability of histones and nucleosomes in blood plasma and serum 258
Direct injection of free histones into rat brains followed by detection in chromatin 260
Initial progress at developing Zebrafish as a model organism for *in vivo* light-driven chemical protein modification 263
Zebrafish as a model organism for light-driven protein modification 263
Initial attempts at *in vivo* protein modification and toxicity screens 265
Development of a platform for histone microinjection, retrieval, and detection in zebrafish larva 270
Conclusion 274
Summary and advantages brought about by this chapter 274
Prospects for future work 276
Experimental Details 279
General Methods 279
Protein modification details 285
Concluding Remarks 286
Outlook and lessons learned 287
References 291

Introduction

“The field of epigenetics, where everything is true, but nothing is accurate”

-Professor Michael Freitag, Oregon State University

Overview of protein post-translational modifications (PTMs)

Background

In the mid-20th century, the Central Dogma of Biology was proposed by Francis Crick[1], describing the transfer of information between the sequences of three biomolecular polymers: DNA, RNA and proteins (**Figure 1A**). This fundamental process underlying all of biology sparked research into how these sequences passed on information. Techniques like mass spectrometry[2] and DNA sequencing[3] allowed scientists to identify specific biomolecules and resolve their sequences, while technologies like NMR[4], X-ray crystallography[5], and Cryo-EM[6] allowed scientists to decipher biomolecular structure, giving insights into how their sequences brought about function.

Recently, relative to the overall history of science, researchers from around the world undertook a massive challenge, sequencing the entirety (disregarding long repeat rich telomere and centromere regions[7]) of the human genome[8-10]. The Human Genome Project cost an estimated $5 billion over 13 years and spurned the development of technology that allowed for entire genomes to be sequenced with base pair precision. However, as science often goes, sequencing the human genome ended up revealing many more questions than answers. Notably, the human genome contained far fewer genes than expected for an organism of our size and complexity[11]. Upon further investigation, the expected increase in complexity required for a fully functional and diverse proteome could be explained via several processes, a few are listed below:

- *RNA processing:* Between transcription and translation, pre-mRNA is processed into mature mRNA[12,13]. This can involve several processes that allow for the mRNA originating from one gene to be formed into several different protein isoforms. If a

gene contains within it several introns and exons, a process called alternative splicing can create variation in the final protein product by altering which introns are removed and which exons are utilized in the final mature mRNA[12,13].

- *Epigenetics:* Throughout the genome, specific regions and genes can be either activated (in euchromatin) or silenced (in heterochromatin) via the covalent modification of both DNA and histones (the protein component of nucleosomes, consisting of an octamer with two each of histone H2A, H2B, H3 and H4)[14,15]. These modifications can either change the physical properties of proximal chromatin to make it more accessible or restricted to the transcriptional machinery, or the modifications can recruit various chromatin remodeling enzymes that change chromatin structure[14,15]. This allows a single genome to differentiate the functions of multiple cell types in an organism, as well as change expression profiles throughout developmental phases, and in response to certain stimuli. This process will be covered in more detail later in the **Introduction**.
- *Post translational modifications (PTMs):* After proteins are translated, their primary sequence only contains up to 20 canonical amino acids (plus selenocysteine[16] and pyrrolysine[17] in special cases), restricting their potential chemical functionality. However, after translation, dozens of chemical modifications can be appended to specific protein sidechains[18,19]. These various modifications (a summary is given in **Figure 2**) can drastically alter the physical and chemical properties of their protein residue, and contain a much higher chemical functional diversity than the canonical amino acids[20]. These PTMs, resulting from glycosylation, methylation, acylation, phosphorylation, oxidation, and many other transformations,

exponentially increase the available structural and functional space that a single protein can access from its primary sequence alone[19,21].

Overall, the processes above help exponentially expand the complexity and diversity of the proteome beyond what can be communicated by the primary sequences of DNA to RNA to protein. We now understand that, while the Central Dogma of Biology is insightful into the inner workings of biological function, it should be interpreted as only part of a much more complicated story (**Figure 1B**). Of particular interest to this thesis is the fact that two distinct processes highlighted above, *epigenetics* and *PTMs*, are intimately connected, as a crucial portion of epigenetic processes are governed by the PTMs that are present on histone sidechains[14,15]. These processes will be further described below.

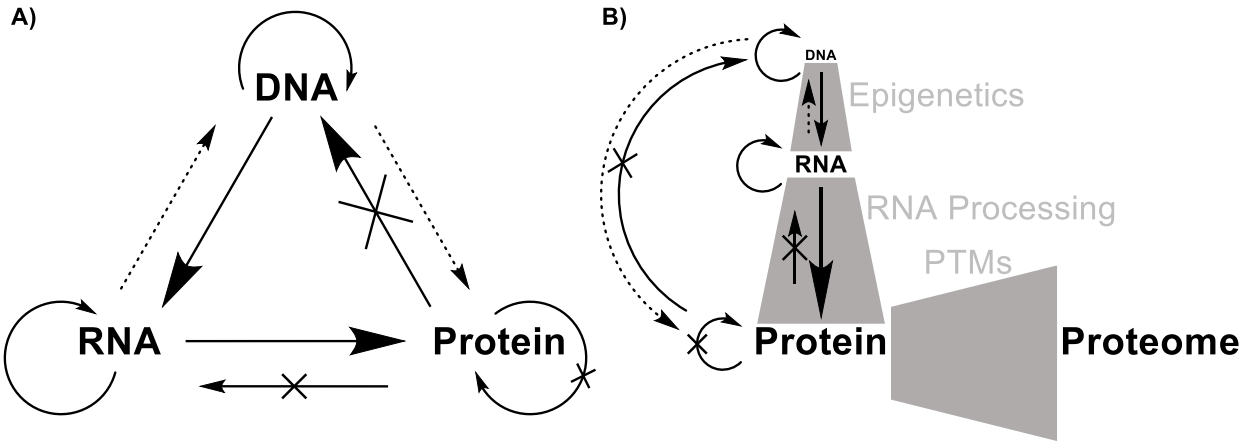


Figure 1: A) The Central Dogma of Biology as proposed by Francis Crick[1]. Big arrows designate major pathways of information flow, dashed arrows indicate process that are possible only under special circumstances. **B)** An expansion on the Central Dogma highlighting how non-sequence derived complexity and diversity can be added during each step of DNA to RNA to protein information transfer (gray).

Types and roles of PTMs

Post-translational modifications are naturally installed on protein sidechains, termini, and backbone sites through a variety of methods[18,19,21]. While a plethora of enzymes are known to both install and remove PTMs (writing and erasing)[19], protein sidechains can also be covalently modified in response to their redox environment, such as in the presence of reactive oxygen species (ROS)[22]. Some protein sidechains can also spontaneously rearrange over time or in the presence of heat or favorable pH[23]. Regardless of their method of installation, enzymatic or not, PTMs change the intrinsic chemical and physical properties of the site they are located, therefore altering the structure and function of their host protein, sometimes drastically. The following list gives a brief overview of several classes of PTMs, with relevant examples. See **Figure 2** for a list (non-comprehensive) of many common PTMs on their associated sidechains.

- *Enzymatically installed sidechain PTMs:* Many of the most studied PTMs can be both installed and removed by specific enzymes[19]. Often, this is done via the utilization of specific cofactors and other compounds as the source of the PTM chemical precursor. For example, ATP is used by kinases to facilitate phosphorylation (on serine, threonine, and tyrosine)[24,25], acetyl-CoA (and many other acyl variants) are used by acetyltransferases to acetylate lysine[26-28], and *S*-adenosyl methionine (SAM) is used as a methyl group donor for lysine and arginine methylation[29,30]. The ubiquitous use of these cofactors and compounds was vital for early studies elucidating the targets of PTM installing enzymes. For example, the phosphorus in ATP[31], or the carbons in acetyl-CoA[32] or SAM[33], could be radio-labeled and therefore tracked once installed onto their targets.

- *PTMs resulting from redox environment:* Many protein sidechains are highly sensitive to their redox environment. Perhaps the best, and most studied, example of this is the interplay between disulfide bridges and the corresponding reduced free thiols in cysteine[34]. Depending on the redox environment, cysteines can be directed to either form disulfide bridges vital for internal structure or protein-protein complex formation, or released as a free thiol, with distinct nucleophilic reactivity often found in active sites[34]. However, there are many other sidechains that are highly sensitive to their redox environment. Reactive oxygen species can drive the oxidation of multiple residues, including cysteine (to sulfenic, sulfinic or sulfonic acid)[34], methionine (to sulfoxide and sulfone)[35,36], and tyrosine (nitration to nitrotyrosine)[37], to name a few. It would be imprudent to judge these various oxidation products as random and unavoidable consequences of oxidation, as they are often mediators, sinks, or detectors of oxidative damage, altering their host protein's function accordingly in response to their redox environment[22,35,38]. Furthermore, there are often highly specific enzymes dedicated to the removal of these oxidized sidechain PTMs, such as in the case of the sulfoxide form of oxidized methionine, which has dedicated methionine reductase enzymes exclusive to removing either the R or S epimers of the sulfoxide[39].
- *Spontaneously formed PTMs:* Some PTMs are formed not due to redox environment or enzymatic activity, but due to unavoidable and spontaneous reactivity of specific sidechains within the protein. These reactions, and resulting PTMs, can determine the lifetime of the protein, by building up protein damage over time from accumulated modifications altering structure and function[40]. Several

examples are the deamidation of asparagine and glutamine (forming aspartic acid and glutamic acid, respectively)[41], the isomerization of aspartic acid to isoaspartic acid[41], and other slow processes such as the racemization of the protein sidechains at increased pH[42,43].

- *Backbone and protein termini associated PTMs:* Functional groups other than the those found on the sidechains of proteins can also be post-translationally modified, such as the N- and C-termini, and amide backbone of the protein. While often found on smaller peptides, backbone *N*-methylation can protect the residue from susceptibility to proteolysis, as well as alter its ability to form secondary structures by removing hydrogen bonding ability[44,45]. On the protein N-terminus, acetylation, and other acylations such as myristoylation and palmitoylation, are commonly found[46]. This drastically alters the size, charge, and hydrophobicity of the N-terminus, and is believed to be irreversible[46]. The C-terminus can also be modified, with both methylation and amidation often occurring in peptides[47,48].

Through the many mechanisms discussed above, a plethora of different PTMs can be installed throughout the many residues and functional groups found in proteins (**Figure 2**). Defining specific roles of each PTM type would be a fruitless task, as PTMs are related to essentially every cellular process involving proteins. PTMs are, through their own composition, inseparable from processes such as metabolism (e.g. acetylation with Acetyl-CoA)[26], energy economy (e.g. phosphorylation with ATP)[24], redox environment (e.g. oxidation of methionine and other residues with ROS)[22], and countless others. In this thesis, we will primarily focus on histone PTMs (discussed below), their association to important epigenetic processes, and how to make and study them using chemical tools.

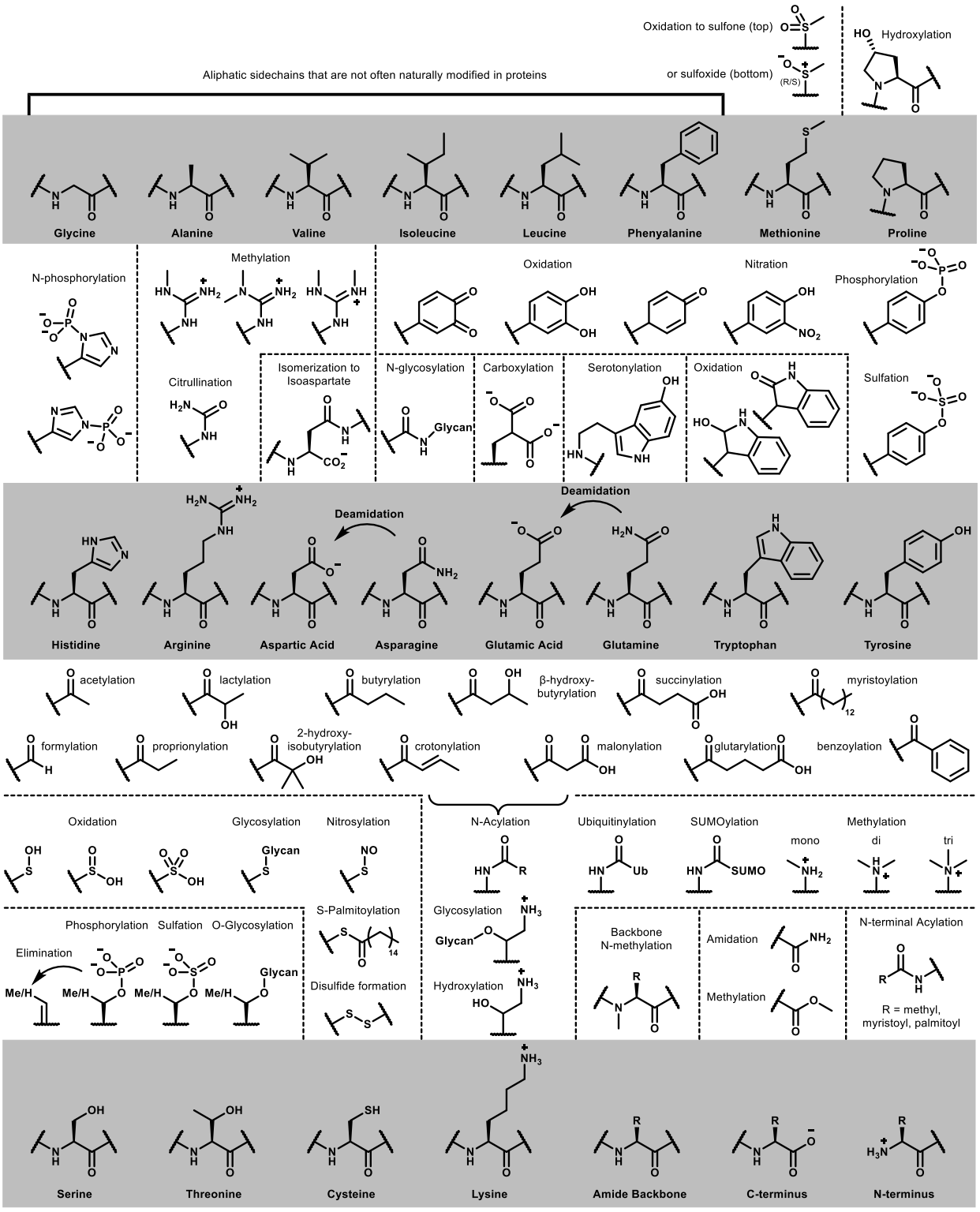


Figure 2: A list of common PTMs above their respective amino acid sidechains (in gray boxes). This non-comprehensive list also covers possible termini and backbone PTMs.

Overview of epigenetics and histone PTMs

Background

Epigenetics is defined as “the stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence[49].” This arises from many mechanisms, such as DNA modification[50], histone PTMs and variants[51], RNA interference[52], and others. All these processes work together to change gene expression independent from the primary sequence of the DNA. Functionally, this happens by controlling both the physical accessibility of chromatin via its level of compaction[53], or by altering expression levels biochemically, such as by driving or blocking recruitment of transcription initiation complexes or controlling the fate of transcribed mRNAs[15,50,51].

Open and accessible chromatin is called euchromatin, while compacted and inaccessible chromatin is called heterochromatin[54]. Heterochromatin is further divided into two subgroups: constitutive heterochromatin consists of tightly packed DNA mostly devoid of genes and primarily containing repeat rich regions in the centromeres and telomeres[55], while facultative heterochromatin contains genes necessary for cell differentiation or development that need to be activated at specific times, hence its ability to become transcriptionally activated given the right conditions[56,57].

Histone PTMs, a major focus of this thesis, are a crucial part of defining and maintaining the correct epigenetic state in proximal regions of chromatin[15,50,51,54]. They go beyond simply promoting gene activation and repression, but are also associated with nucleosome stability, DNA damage, and many unknown functions. A substantial effort has been made to decipher their various functions and synthesize them chemically.

Types and roles of histone PTMs

Histones are modified with a wide variety of PTMs[51], including acetylation[58] (and other acylation variants)[59], methylation[60,61], phosphorylation[62], glycosylation[63], oxidation[64], and others[51]. It is important to note that the type of PTM does not always define its expected effect on chromatin. Some modifications, such as lysine acetylation, are always associated with increased transcription, in part due to their neutralization of the positive charge on lysine, reducing the binding strength between the histone and DNA and therefore increasing chromatin accessibility to transcriptional machinery[58]. However, other common PTM types, such as lysine methylation, are highly dependent on the site, extent of methylation, and context of surrounding residues and their PTMs[61]. These PTMs are precisely installed by enzymes called “writers,” removed by “erasers”, and detected by “readers,” in a complicated interplay that governs chromatin state, responds to DNA damage, differentiates cell cycle stages, and distinguishes important genomic elements such as transcriptional start sites and kinetochore attachment sites in the centromere[65-67]. Some examples are given below showing the complex roles of histone PTMs.

- *Extent of PTM installation:* Histone H4 can be mono-, di-, or trimethylated at the K20 site, with all three methylation states conferring distinct functional outcomes[68-72]. Each lysine methylation state can be distinguished by readers, writers, and erasers by their different hydrogen bonding abilities and cation-π interactions, allowing for different functional outcomes from each methylation (**Figure 3A**).
- *Histone PTM crosstalk:* PTM status be affected by surrounding PTMs, called PTM crosstalk. For example, the phosphorylation of histone H3 S10 blocks the reading of trimethylation of the proximal residue K9 by Heterochromatin Protein 1 (HP1), a

transcriptional repressor (**Figure 3B**)[73]. Crosstalk can also occur between histones in the same nucleosome, and between different nucleosomes. A well-known example of this is where the histone H3 K27 methyltransferase complex, called polycomb repressive complex 2 (PRC2), requires one of its protein subunits to bind its PTM product (K27me3) on one nucleosome to activate methylation on the next nucleosome, in effect "walking" the complex along the chromatin and spreading the facultative heterochromatin mark (**Figure 3C**)[74,75].

- *Site dependency on the effects of PTMs:* Unlike acetylation, many PTMs have different effects depending on their site. A striking example of this is the completely different functions of trimethylation at histone H3 at sites Lys4, 9, and 27. Me_3Lys4 marks euchromatic regions, Me_3Lys9 marks constitutive heterochromatic regions, and Me_3Lys27 marks facultative heterochromatic regions (**Figure 3D**)[68]. What makes these differences more fascinating is the fact that, like many histone PTMs, they are located on the disordered N-terminal tail of histone H3, and two of them, K9me3 and K27me3, have the same surrounding sequence (AR<u>K</u>S). The ability of epigenetic machinery to discriminately read, write, and erase these marks speaks to the level of intricacy and precision governing epigenetic processes.
- *Novel PTMs and unknown functions:* Perhaps unsurprisingly, many PTM types and sites are currently unknown or uncharacterized. Just in the last few years, two new PTMs have been discovered and shown to have distinct effects on chromatin. Histone benzoylation was found to occur in response to eating the common food preservative sodium benzoate[76], and histone serotonylation at histone H3-Gln5 crosstalks with histone H3-Me_3Lys4 in euchromatin (**Figure 3E**)[77].

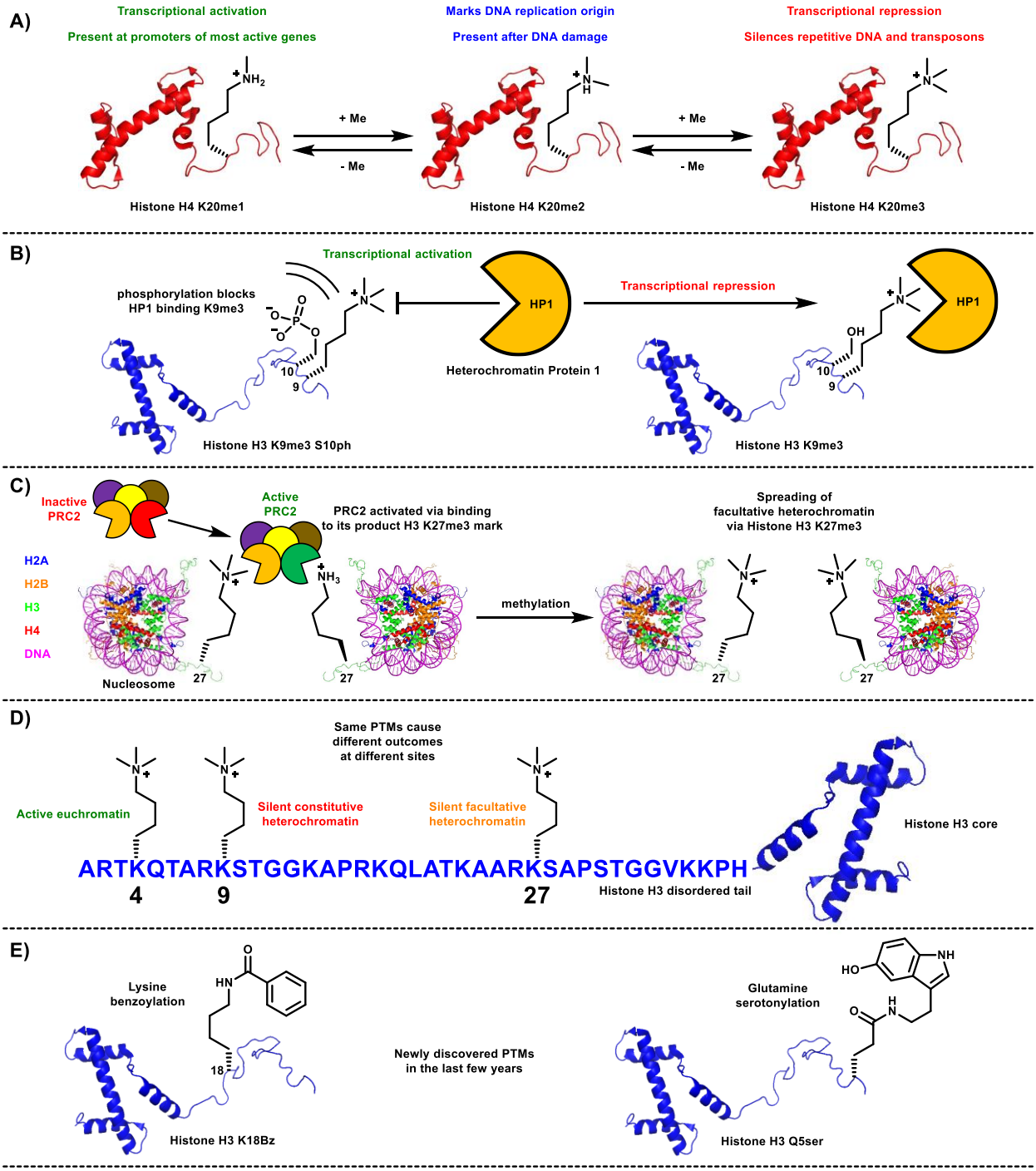


Figure 3: A) Each successive methylation can have a different function. **B, C)** Histone PTMs can crosstalk to other PTMs, both with nearby residues **(B)** and between adjacent nucleosomes **(C)**. **D)** The same PTM can have different functions at different sites. **E)** Two histone PTMs, serotonylation and benzoylation, were discovered only recently.

Outstanding challenges in studying histone PTMs

Due to the importance of histone PTMs in controlling transcription, and by consequence, prevalence in disease and cancer when PTM mismanagement occurs[78], there has been a tremendous effort to develop techniques that allow for their study. Methods have been developed that allow for the precise identification of PTMs and the detection of new ones, locating genomic regions where various histone PTMs reside, and measuring the transcriptional and chromatin environments that histone PTMs are associated with, among others. These methods, some of which are described further below, are all used to try and decipher the "histone code," that is, how the various PTMs imparted on histones encode their ability to control how genetic information is transcribed[54].

One of the best ways to study histone PTMs is to first isolate the specific chromatin regions they are associated with. Chromatin immunoprecipitation (ChIP) does just that by extracting the chromatin fraction (often crosslinked with formaldehyde) from a cell or tissue, shearing the DNA between nucleosomes to obtain a soluble nucleosome mixture, then pulling down nucleosomes with an antibody raised against a specific histone PTM of interest[79,80]. This provides a mixture of nucleosomes that all contain the histone PTM of interest. Then, methods such as mass spectrometry and western blotting can determine the presence and quantity of other PTMs or crosslinked binding partners[81-83].

Furthermore, the DNA from these nucleosomes can be isolated, sequenced, and aligned with a reference genome in a method called ChIP-seq to quantify where the histone PTM is localized in the genome[84]. These results, paired with RNA-seq, which reveals transcription levels across the whole genome, provide genome-wide correlations between specific histone PTMs and gene expression[85]. Other sequencing methods such as DNase

I hypersensitive sites sequencing (DNase-seq)[86] and Chromosome conformation capture (3C)[87] can determine chromatin accessibility and spatial organization, respectively. All these methods combined provide a strikingly precise (often down to the atom) characterization of chromatin across the entire genome. However, there are many caveats to these methods that leave crucial details about the mechanisms of histone PTM function unresolved.

First, as described above, many of the techniques require the use of a histone PTM specific antibody to enrich the PTM of choice from a mixture of chromatin. While histone PTM antibodies are often very precise[88], they require herculean standards of specificity. Antibodies must be able to consistently bind their target PTM at the correct site and histone, while ignoring the same PTM type on nearby residues or histones. Often this means discriminating between changes as small as a single methyl group on a lysine. Further compounding the problem is the prevalence of multiple PTMs on the same histone. Important instances of PTM crosstalk between adjacent modified residues, such as in the case described above between histone H3 K9me3 and S10ph, can be entirely missed as one PTM may sterically prevent antibody binding to the other[89]. Additionally, high quality antibodies do not exist for all PTMs, and many antibodies in past and current use are known to have cross-reactivity with other PTMs[88,90,91].

Secondly, while sequencing methods are incredibly powerful, they often only provide a correlative picture between the localization of a specific histone PTM and the measured chromatin state[92]. This correlation vs causation problem is hard to resolve, as many methods do not decipher whether the existence of a PTM at a specific site is causing the observed chromatin state or was installed in response to the chromatin state.

Furthermore, if a multitude of histone PTMs are contributing to an observed chromatin state, then methods that can only resolve one PTM at a time will only reveal a small portion of the mechanistic picture[93].

An increasingly popular approach to addressing the above problems is to synthesize the desired PTM containing histone substrates, then assemble them into more complex chromatin structures such as mono- and polynucleosomes[94,95]. This overcomes the problems associated with antibodies as the synthetic substrates can be made homogenously purified and fully characterized. Furthermore, observed effects resulting from PTM containing chromatin substrates can show causation when compared with unmodified controls. The next section will further explore current chemical methodologies for synthesizing PTM containing histones.

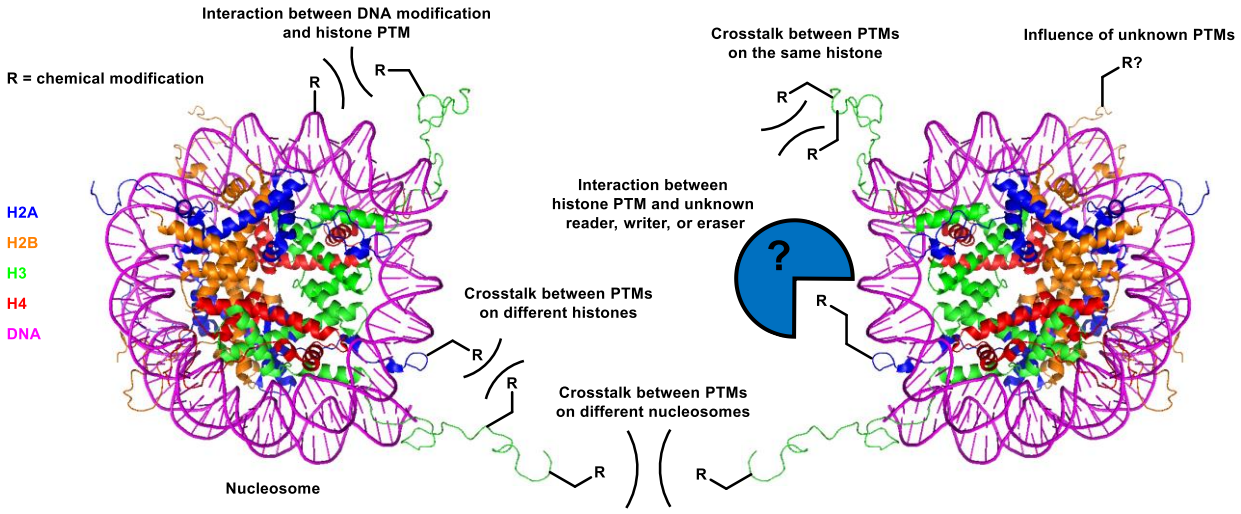


Figure 4: An overview of important histone PTM interactions in chromatin that are difficult to study with current techniques.

Current approaches to creating desired histone PTMs

As discussed in the previous section, chemical tools that allow for the synthesis of homogenously modified histones are highly sought after in the field of epigenetics[94,95]. However, as is common throughout the field of Chemical Biology, there are a wide range of tools for creating histone PTMs, each with its own distinct advantages and disadvantages. Furthermore, there is a spectrum of complexity in the epigenetic substrates of interest one can work with. In the simplest case, short peptides mimicking PTM containing histone tails are easily accessible, but as you move up the scale of complexity (and biological relevance) to full histone proteins, histone complexes, nucleosomes, polynucleosomes, and eventually to chromatin within living cells or organisms, the number, capabilities, and simplicity of the tools available to chemically create and study specific histone PTMs is drastically reduced[93]. Pairing both a practical yet usefully complex epigenetic substrate with the right chemical tool to install the PTM of interest requires careful consideration. Additionally, when designing new chemical tools for synthesizing histone PTMs, one must always consider the compatibility of the chemistry with the target epigenetic substrate, which in the most complex applications may be chromatin within a living cell. A summary of the major approaches used to create histone PTMs is given below, with both their advantages and disadvantages listed.

Genetic Code Expansion (GCE)

GCE is a technique that reassigns a rarely utilized redundant codon, often the amber stop codon in amber codon suppression, to allow for a noncanonical amino acid (ncAA) to be incorporated in its place during cellular ribosomal protein translation (**Figure 5A**)[96-98]. This has been achieved by the development of an amber codon targeting tRNA/aaRS (transfer RNA/aminoacyl-tRNA synthetase) pair that can efficiently charge the ncAA of interest while discriminating against native amino acids, often accomplished by mutating an orthogonal tRNA/aaRS pair from an evolutionarily distant organism like archaea. Common archaea derived tRNA/aaRS pairs such as $tRNA_{CUA}$/PylRS and $tRNA_{CUA}$/TyrRS (subscript details the anticodon to the amber stop codon) can efficiently incorporate over 100 structural ncAA variants of their original amino acid targets[96-98]. These GCE systems can work in a variety of different cell types, from the protein production workhorse *E. coli* to more complex living organisms[96-98].

- Advantages: The main advantage of using GCE is that, due to the necessity of hijacking *in cellulo* translational machinery[96], the method can express histones containing PTMs of interest *in vivo*, allowing histone PTM function to be studied in its native chromatin environment in living cells. In addition, if the PTM precursor amino acid is easily accessible, the method is fairly straightforward for a non-chemist, requiring only a slightly augmented protein expression protocol with extra plasmids for the tRNA/aaRS pair[98]. Furthermore, there are many tried and tested ncAAs to choose from that can be incorporated as histone PTMs[99], such as a multitude of acyllysine variants (**Figure 5B**)[99-101], and phosphorylated versions of Ser[102], Thr[103], and Tyr[104].

- Disadvantages: Historically, ncAA containing proteins created by GCE have had very low expression yields[98], as the orthogonal tRNA/aaRS pair is not as efficient as the native system and must compete with Release Factor 1, which causes translation termination and a truncated protein product[105]. In addition, many PTMs are not easily accessible as their respective ncAAs are either synthetically challenging to make in the quantities needed, toxic to the cell, have difficulty passing through the cell membrane (such as in the case of charged ncAAs), or have simply not been shown to work yet[94,97,98]. Notably, the crucial lysine methylation histone PTMs have been particularly difficult for GCE to address, as the minute change in size of each added methyl group is difficult for a tRNA/aaRS pair to select for against a nonmethylated (unmodified) lysine. Only the mono- and dimethyllysines have been created in a protected form, requiring further chemical steps to reveal the desired PTM, incompatible with *in vivo* experiments[106,107].

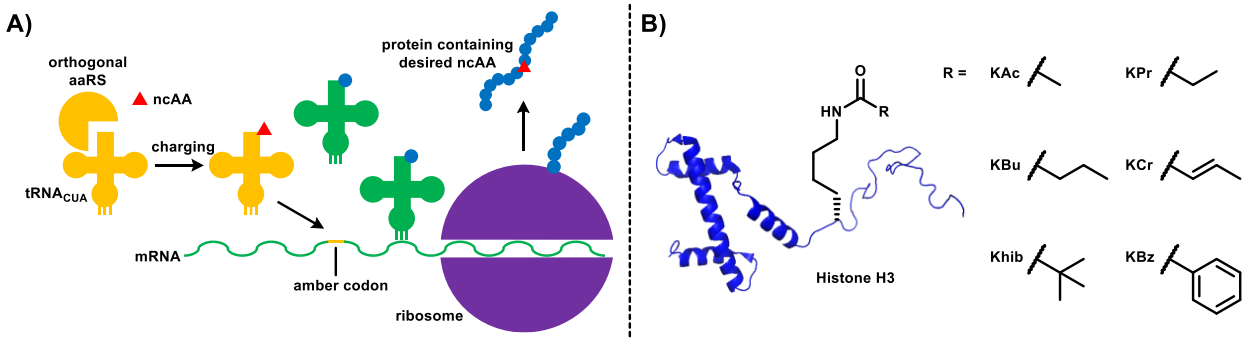


Figure 5: A) Schematic showing the mechanism of amber codon suppression allowing for the incorporation of a desired ncAA into a protein of interest. **B)** Several examples of epigenetically relevant acyllysine histone PTMs successfully installed with GCE.

Protein semi-synthesis

Perhaps the most chemically straightforward approach to making PTM containing histones is to avoid ribosomal translation altogether and construct the protein through solid-phase peptide synthesis (SPPS)[108]. However, while SPPS can incorporate a vast number of amino acid derivatives beyond the canonical 20, it is notoriously limited in polypeptide length[109]. The invention of native chemical ligation (NCL)[110] and expressed protein ligation (EPL)[111] dramatically improved the limitations on length by allowing for peptide fragments to be ligated to other peptide fragments (NCL) or expressed truncated proteins (EPL). This is conventionally achieved via an N-terminal Cys and C-terminal thioester undergoing a *trans*-thioesterification followed by an *S* to *N* acyl shift to give the native amide bond (**Figure 6A**). Many variants of this strategy allow for residues other than Cys to be used[112].

A more recent approach uses split inteins to ligate together two peptide fragments (one half may be an expressed protein fragment) of choice while simultaneously excising themselves from the product peptide chain[113,114]. Impressively, the split intein approach has potential for installing histone PTMs *in vivo*. A truncated histone bearing one of the intein halves was expressed in living cells, and upon endocytic uptake of a peptide containing the other intein half and a histone PTM laden tail, the PTM containing tail was successful fused onto the rest of the histone body (**Figure 6B**)[115,116].

- Advantages: In contrast to GCE (above), protein semi-synthesis methods are not limited by what ncAAs are compatible with orthogonal tNRA/aaRS pairs or what can fit through the ribosome. Therefore, more protein sidechains are accessible, even with large PTMs such as glycosylation or ADP-ribosylation[95]. Furthermore,

yield is not lost with multiple PTM installations, and many different PTMs can be incorporated into the same protein. While most PTM containing proteins are created by semi-synthesis *in vitro*, recent work on ultrafast trans-splicing inteins is a promising platform for more installing histone PTMs *in vivo*[115,116].

- Disadvantages: Peptide synthesis and ligation chemistry are technically challenging and require specific equipment. Large proteins may need complicated protection and deprotection schemes to link all ligation fragments while leaving PTMs intact and may be difficult to refold properly after ligations[94,95]. Some PTM containing building blocks are also difficult to couple during peptide synthesis and are prone to epimerization[94]. Lastly, while recent work on ultrafast trans-splicing inteins is encouraging for installing histone PTMs *in vivo*, it requires chromatin-wide incorporation of engineered histones truncated with intein fragments, potentially disrupting native chromatin structure[115,116].

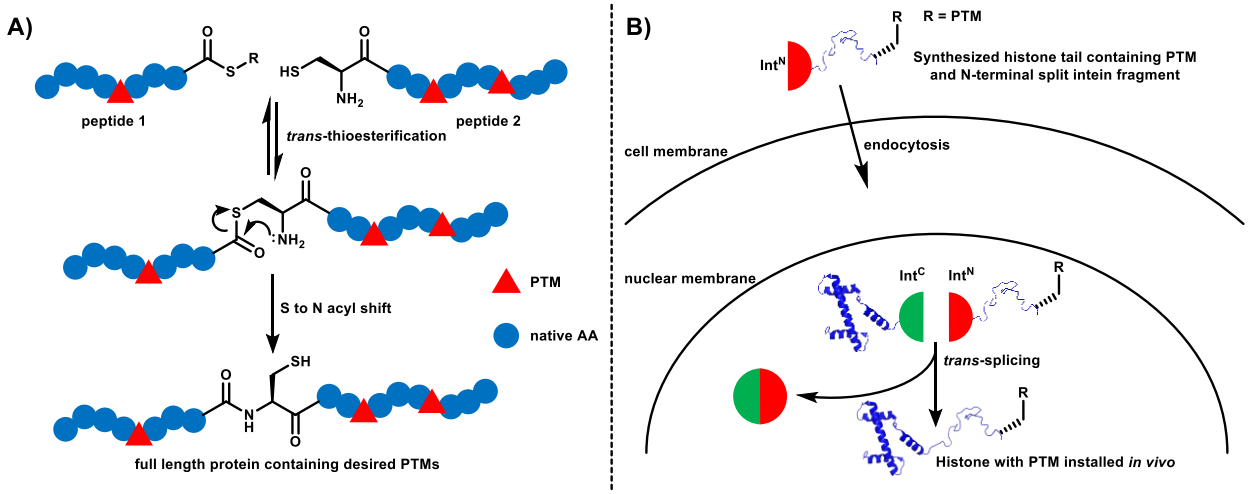


Figure 6: A) The mechanism of native chemical ligation. **B)** Ultrafast trans-splicing inteins allow for the installing of histone PTMs *in vivo*.

Chemical mutagenesis

A third strategy for creating histones bearing PTMs utilizes precise chemical tools to selectively react at specific residues within the protein, in essence, “chemically mutating” the site to another residue[117,118]. Cysteine is commonly targeted due to its unique chemical properties (highly nucleophilic and relatively low pKa ~ 8.5) and rarity on the protein surface[118], although other residues such as lysine[119], tryptophan[120], tyrosine[121], methionine[122,123], and the C-terminus[124] have all been successfully targeted as well. Early versions of Cys selective protein chemistry utilized Cys alkylation with ‘off-protein’ electrophiles such as alkyl halides to install PTM mimics such as thia-lysine and thia-arginine analogues[95,118,125]. However, the sulfur at the sidechain gamma position imparts physical and chemical distortions to the desired PTM containing sidechain[94,95].

Later attempts used a ‘tag-and-modify’ approach (**Figure 7A**)[126], where the Cys was first converted to the unique and reactive electrophilic residue, dehydroalanine (Dha), via a bis-alkylation/elimination reaction, among other methods[127]. ‘Off-protein’ *S*-[128,129], *Se*-[130], *P*-[131], or *N*-nucleophiles[132,133] could then be selectively added to the Dha residue, creating novel sidechains and avoiding the unwanted reactivity of ‘off-protein’ electrophiles with the plethora of other nucleophiles found in proteins. More recently, a radical based approach used the SOMOphilic properties of Dha to capture ‘off-protein’ carbon-centered alkyl radicals generated from alkyl halides and sodium borohydride[134-136]. While the method required use of a glovebox (O_2 free), it succeeded in recreating the native βC-γC bond of the protein sidechain, allowing for the installation of dozens of native, modified, and unnatural sidechains into a range of protein substrates (**Figure 7B**). Furthermore, histones were easily modified with PTMs such as lysine and arginine methylation.

- Advantages: Chemical mutagenesis methods are often cheap and technically simple for non-chemists, as they do not require special protein expression or partial protein synthesis and involve few steps with selective and robust chemical reagents[136]. While Cys is commonly targeted for Dha based chemistry, many other residues can be used as Dha precursors as well, such as pSer and several SeCys variants[127]. Furthermore, protein size is not limiting, as in protein semi-synthesis, and the methods can install a huge variety of protein sidechains[136].
- Disadvantages: A large disadvantage of using Dha based chemical mutagenesis is the creation of a mixture of diastereomers (D and L) as there are currently no methods for stereoselective Dha functionalization[127]. Furthermore, radical based Dha modification methods require the use of a glovebox[135,136], are not currently tolerant of a wide variety of redox sensitive functional groups, and only work *in vitro*, as the chemistry is not yet biocompatible.

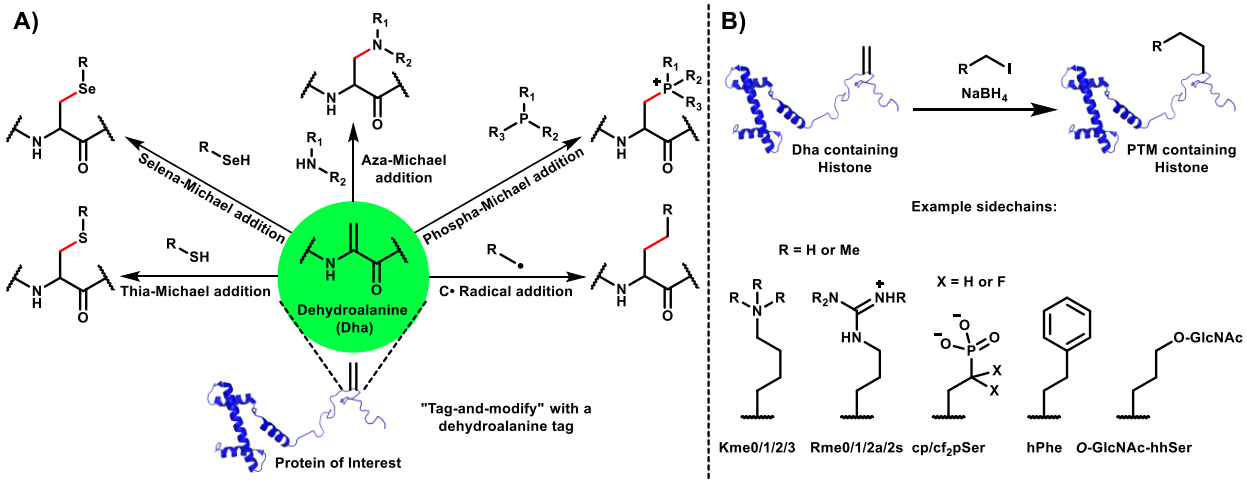


Figure 7: A) The Dha 'tag' can be 'modified' via several reaction manifolds. **B)** Carbon-centered alkyl radical addition to Dha can create many natural, modified, and unnatural sidechains in proteins.

“Histonefection”: an underappreciated ability of histones to penetrate cell membranes and localize to chromatin

The bulk of study surrounding the biological roles of histones has focused on their critical jobs as DNA binders in the nucleus, specifically, their function as an octamer within the smallest subunit of chromatin, the nucleosome[137,138]. This makes sense, as although they are expressed in the cytoplasm, they are immediately bound by chaperones and transported into the nucleus, where their role in controlling gene expression through regulating chromatin accessibility and accumulating PTMs is of intense study[139]. However, there is a small area of research that has investigated the functions of extracellular histones, which are released after cell death due to injury or disease[140-142]. These extracellular histones have distinct abilities that are perhaps unexpected for a protein that is almost entirely confined to the nucleus under normal conditions.

Firstly, free extracellular histones strongly activate the immune system, leading to inflammation and cell death[140-142]. Secondly, and perhaps more intriguingly, extracellular histones can rapidly penetrate cell membranes, and by virtue of their native nuclear localization sequence[143], can localize to the chromatin of the cells they have entered[144]. These two qualities of extracellular histones have both been noted independently in literature, but a link has yet to be established between the ability of extracellular histones to incorporate themselves into the chromatin of surrounding cells and those cells’ robust inflammatory response to the uptake of extracellular histones.

The ability of free histones to penetrate cell membranes has been known for decades[144,145]. As far back as the 1960s, scientists started noticing the effects of adding calf thymus histones to cells[146,147]. Plasmid DNA[148-150], RNA[150] and proteins such as

BSA[147,151] could all be dragged into cells along with the free histones, earning the method the name “histonefection,”[144] an alternative method for facilitating uptake of desired DNA or proteins into cells than more conventional means at the time.

All core histones (histones H2A, H2B, H3 and H4) as well as linker histone H1 exhibit the ability to penetrate cell membranes in multiple mammalian cell lines[144,151], yet their mechanism of membrane penetration remains unresolved. Most carrier/DNA complexes are understood to enter cells via an endocytic pathway[152,153] and using histones as carriers for DNA and other proteins may function similarly. A histone H1/DNA complex was found to localize within endosome vesicles[154], and a histone/BSA protein complex (attached covalently) was reduced by energy depletion[151], both suggesting endocytosis as the mechanism for cell penetration. However, studies on free histone penetration suggest an alternate pathway. Free histones have been observed penetrating cells unrestrained by low temperature, ATP depletion, high sucrose molarity, or in the presence of various endocytosis and phagocytosis inhibitors[151,155]. Studies on both protists and plant cells also suggest a non-endocytic uptake mechanism[156,157]. Furthermore, a handful of studies claim that histones directly cross cell membranes by passive diffusion, shown both in mammalian cell culture with HeLa cells and a model system of lipid bilayers in liposomes[157], though further validation is needed to confirm these findings.

Perhaps the best comparison of free histone uptake is with the popular poly-Lys and poly-Arg peptide sequences commonly used as cell penetrating peptides[158]. The N-terminal disordered histone tails are abundant with basic residues such as lysine and arginine and could therefore be incorporated via a similar mechanism as that of the heavily studied poly-Lys/Arg sequences, which are believed to enter via endocytosis[159,160]. Additional

validation of this hypothesis is the fact that, in some cases, the disordered and basic N-terminus is required for free histone penetration, as evident by the loss of cell penetration ability by histone H2A when the 37-residue N-terminal tail is removed, and ability of the 37-residue section to facilitate membrane penetration on its own[148].

Regardless of the mechanism, free histones do penetrate cell membranes, bringing them into the cytosol. From there, either their native nuclear localization sequence brings them to the nucleus[143], or the cell fails to differentiate them from newly expressed histones and transports them into the nucleus assisted by chaperones[139]. Histone H3 and H4 are known to be chaperoned together as a tetramer into the nucleus[139], and interestingly, mixing the two histones together prior to adding to the cells facilitates more efficient cell penetration[151].

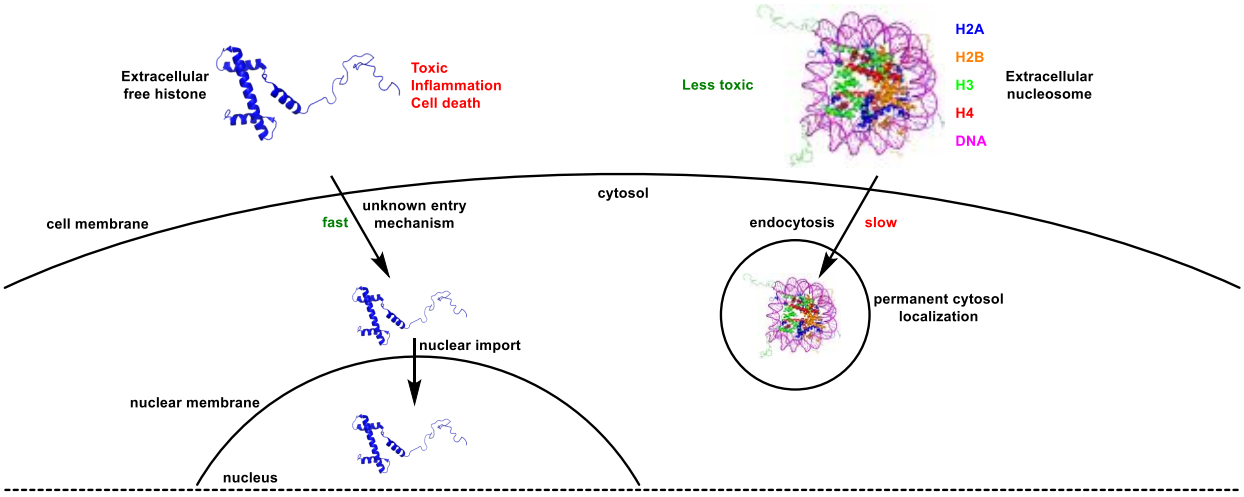


Figure 8: Extracellular free histones are toxic to cells, rapidly cross cell membranes by an unresolved mechanism, and localize to the nucleus. Extracellular nucleosomes are less toxic, enter cells slowly via endocytosis, and stay localized in the cytosol.

Beyond this, the fate of incorporated extracellular histones remains unknown and understudied. As the name suggests, the study of “histonefection” has primarily been studied as a technique for dragging other biomolecules of interest, such as DNA or proteins, into cells[144]. It is likely, but not confirmed, that extracellular histones are incorporated into the chromatin of the cells they enter. If so, their prior modification states, which may carry an ensemble of PTMs from the diseased or damaged cell that they came from, could influence the gene expression of the chromatin into which they become incorporated. This leads to some intriguing questions:

- Do extracellular histones get incorporated into cells that they penetrate?
- What is the response of the cells to extracellular histone incorporation?
- How long do incorporated histones last in the cells they penetrate?
- Do histone PTMs affect their ability to penetrate cells or affect gene expression of the cells they enter due to their PTM status?
- If extracellular histones get incorporated into nucleosomes, are they localized in specific areas of the genome?
- Is there a defense mechanism against the uptake of extracellular histones to avoid the potentially damaging effects of random incorporation of extracellular histones?
- Why do histones possess such a robust ability to penetrate cells?
- Could exchange of histones between cells, with PTMs, be a completely uncharacterized communication pathway, or mechanism of disease spread through the sharing of epigenetic states via histone PTMs?

There are a few hints in previous literature that let us partially address some of these questions. First, fluorescently tagged histones (such as GFP tagged histones) have been

visualized in the nucleus just minutes after addition to the media, showing rapid uptake and at least close proximity to chromatin after internalization[161]. But perhaps more interestingly, GFP-tagged histones in a diseased cell could be seen migrating into chromatin of co-cultured healthy cells after the diseased cells died[162]. This suggests that, at least in the confines of mammalian cell culture, histones can be exchanged between diseased and healthy cells. How prevalent or possible this is in a more complex organism has yet to be studied.

Even less has been studied about how PTMs affect the ability of histones to penetrate cells or their effects on cells once incorporated, perhaps due to the lack of chemical methods to create homogenously modified histones of interest when these investigations were carried out. However, a recent preprint paper hints that histone H3 in particular is secreted by cells with specific PTMs, and later incorporated into nuclei of surrounding cells[163]. This hints that the PTMs of the extracellular histones may fulfill some function once incorporated, but further study will need to elucidate what that function is.

Perhaps the most studied aspect of extracellular histones is their toxicity. Following cell death from damage or disease, both free histones and nucleosomes can be released into the circulatory system[140-142,164]. While this area is generally understudied and the differentiation between free histones, nucleosomes, free DNA, and neutrophil extracellular traps[165] in the blood is lacking in older work, there is a clear distinction between the toxic effects of free histones and nucleosomes[141]. Cell free nucleosomes, present enough in the blood to be often used as diagnostic disease biomarkers[166-168], are known to be far less toxic than free histones, both in mammalian cell tissue culture and in injected organisms[141,169]. In fact, while no toxicity could be observed for rats injected

with 1 mg of nucleosomes[170], a similar amount of free histones caused death in just minutes[171]. Whether this marked difference in toxicity between free histones and nucleosomes in the blood is due to difference in ability to penetrate cells and incorporate themselves into chromatin has yet to be investigated conclusively. Of note, one paper mentions that nucleosomes can penetrate cells as well, but it takes far longer than free histones (around 1 day) and the nucleosomes are localized in the cytosol, likely due to endocytosis, and do not enter the nucleus[172].

Methods to characterize and alleviate the toxic effects of free histones in the blood have been investigated. Interestingly, the toxic effects of free histones could be drastically reduced in injected animals via the addition of either α-histone antibodies[173] or activated protein C[171], which targets and proteolyzes histones. The addition of heparin, a negatively charged glycosaminoglycan and common anticoagulant, also reduces histone toxicity[174]. It is possible that the binding of the α-histone antibody and heparin to the free histones either restricts their ability to penetrate cells, inhibits their detection, or both. Toll-like receptors TLR2 and TLR4 have both been suggested to recognize the free histones in circulation and initiate an immune response[175-177], but other studies which pre-blocked those TLRs with specific antibodies found that the toxic effects of free histones remained[178]. In summary, the toxic effect of free histones is robust, different than nucleosomes, but lacking in characterization.

As noted earlier, there has been no obvious connection in the literature between the toxic nature of free histones in the blood, and the ability of extracellular histones to penetrate cells and potentially incorporate themselves into chromatin. As of writing this thesis, I could find no mention of the ability of histones to penetrate cells in any literature on

extracellular histone toxicity, and the bulk of literature on “histonefection” commonly refers to histones as non-toxic, citing their low toxicity as an important feature for supporting their use as delivery agents for gene therapy[144]. Overall, this stark disconnection between fields comes down to the historical lack of tools to characterize free histones in the blood, track where and how they penetrate cells and incorporate themselves into chromatin and test the effects of different PTMs[141]. Investigating the role of free histone uptake and its connection to toxicity is a rich unexplored area that is now possible to characterize using recent technologies in sequencing, imaging, and chemical protein modification.

Goals and scope of this Thesis

Goals and scope of Chapter 1

Chapter 1 seeks to explore, scope, optimize and characterize a new method for the chemical modification of proteins, with a particular focus on installing useful sidechains for studying histone PTMs. The objective will always be to develop protein modification methodology that either provides a practical improvement to a previous method or allows for the installation of novel sidechains into proteins.

Chapter 1 begins by continuing previous work within the group on developing a light-driven approach to generating carbon-centered radical sidechain precursors in water for Dha modification in proteins. After reagents and conditions were optimized, the scope of the reaction was tested on both proteins and sidechain precursors, with a focus on finding novel chemical functionality that could be incorporated into proteins.

Further work explored novel ‘on-protein’ chemistry previously made available through the installation of electrophile containing sidechains such as alkyl halides. A range of ‘off-protein’ nucleophiles were tested to screen alkyl halide reactivity, allowing for the installation of new functional groups as well as various histone PTMs.

Goals and scope of Chapter 2

Chapter 2 seeks to apply novel chemical tools, developed either in **Chapter 1** or by colleagues within the group, to the study of histone PTMs and epigenetic processes. A major focus is the use of novel chemical tools to either create difficult to access histone PTMs, or to allow for the study of histone PTMs in new ways. In particular, protein sidechains containing elements and functional groups that are rare in biology, such as fluorine, boron, and alkyl halides, are explored for their use as novel sites of 'on-protein' reactivity or background and noise-free reporters of histone PTM function.

First, alkyl halide containing sidechains are investigated for their ability to mimic the connectivity of relevant histone PTM sidechains such as methylated lysines, allowing for binding of relevant histone PTM readers followed by proximity-driven covalent crosslinking between the alkyl halide and nucleophilic residues in the binding pocket. The crosslinking between an alkyl halide containing histone and the histone demethylase KDM4A is explored and characterized by gel, western blot, LC-MS/MS, and zinc ejection assays. Finally, alkyl halide containing histones are explored for their use in capturing unknown interaction partners after incubation in HeLa cell nuclear lysate.

Next, a novel method developed in the group is utilized to install Boronoalanine (Bal) into histone sidechains. Bal containing histones are reconstituted into nucleosomes and phosphorylated by Aurora B Kinase. Crucially, the installed Bal sidechain transiently interacts throughout the disordered histone tail via its Lewis acidity, in a process that can be tracked via NMR. This allows for NMR tracking of the structural changes undergone by the histone tail in the nucleosome after Aurora B Kinase phosphorylation, providing novel functional insights into an important epigenetic event.

Then, various natural and unnatural histone PTMs are installed to study either their interactions with DNA in the nucleosome, or the substrate tolerance of epigenetic enzymes. Various lysine methylation states are installed on histone H3-Lys36, followed by nucleosome reconstitution, to study how the extent of methylation impacts sidechain interaction with nucleosomal DNA, affecting its solvent exposure and ability to bind to chromodomain reader MRG15. In addition, various acyllysine derivatives are installed at histone H3-Lys18, a recognition site of the histone deacetylase Sirt2. A recently discovered unnatural histone PTM, lysine benzoylation, is shown to be removed by Sirt2, although at a slower rate than the native acetyllysine PTM.

Next, γCF_2-containing sidechains are investigated for their ability to track essential protein processes via ^{19}F-NMR. Fluorinated histone PTMs, such as γCF_2-AcLys, are shown to be processed by the Sirt2 deacetylase to γCF_2-Lys, with ^{19}F-NMR easily tracking the change in PTM status. Furthermore, ^{19}F-NMR measurements on a collection of relevant γCF_2- containing histone PTMs such as oxidized methionines are shown to be sensitive enough to resolve the different PTM states. Then, histone octamer assembly was shown to be tracked by ^{19}F-NMR at quantities below 1 mg.

Goals and scope of Chapter 3

Chapter 3 seeks to investigate and characterize the understudied ability of extracellular histones to penetrate cells and incorporate into chromatin. Using chemistry developed in **Chapters 1** and **2**, and an epitope tagged “trackable” human histone H3, extracellular histone incorporation will be studied extensively in a model system, using WT HeLa cells. Histone localization, half-life, and effects of PTMs will be screened on HeLa cells to characterize what happens when extracellular histones enter the cells.

The effect on transcription will be extensively investigated by following extracellular histone incorporation with RNA-seq, both with native histone H3.1 and a plethora of PTM containing variants. This will seek to determine the global transcriptional response of cells to the uptake of free histones, and what effect, if any, PTM status has on gene expression. In addition, a robust procedure for crosslinking, extracting chromatin, and shearing it via sonication to obtain soluble nucleosomes will be used to determine if extracellular histones are indeed incorporated into nucleosomes. The soluble chromatin will be used for ChIP-seq in future work to determine where in the genome the incorporated histones are preferentially located.

A novel chemical method for introducing the radioisotope ^{18}F into histone sidechains will be used to make native histone H3 trackable by positron emission tomography (PET) imaging. This will be used to follow the fate of histones after tail injection into a rodent animal model, to see if histones are able to penetrate cells and enter chromatin before they are proteolyzed and cleared from the blood. Future work will contrast these findings with full nucleosomes, providing a valuable tool for extracellular histone and nucleosome studies. Free native histones will also be injected into rat brains to simulate extracellular

histone release after injury, followed by both imaging against the epitope tagged histone and chromatin extraction to determine where the injected histone localizes, how its fate is different than that of free histones injected in the blood, and if the histones are present in soluble chromatin after extraction and shearing. Future work will perform ChIP-seq on the soluble chromatin fractions to determine where in the genome the incorporated histone localizes preferentially and how it differentiates from the ChIP-seq results from cultured HeLa cells.

Lastly, zebrafish embryos and larvae will be investigated as a model organism for establishing a platform for light-driven protein modification chemistry (developed and applied in **Chapters 1** and **2**). By virtue of their translucence and wide acceptance as an excellent model for studying cell differentiation and development, zebrafish are an ideal model organism for studying histone PTMs and optimizing light-driven protein modification methods *in vivo*. In this chapter, various light-driven chemical modification methods will be screened on microinjected Dha containing histones in zebrafish embryos and larvae to both determine their biocompatibility, and ability to modify histones *in vivo* with relevant PTMs.

Chapter 1: Development of Novel Chemical Methods for Introducing PTMs into Histones

"The logical order of how you write a paper does not always follow the chronological order that you performed the experiments"

- Professor Benjamin Guy Davis

Relevant publications (* indicates equal contribution):

Josephson, B.*, Fehl, C.*, Isenegger, P.G.* *et al.* Light-driven post-translational installation of reactive protein side chains. *Nature* **585,** 530–537 (2020).

Introduction

Recent strategies in chemical protein modification

New chemical protein modification strategies are constantly being developed, always adapting to the needs in the field for the creation of novel protein PTMs, reactive or useful chemical tags, or ideal reaction conditions. Of particular note, and importance to this thesis, is the recent focus on two often related protein modification strategies: radical based protein modification, and light-driven protein modification.

Radical protein modification

Many traditional protein modification techniques rely on heterolytic ($2e^-$) reactions between nucleophilic protein residues and 'off-protein' electrophiles[118]. These methods are often limited in their chemo- and regioselectivity, as there are a plethora of possible nucleophilic residues (e.g. Cys, Lys, N-terminus, Ser)[118,179]. Radical ($1e^-$) reactions offer an attractive alternative (**Figure 9A**), as both water and most function groups present in proteins are unreactive towards many radical species[180]. Radical protein modification often occurs via one of two pathways: 1) generation of 'on-protein' radicals via the homolytic bond cleavage of specific functional groups in native or synthetic protein residues, followed by reaction with 'off-protein' radical acceptors (**Figure 9B**), or 2) generation of 'off-protein' radicals followed by reactions with specific native or synthetic 'on-protein' radical acceptors (**Figure 9C**).

1) 'On-protein' radical formation can utilize several specific residues within proteins, allowing for residue selective protein modification after radical addition to an 'off-protein' radical acceptor (**Figure 9B**). Benzophenone has been used extensively as a photoactivatable aryl ketone which preferentially reacts with C-H bonds in aliphatic amino

acids to initiate an ‘on-protein’ radical which then adds to the ‘off-protein’ benzophenone[181]. Additionally, the benzophenone functional group can be installed into proteins via GCE to facilitate radical protein-protein crosslinking[182]. Cys can be targeted as well, either through initiation to a thiyl radical[183], or desulfurization revealing an alanyl radical[184], which can both react with ‘off-protein’ radical acceptors. In the presence of hydrogen peroxide and horse radish peroxidase, radicals can form on Tyr residues allowing for crosslinking to other Tyr residues or ‘off-protein’ Tyr analogues[185]. An α-thio radical on Met can be prepared after single electron transfer (SET) between Met and the highly oxidizing excited state of a photoactivated lumiflavin photocatalyst, which can then react with ‘off-protein’ radiacal acceptors[186]. Radical acceptors can also react with radicals formed at the beta position of Trp residues[187], and with radicals formed after C-terminal decarboxylation, although side-reactivity with Asp, Glu and Trp residues was observed at more neutral pH[124].

In addition, ‘on-protein’ radical formation can be assisted by first installing non-canonical residues into proteins that are superior radical precursors when compared to native protein residues. This has been extensively used for the formation of protein-polymer constructs by first installing an atom transfer radical polymerization (ATRP) initiator onto one of many possible native protein residues, followed by radical initiation and polymerization[188,189]. The resulting protein-polymer products are valuable as pharmaceuticals due to their improved solubility, stability, immunogenicity, and half-life in the blood[188,189]. Another recent approach first installed diazo-containing motif during a Met specific bioconjugation reaction, which could then be further activated to a radical ylide, allowing for the capture of several ‘off-protein’ benzyl radical derivatives[123].

2) Another strategy instead initiates ‘off-protein’ radicals which then react with ‘on-protein’ radical acceptors (**Figure 9C**). This strategy has been far less utilized, as there are few native protein residues which can serve as efficient radical acceptors, limited only to the aromatic amino acids. The most common targets are the electron rich aromatic residues Trp and Tyr, which are susceptible to reacting with ‘off-protein’ generated radical species of low nucleophilicity[190]. These electron poor ‘off-protein’ radicals are often alkyl radicals with several fluorine substituents which provide the necessary electron withdrawing character[120,191-194]. By comparison, there are fewer examples of ‘off-protein’ radicals targeting the comparatively electron poor aromatic protein residue His, with only one example working on simple peptides, with harsh conditions of low pH and high temperature needed to ensure His selectivity and reactivity[195]. Phe, with no activating substituents, is even less favorable, with only one example where a small peptide containing no competing aromatic residues was fluoroalkylated by radicals generated from an ‘off-protein’ hypervalent iodine salt[194].

Alternatively, similar to the previous section, synthetic amino acids with superior radical accepting ability can be installed prior to ‘off-protein’ radical generation, providing an exquisitely site-selective target for radical protein modification. Dha is an excellent SOMOphile, with high reactivity towards both electrophilic and nucleophilic radicals due to the proximity of the double bond to both electron donating and withdrawing groups in the peptide backbone[127]. Additionally, the radical formed on the α-carbon after radical addition to Dha is stabilized by the captodative effect of those same groups[196]. Crucially, the βC-γC bond formed after ‘off-protein’ radical addition to Dha creates the authentic bond found in many native protein residues and PTMs[134-136].

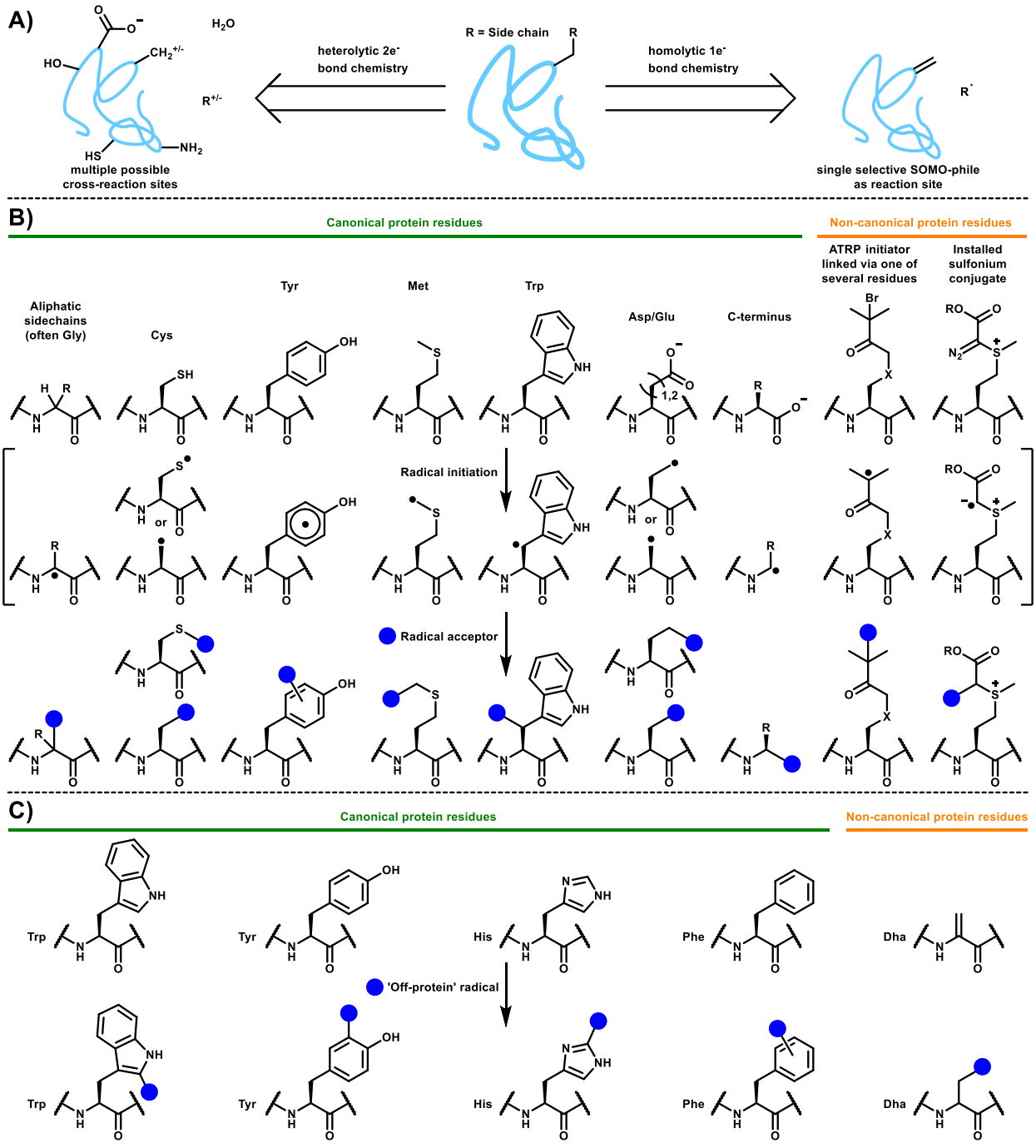


Figure 9: A) Retrosynthetic analysis of chemical protein modification using either heterolytic (left) or homolytic (right) modification chemistry. **B, C)** Examples of 'on-protein' **(B)**, and 'off-protein' **(C)** radical formation followed by protein modification on several canonical and noncanonical residues.

Light-driven protein modification

There are many advantages to using light to initiate the desired chemistry used for protein modification. In addition to the unrivaled spatiotemporal control afforded by light, visible light in particular possesses excellent biocompatibility[197,198]. Not only is visible light often safe to use on cells or animals, but most biogenic compounds do not interact with visible light at room temperature, making visible light-driven chemical approaches very selective[197]. While some organic functional groups can be activated by light directly, this often necessitates the use of harsher, less biocompatible light sources such as UV light (365 nm) to overcome the necessary activation barrier of the target substrate (**Figure 10A**)[197]. Nevertheless, this direct approach has been used for a variety of *in vitro* protein bioconjugation reactions[183], crosslinking experiments[199,200], and *in vivo* decaging of protected amino acids installed via GCE into proteins[201].

Another approach that has gained popularity in recent decades makes use of an organic or transition-metal-based photocatalyst to convert visible light to the electrochemical potential needed to activate target substrates (**Figure 10B**)[202]. While reductively or oxidatively benign in the ground state, these photocatalysts, once excited by light absorption, exist as strong and relatively long-lived single electron transfer (SET) reagents[202]. This ability, paired with the advantages of light discussed above, have made them a popular choice for driving a wide variety of radical organic reactions[202]. In fact, many of the radical protein modification strategies discussed above make use of photocatalysts to initiate the desired protein modification chemistry[120,123,186,187]. Previous work using visible light-driven peptide and protein chemistry successfully used various photocatalysts to promote amino acid transformation such as Cys to Ala desulfurization[203]

and drive bioconjugation reactions, sometimes with spatio-control via ligand-directed placement of the photocatalyst[204]. At the beginning of this thesis work, there was one example in which a small molecule Dha derivative was photocatalytically modified with heteroarenes[205], suggesting the possibility of similar Dha modification in proteins. We reasoned that proper choice of light, photocatalyst, and 'off-protein' radical precursors would provide an ideal platform for modifying Dha residues in proteins.

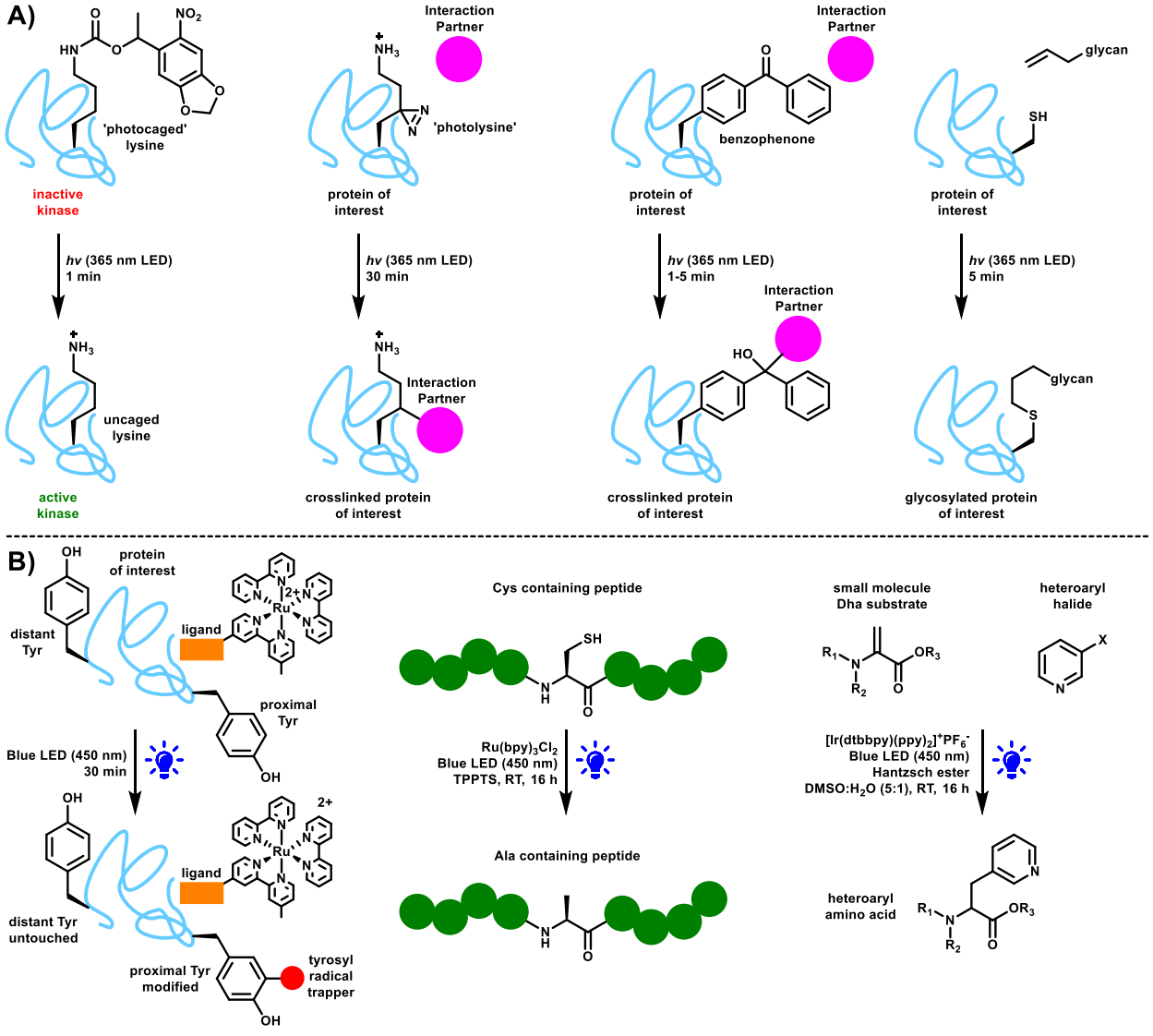


Figure 10: Protein modification examples using **A)** direct activation of target substrates with light, and **B)** visible light paired with photocatalysts to activate desired substrates.

Optimizing a light-driven radical approach to dehydroalanine based PTM installation

Previous work within the group

Given the benefits of light-driven chemistry and the utility of photocatalysts for initiating radical reactions (discussed above), we sought to improve upon our previous work on radical protein modification via Dha residues[135,136]. Our group's original strategy for sidechain radical generation relied on alkyl halides and sodium borohydride, each with their own limitations in terms of solubility and potential side reactions via off-target protein alkylation (alkyl halides), and lack of chemical functional group tolerance (redox sensitive groups to sodium borohydride)[135]. Furthermore, any reaction utilizing sodium borohydride would not be biocompatible due to the rapid generation of hydrogen gas when mixed with water, an attribute which also made radical initiation difficult when mixed with water resulting in intense bubbling, which also caused difficulties with repeatability[135]. Also, while the original radical generation strategy allowed for the installation of many native and PTM containing amino acid sidechains, the harsh reductive potential of sodium borohydride (E_{red} = -1.24 V)[206] made installing many redox sensitive functional groups incompatible. Protein sidechains containing reactive, unnatural groups such as azides, alkyl halides, and unsaturated alkyl groups prone to reduction are useful tags for further 'on-protein' chemistry[179], and we reasoned that a milder and more controlled reductive or oxidative initiation afforded by the pairing of light, photocatalyst and suitable 'off-protein' radical precursor would allow for the installation of more sidechain functionality into Dha containing proteins.

Early work (performed by Professor Charlie Fehl and Dr. Tom Wright) explored multiple avenues for photocatalytic 'off-protein' radical generation, including both reductive and oxidative reaction manifolds. A model protein, mutant *Xenopus laevis* histone H3-Cys9 was expressed and converted to H3-Dha9 as previously described[135] for use as the protein substrate for the photoredox reactions. Some success was found pairing various alkyl iodide 'off-protein' radical sidechain precursors with the photocatalyst Ir[*d*F(CF_3)ppy]$_2$(bpy)PF_6 (*E_{red} = -1.33 V)[207] and the sacrificial reductant *N,N*-Diisopropylethylamine (DIPEA) upon illumination with a 9 W Blue LED. However, the reactions were plagued with oxidative damage (+16 Da oxygen adducts likely due to Met oxidation, judged by intact protein LC-MS, even occurred in degassed conditions), poor ion series, low conversions, limited photocatalyst solubility, and long reaction times (4 h)[208]. Photocatalytic reductive initiation of these alkyl iodide sidechain precursors was thus abandoned, and oxidative initiation options were investigated instead.

Recent literature had identified alkyltrifluoroborates as promising reagents for oxidative photocatalytic radical initiation[209]. Their intrinsic charge was encouraging in terms of their solubility, but their drastically differing rates of hydrolysis were initially concerning[210]. However, when paired with the strongly oxidizing photocatalyst Ir[*d*F(CF_3)ppy]$_2$(bpy)PF_6 (*E_{ox} = +1.32 V)[207], two early alkyltrifluoroborate candidates, isopropyltrifluoroborate and benzyltrifluoroborate, were successfully installed onto histone H3-Dha9, creating the sidechains Leu and hPhe, respectively (**Figure 11**). Intriguingly, while isopropyltrifluoroborate gave nearly full conversion to the desired single addition product, benzyltrifluoroborate mostly gave the double addition product, indicating that the α-carbon radical generated after the initial radical addition to Dha was quenched by an undesired

second benzyl radical. It was reasoned that this was due to a more rapid initiation of benzyl- vs isopropyltrifluoroborate due to its lower E_{ox} value (E_{ox} = +1.10 vs +1.50 V, respectively)[211]. Therefore, various additives and alternative photocatalysts were tested to mediate the reaction and prevent overalkylation.

As expected, milder photocatalysts $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ either failed or poorly initiated the benzyltrifluoroborate radical precursor, due to insufficient oxidation potentials ($^*E_{ox}$ = +0.99 V and +0.77 V, respectively)[207]. Pleasingly, $NaBH_3CN$, when added with the initial photocatalyst $Ir[dF(CF_3)ppy]_2(bpy)PF_6$, successfully suppressed double addition of the benzyl radical by quenching the α-carbon radical (**Figure 11**). But most interestingly, catechol, initially tested as a hydrogen atom transfer (HAT) reagent[212], not only increased double addition of benzyltrifluoroborate when added with $Ir[dF(CF_3)ppy]_2(bpy)PF_6$, but allowed for radical initiation with the milder photocatalysts $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ (**Figure 11**). This was especially advantageous as those milder photocatalysts are cheaper, more water soluble, and drastically reduce unwanted oxidative protein damage due to their lower potentials (see below).

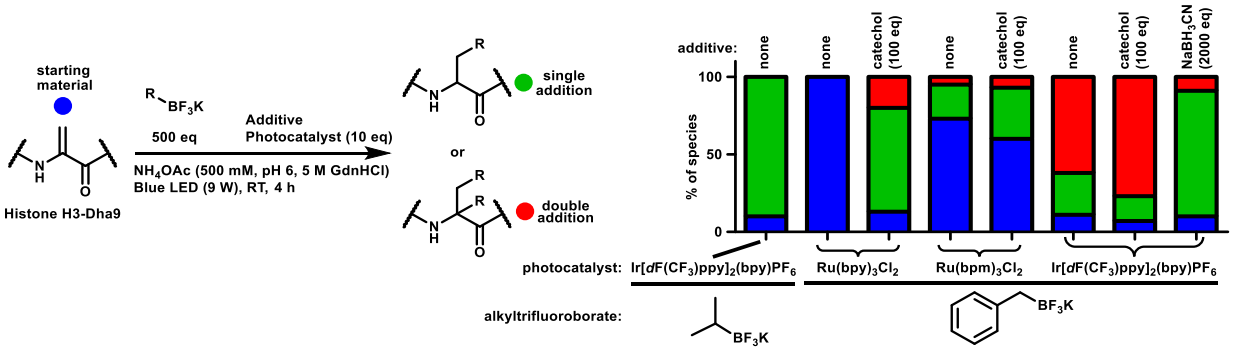


Figure 11: Initial screens of alkyltrifluoroborate radical initiation with photocatalysts of varying strengths and additives, with the goal of producing selective single addition products.

Several hypotheses were tested to elucidate the role of catechol in effectively lowering the necessary photocatalyst oxidative potential (**Figure 12**). First, its role as a potential redox mediator[213] was ruled out by screening various catechol and quinone derivatives. Only 1,2-vicinal diols catechol and oxidized variant 1,2-quinone *o*-chloranil promoted oxidation of model substrate phenethyltrifluoroborate with the mildest photocatalyst $Ru(bpy)_3Cl_2$ (**Figure 12A**). Second, modification of the photocatalyst $Ru(bpy)_3Cl_2$ by catechol was ruled out when no reactivity was observed when using a pre-synthesized catecholo-$Ru(bpy)_2$ photocatalyst[214], even when catechol was spiked into the reaction (**Figure 12B**). Third, the possible *in situ* formation of alkylboronic acid catechol esters, post trifluoroborate hydrolysis, was tested by using two preformed esters, which both reacted successfully without the addition of catechol (**Figure 12C**).

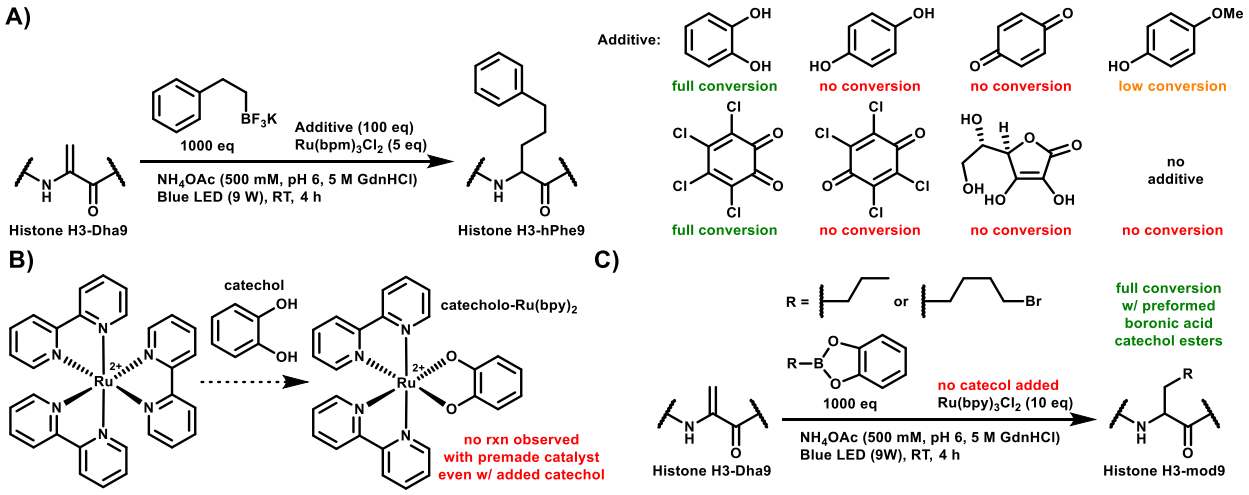


Figure 12: Testing the role of catechol in promoting alkyltrifluoroborate oxidation with **A)** catechol derivatives **B)** catalyst modification and **C)** boronic acid catechol ester formation.

To further probe the role of catechol we performed cyclic voltammetry experiments (done by Professor Charlie Fehl in collaboration with Dr. Lifu Chen and Dr. Christopher Batchelor-McAuley of the Professor Richard Compton group) on phenethylboronic acid with and without catechol in aqueous media[215]. Pleasingly, in agreement with the experiments described above, adding catechol resulted in a drastically reduced E_{ox} value (E_{ox} = +0.49 V vs +1.47 V) (**Figure 13**). We reasoned that this evidence, along with the experiments listed above, were sufficient evidence that the addition of catechol results in boronic acid catechol ester formation with *in situ* hydrolyzed alkyltrifluoroborates[216], lowering the oxidation potentials of 'off-protein' radical sidechain precursors to within the activation range of milder photocatalysts such as $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$.

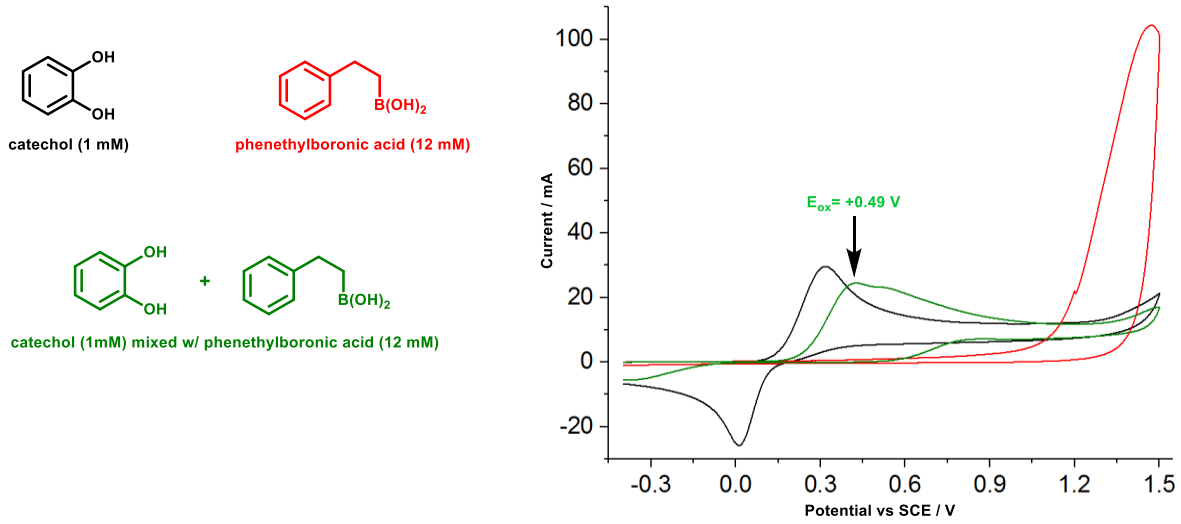


Figure 13: Independent and mixed voltammetric responses of phenethylboronic acid (12 mM) and catechol (1 mM) on glassy carbon (GC) in PBS (pH 7.10) show clear E_{ox} reduction in the presence of catechol.

With these early results, and the beneficial qualities of catechol, the development of a useful light-driven radical Dha modification reaction platform was looking promising. However, many issues required addressing: the relationship between photocatalyst strength and alkylboronate ‘off-protein’ radical sidechain precursor, interplay between photocatalyst strength and oxidative protein damage, full optimization and characterization of the reaction and modified protein reaction products, flexibility of catechol derivative, tolerance of the reaction manifold to redox sensitive functional groups, investigation of novel ‘on-protein’ chemistry afforded by installation of reactive functional group containing sidechains, protein compatibility and scope, and the implementation of novel bio-applications with the newly created protein modifications. These investigations will be the concern of the majority of this thesis chapter, and were part of a major publication from my DPhil studies[208].

Matching photocatalyst strength with sidechain precursor

As discussed above, initial work had identified a promising set of ‘off-protein’ radical precursors in alkylboronates, which once activated, released alkyl radicals which could selectively react with Dha residues installed in proteins. These compounds, when paired with catechol, exhibited a far lower activation potential, allowing them to be oxidized by milder photocatalysts such as $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ in the presence of a blue LED light source. However, further work needed to be done to define the relationship between choice of alkylboronate radical precursor and photocatalyst and its effect on the efficiency and conversion of the reaction, as well as the relationship between catalyst strength and protein oxidative damage. Previous work (**Figure 11**) clearly highlighted that different alkylboronates (benzyl vs isopropyl alkyl substituent) exhibited drastically different reactivity with different strength photocatalysts.

To begin, the relationship between photocatalyst strength and oxidative protein damage was investigated to find the upper limit (in terms of $*E_{ox}$ value) that could be tolerated in the reaction conditions while restricting oxidative protein damage and controlling reactivity by selectively providing single over double addition products. As a model system (used here and throughout the rest of the reaction development and optimization section), the histone H3 protein from *Xenopus laevis* was modified with Dha at position 9 (histone H3-Dha9) to be used as the protein substrate for ‘off-protein’ radical addition[135]. For the model ‘off-protein’ radical precursor, the commercially available potassium phenethyltrifluoroborate was chosen. Five commonly used photocatalysts exhibiting increasing $*E_{ox}$ values were screened for their ability to promote desired reactivity and/or prevent protein oxidative damage: $Ru(bpy)_3Cl_2$ ($*E_{ox}$ = +0.77 V), $Ru(bpm)_3Cl_2$ ($*E_{ox}$ =

+0.99 V), Ir[*d*F(CF$_3$)ppy]$_2$(bpy)PF$_6$ (*E$_{ox}$ = +1.32 V), Ru(bpz)$_3$PF$_6$ (*E$_{ox}$ = +1.45 V) and Acr$^+$-MES$^-$ (*E$_{ox}$ = +2.06 V)[207]. These photocatalysts were tested in four variations of reaction conditions: open to air with no substrate or catechol, open to air with substrate and catechol, degassed with substrate but no catechol, and degassed with substrate and catechol. Reactions were assessed by intact protein LC-MS, by both interpreting the quality of the ion series and the content of deconvoluted spectra, which contained easily identifiably peaks corresponding to either starting material Dha, single vs double addition products, or oxidative damage adducts (+16 Da, presumed to be Met oxidation). See **Figure 14** for full reaction details.

The most obvious trend seen in the results of the photocatalyst screen was the increasing oxidative damage and lack of reaction control when photocatalyst strength was increased. In the first condition, where the reaction was open to air and contained no phenethyltrifluoroborate substrate or catechol, there was oxidation damage in all cases demonstrated by the mass of multiple +16 Da adducts accumulating on the protein. However, catalysts with *E$_{ox}$ > +1.00 V caused protein damage significant enough to affect the ion series of the protein, making the spectra difficult to deconvolute and analyze further. Interestingly, the addition of substrate and catechol to the reaction in open air reduced oxidative damage in all cases, potentially by providing the excited photocatalysts with an abundance of targets other than the protein component in the reaction. The photocatalyst Ru(bpm)$_3$Cl$_2$ (second in strength) yielded the best result in these conditions, with minimal oxidative damage in open air and the major reaction product being the desired single addition adduct. Higher strength photocatalysts had more oxidative damage, double addition and starting material adducts (**Figure 14**).

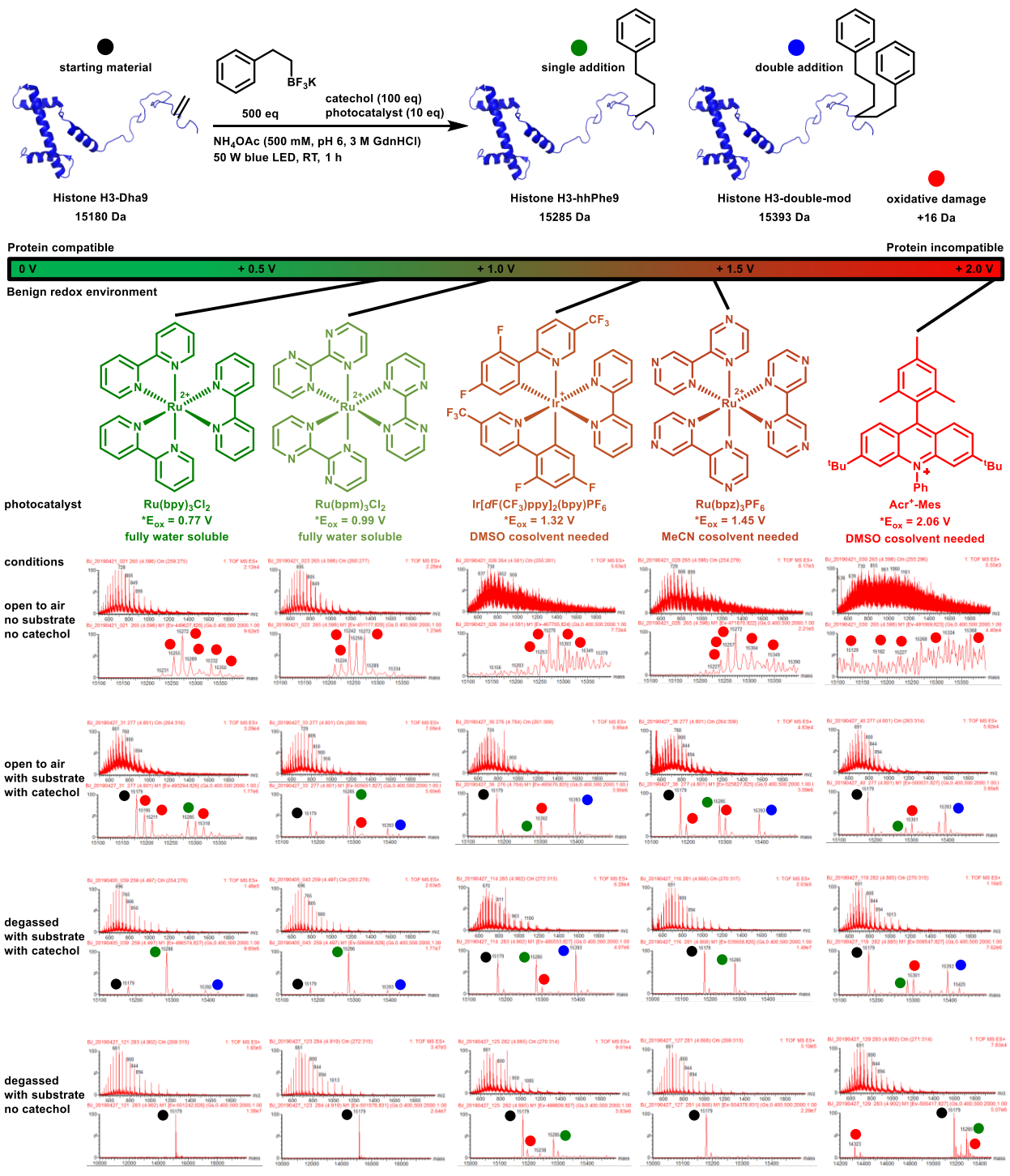


Figure 14: Photocatalyst strength screening to assess oxidative protein damage and reactivity control in multiple environments. Reaction conditions (top) were tested with a range of photocatalysts (middle) and analyzed by intact protein LC-MS (bottom).

When the reactions were degassed and carried out in a glovebox (< 6 ppm O_2) with substrate and catechol, the milder photocatalysts $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ (each with $^*E_{ox}$ < +1.00 V) performed excellently, with nearly full conversion to the desired single addition product, negligible oxidative damage, and trace amounts of undesired starting material or double addition adducts. In contrast, the higher strength photocatalysts had lower conversion, some oxidative damage even in the low oxygen environment, and significant double addition, except in the case of $Ru(bpz)_3PF_6$ which likely exhibited lower reactivity due to solubility issues. When catechol was removed, $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ were unreactive as expected as they could not directly activate the hydrolyzed phenethylboronic acid (+1.47 V, see above). While two of the higher strength photocatalysts did exhibit some reactivity ($Ru(bpz)_3PF_6$ again had low reactivity likely due to poor solubility), conversion was poor and included some oxidative damage and evidence of protein cleavage (**Figure 14**).

Based on the results of the photocatalyst screening experiment, the decision was made to focus on the two milder photocatalysts $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$, as they provided the highest conversion to the desired single addition product while also avoiding oxidative damage on the protein. Directly activating alkylboronate ‘off-protein’ radical precursors with stronger photocatalysts was ruled out as oxidative protein damage was difficult to avoid under even degassed conditions, and reactivity was hard to control, giving mixtures of single and double addition. The source of this lack of control is currently not resolved but may be due to rapid radical initiation by the stronger photocatalysts leading to overalkylation, versus a milder, more controlled release of radicals by the weaker catalysts giving preferably single addition products.

Next, varying the alkyl component of the alkylboronate 'off-protein' radical precursors was tested with the two photocatalysts of choice, $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$. It is known that varying the alkyl component of alkylboronates can drastically change their E_{ox} values (**Figure 15A**)[211,217], a trend that may be transferred to the target *in situ* formed alkylboronic acid catechol ester formed during the reaction (see previous work section above). Therefore, the subtle differences in strength between $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ ($^*E_{ox}$ = +0.77 vs +0.99 V, respectively) might have provedd essential for successfully reacting with different alkylboronate substrates. This trend was initially observed with the stark differences in reactivity of each photocatalyst with potassium benzyltrifluoroborate under the same conditions (**Figure 11**).

To test the reactivity of different 'off-protein' alkylboronate radical precursors with either $Ru(bpy)_3Cl_2$ or $Ru(bpm)_3Cl_2$, identical reactions (**Figure 15B**) were carried out with both photocatalysts and a range of alkylboronate substrates with both known and unknown oxidation potentials (**Figure 15A**). In some cases, reaction times were altered to give the best conversions to compare between photocatalysts, but in every case the times and other conditions were kept identical for each photocatalyst. Alkyl constituents consisted of primary, secondary, and benzyl radical precursors, and longer primary alkyl radical precursors contained a variety of functional groups such as bromine, azide, and alkenes. Furthermore, the substrates possessed a range of boronate types, including trifluoroborates, boronic acid pinacol esters, and boronic acid catechol esters. The reactions were again monitored by intact protein LC-MS (**Figure 15C**), and conversions to either the desired single addition or undesired double addition product were calculated and compared (**Figure 15D**).

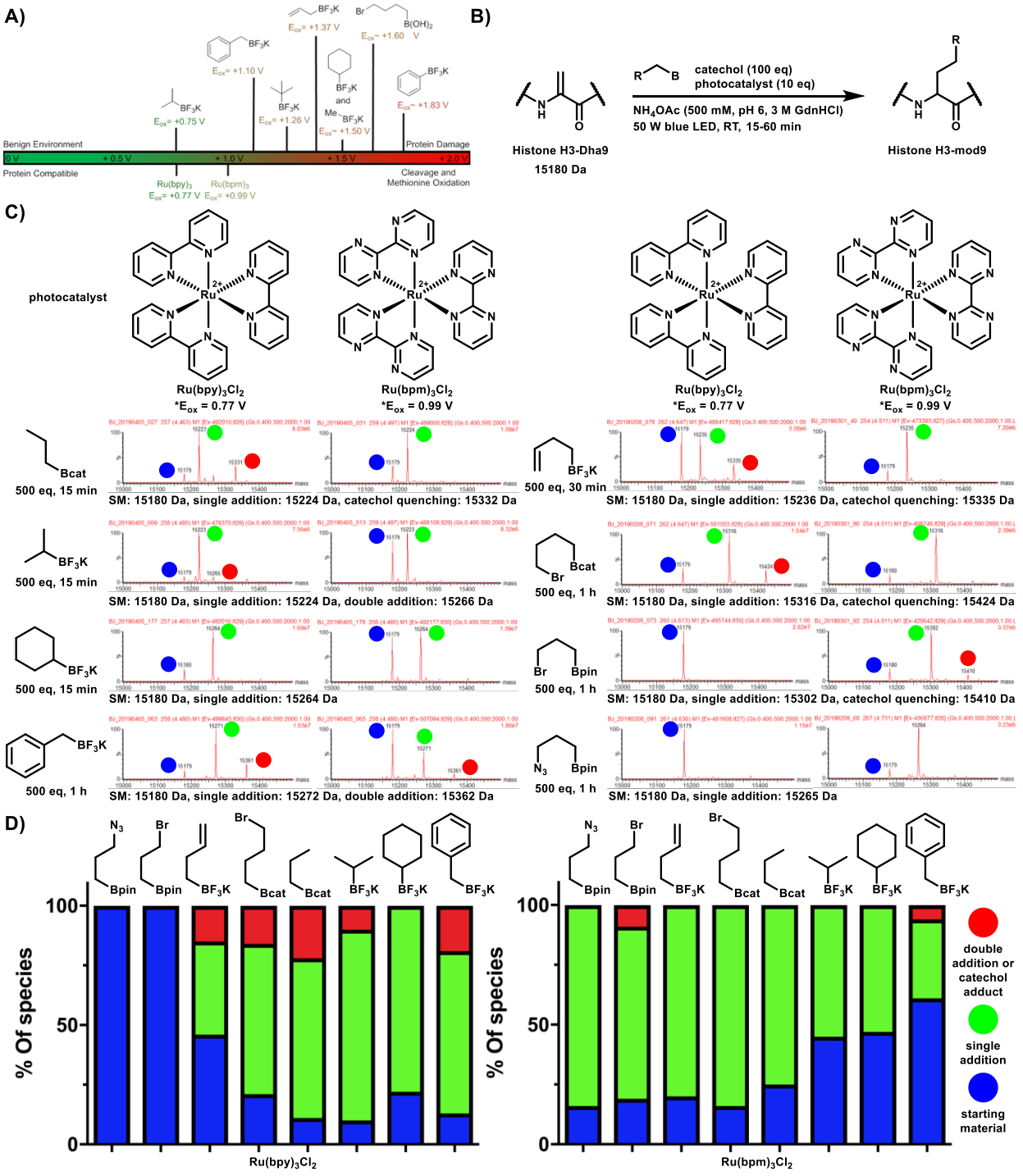


Figure 15: A) E_{ox} values of various alkylboronates. **B)** Reaction conditions for photocatalyst comparison with different alkylboronate substrates. **C)** Intact protein LC-MS deconvoluted spectra. **D)** Stacked bar chart of reaction products from **(C)**.

The results of the alkylboronate screen comparing photocatalysts $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$ ($^*E_{ox}$ = +0.77 V vs +0.99 V, respectively) was most obvious when a stacked bar chart was created for each photocatalyst (**Figure 15D**). The sequence of alkylboronates that gave increasing conversions to the desired single addition product for $Ru(bpy)_3Cl_2$ (**Figure 15D**, left) showed a decreasing trend for $Ru(bpm)_3Cl_2$ (**Figure 15D**, right). Immediately, it was clear that some alkylboronate substrates were not accessible to the weaker $Ru(bpy)_3Cl_2$ photocatalyst, as both 3-azidopropyl- and 3-bromopropylboronic acid pinacol ester gave no conversion, while giving nearly full conversion when $Ru(bpm)_3Cl_2$ was used. This suggested that the longer primary alkylboronate radical precursors were best paired with the slightly stronger $Ru(bpm)_3Cl_2$ photocatalyst, further supported by the low conversion of but-3-enyltrifluoroborate with $Ru(bpy)_3Cl_2$ in contrast to the high conversion with $Ru(bpm)_3Cl_2$.

Conversely, several alkylboronate substrates, notably secondary alkyl and benzyl radical precursors, exhibited excellent conversion with the weaker $Ru(bpy)_3Cl_2$ photocatalyst yet gave mediocre conversion with the stronger $Ru(bpm)_3Cl_2$ photocatalyst. There are at least two explanations, which possibly work in combination; for one, several secondary alkyl- and benzyltrifluoroborates possess relatively low E_{ox} potentials (**Figure 15A**)[210], and two, the resulting secondary and benzyl radicals are more stable than their primary alkyl radical counterparts. A stronger photocatalyst may therefore rapidly initiate radical formation for these substrates which are then longer lived due to their inherent stability, leading to double addition and possible self-quenching causing lower conversions. This was most evident when the strong photocatalyst $Ir[dF(CF_3)ppy]_2(bpy)PF_6$ ($^*E_{ox}$ = +1.32 V) was tested with catechol and benzyltrifluoroborate, leading to nearly full conversion to the

double addition product (**Figure 11**), a result that has yet to be reproduced with radical precursors that lead to less stable radicals. A weaker photocatalyst such as $Ru(bpy)_3Cl_2$ may negate these detrimental effects by slowly generating the desired radicals.

These results, in combination with previous work (see above), provide a platform for pairing the right photocatalyst with desired 'off-protein' sidechain radical precursor. The two mildest photocatalysts investigated, $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$, avoid unwanted oxidative protein damage, yet, when paired with catechol, can react with a range of alkylboronate substrates. The stronger photocatalyst, $Ru(bpm)_3Cl_2$, is best paired with primary alkylboronate substrates to overcome their higher oxidation potentials, while the weaker $Ru(bpy)_3Cl_2$ photocatalyst is more suitable for slowly releasing the stable radicals from easily activatable secondary and benzyl radical precursors.

Altogether, these findings allowed us to propose a mechanism for the light-driven modification of Dha residues in proteins with catechol mediated radical formation from alkylboronate sidechain precursors (**Figure 16**). First, an alkylboronate 'off-protein' sidechain radical precursor is added to aqueous media. Throughout this work, trifluoroborates, boronic acids, boronic acid pinacol esters and boronic acid catechol esters were all used successfully[208]. Once immersed in water, the alkylboronate hydrolyses to its corresponding boronic acid[210], combining with the additive catechol resulting in *in situ* formation of a boronic acid catechol ester that drastically lowers the E_{ox} value to within the range of the mild photocatalysts $Ru(bpy)_3Cl_2$ and $Ru(bpm)_3Cl_2$. Under blue light illumination, the photoactivated excited state of the photocatalyst oxidatively initiates alkyl radical formation from the boronic acid catechol ester, which then selectively reacts with the pre-installed SOMOphile Dha residue in the target protein of interest. The

subsequently generated α-carbon radical is then likely quenched by the Ru^{I} state of the photocatalyst, completing the catalytic cycle and revealing the desired sidechain after protonation from water. For the model 'off-protein' radical precursor potassium phenethyltrifluoroborate, conversion to the desired single addition product was >95% as determined by analyzing the deconvoluted spectra from intact protein LC-MS measurement of the crude reaction mixture (**Figure 16**). If catechol, substrate, Dha containing protein, photocatalyst, or light were absent, the reaction failed[208].

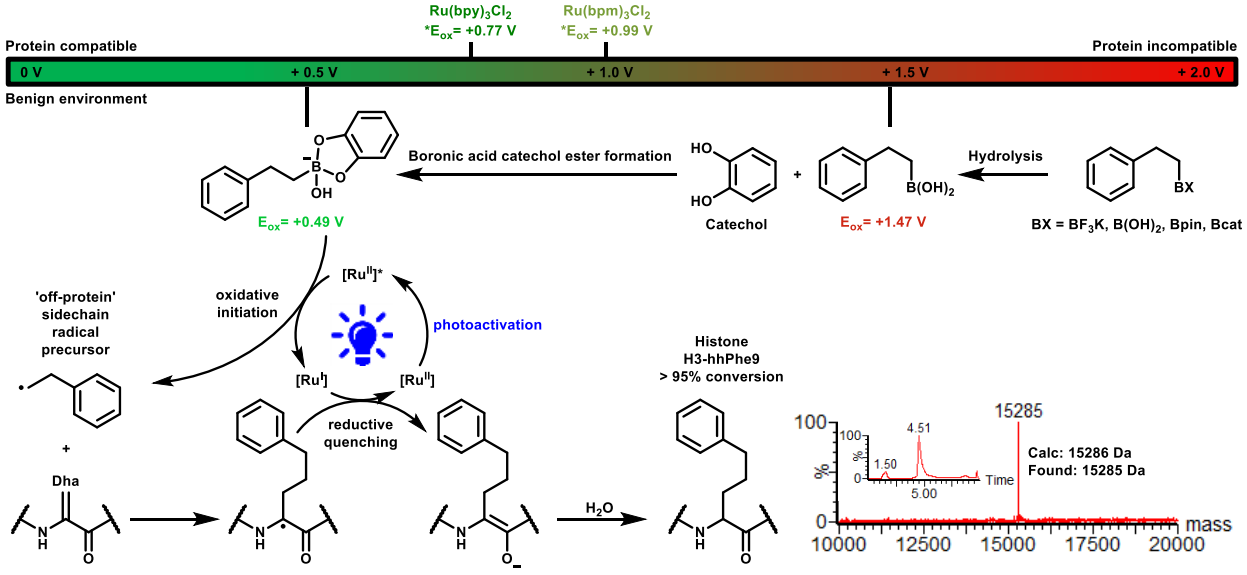


Figure 16: Proposed mechanism of light-driven modification of Dha residues in proteins with catechol mediated radical formation from alkylboronate sidechain precursors.

Testing catechol variants

While the reaction conditions developed and optimized above were sufficient for installing various sidechains with excellent conversions, we were curious if the catechol additive could be varied to include several biorelevant compounds. L-DOPA, a hydroxylated version of Tyr, and dopamine, the related neurotransmitter, both contain the catechol moiety[218]. Furthermore, so do many natural products such as gallic acid[219] and caffeic acid[220]. While currently ambitious, the use of *in situ* biogenic catechol derivatives in living systems, such as dopamine in the synapses of neurons in the brain, to drive desired protein modification locally would be of extreme interest for future exploration. Luckily, of the many derivatives tested, most provided excellent conversion with $Ru(bpm)_3Cl_2$ and 4-bromobutylboronic acid (**Figure 17**).

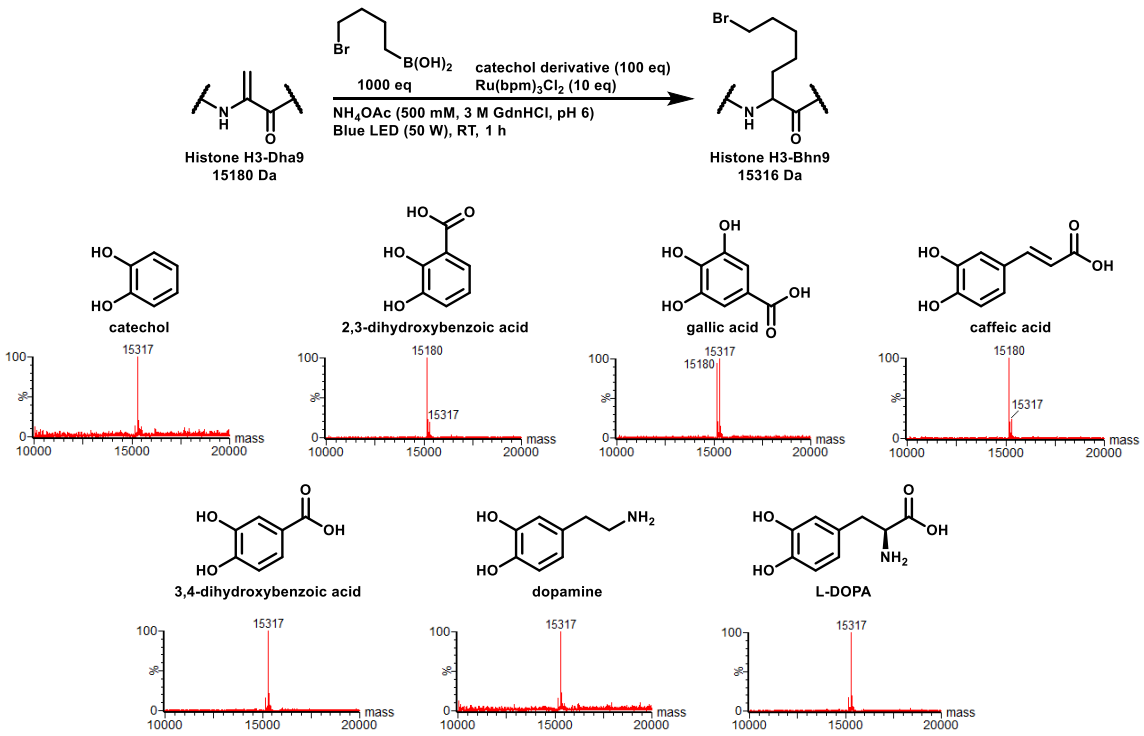


Figure 17: Testing reactivity with different biologically relevant catechol derivatives.

Protein scope, compatibility, and scalability

Next, we decided to explore the limitations of the light-driven radical Dha modification reaction developed and optimized above. While it performed well in ideal conditions, giving excellent conversions to many desired sidechains on the model protein histone H3-Dha, a useful protein modification reaction needs to be flexible enough to tolerate different proteins in a range of conditions allowing for proper protein folding and function, and at different scales as required for future applications.

Five proteins were selected to test the protein scope of the light-driven radical chemistry with the model substrate phenethyltrifluoroborate (**Figure 18A**), with expression, purification, and Dha installation carried out as previously described[133,135]. Nuclear protein histone H4-Dha16 and enzyme PanC were tested under analogous conditions to previously tested histone H3-Dha9, both giving nearly full conversion to the desired product containing the sidechain hhPhe. Periplasmic protein AcrA and beta-sheet rich NPβ were tested in buffers allowing them to retain their folded structure, with circular dichroism (CD) measurements revealing retention of their folded state consistent with their WT or Cys (mutated at the residue of interest) forms (**Figure 18B**). Interestingly, NPβ exhibited a drastic change in folding upon Dha formation, possibly due to the planar Dha structure disrupting the turn at its installed site, yet it regained proper folding after sidechain installation during the reaction.

A large scale (5 mg) reaction (**Figure 18C**) was carried out on histone H3-Dha9 with 4-bromobutylboronic acid, giving good conversion (82%) to the desired single addition product. Furthermore, in future applications (**Chapter 2**), several different histone H3 sites were successfully modified as well (sites 4, 9, 18, 27, 36 and 56).

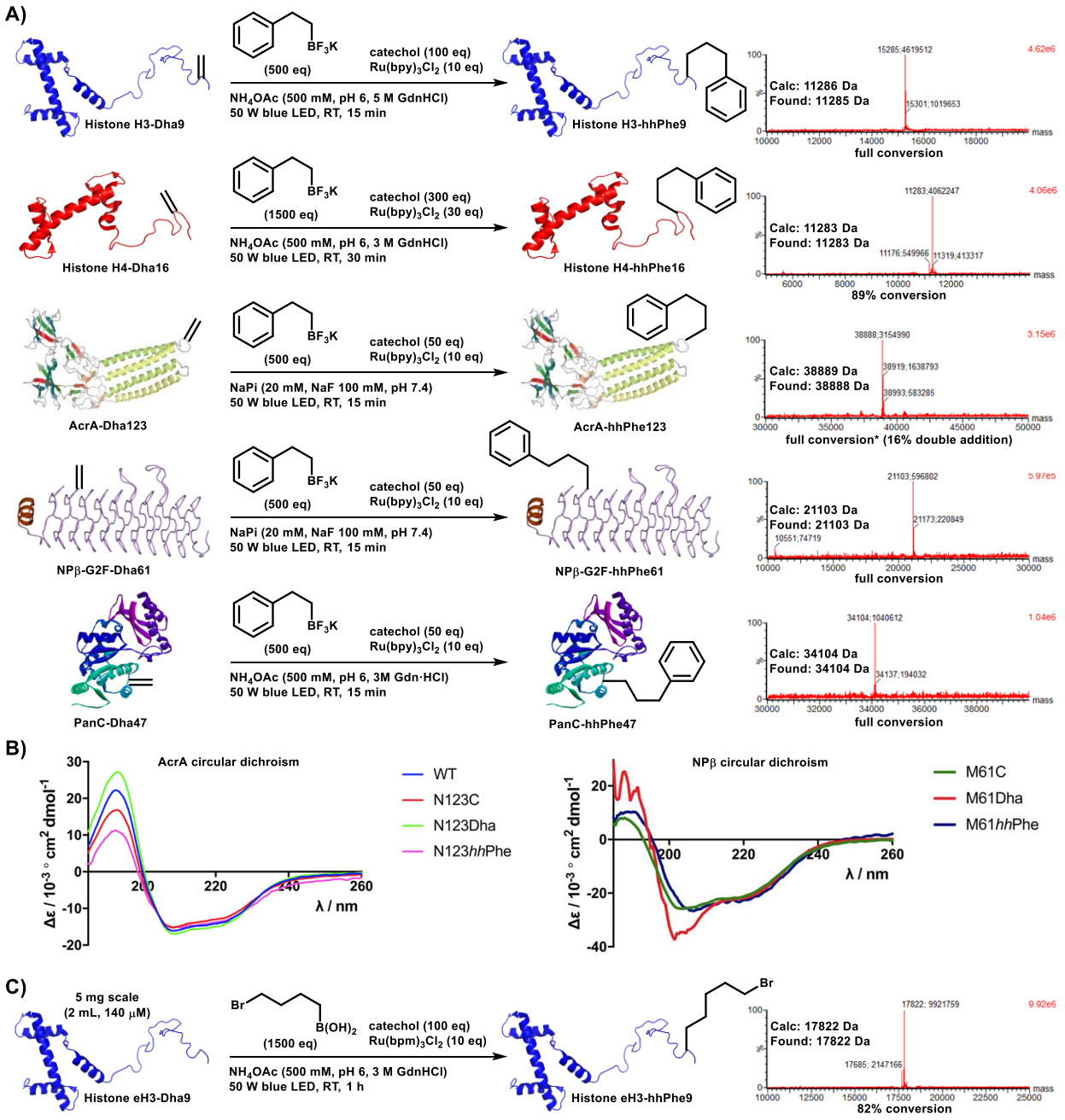


Figure 18: A) Protein scope reaction conditions and intact protein LC-MS deconvoluted spectra results. *38919 Da AcrA adduct was made during Dha formation and was disregarded. **B)** CD measurements for AcrA and NPβ. **C)** Large scale (5 mg) reaction with intact protein LC-MS deconvoluted spectra of the product.

Substrate scope

Collectively, the light-driven radical Dha modification reaction developed herein had proven flexible to catechol variation, protein and protein site choice, scale, and photocatalyst choice to match 'off-protein' radical precursor substrate. While many 'off-protein' sidechain radical precursors had been tested successfully in the initial investigations of the reaction, we had yet to fully probe the functional group tolerance. Ideally, a reaction that installs protein sidechains should be able to tolerate native and PTM containing sidechain precursors as well as various reactive chemical groups allowing for bio-orthogonal chemistry[135,179]. We investigated the functional group tolerance and substrate scope of our reaction (**Figure 19**) by following the guidelines established above, using $Ru(bpy)_3Cl_2$ for benzyl or secondary alkylboronates, and $Ru(bpm)_3Cl_2$ for primary radical precursor alkylboronates. Full reaction details for each substrate can be found in the Experimental Details section for this chapter. Alkylboronate substrates were either commercially available, or kindly provided by Dr. Patrick Isenegger, Adeline Poh, Dr. Tim Mollner, Dr. Andrew Giltrap, and Dr. Tobi Arisa.

Initial investigations showed that simple alkyl groups could be successfully added to give native and unnatural sidechain variants such as norvaline (Nva), norleucine (Nle), leucine (Leu), cyclohexyl-alanine (Cy-Ala), homo- and homohomophenylalanine (hPhe and hhPhe, respectively). Ether groups were tolerated as well, allowing for phenyl-homoserine (ph-hSer) and PMB protected homoserine (PMB-hSer) installation. Next, alkene functional groups were tested, allowing successful installation of allylhomoalanine (allyl-hAla) and allylnorvaline (allyl-Nva). Alkene functionality, rare in proteins and throughout biology, has previously been leveraged as a bio-orthogonal reactive handle for cross-

metathesis reactions on proteins[130], showing the compatibility of this method with other orthogonal protein chemistries.

In previous radical Dha modification chemistry, esters were too reductively sensitive to the reaction conditions, and could not be installed into proteins[135]. However, this mild oxidative radical initiation approach left esters untouched, allowing for methylhomoglutamic acid (Me-hGlu) and tertbutylhomoglutamic acid (^{t}Bu-hGlu) incorporation, as well as the free acid homoglutamic acid (hGlu). Azides, another reductively sensitive functional group and popular reactive tag for bio-orthogonal chemistry[179], were also tolerated, exhibited by the installation of azidonorleucine (Anl). The canonical residue lysine (Lys), and several relevant lysine PTMs such as acetyllysine (AcLys) and trimethyllysine (Me_3Lys), were easily installed, showcasing the strength of this method in studying relevant histone modifications[58,61]. In addition, the recently identified unnatural lysine PTM benzoyllysine (BzLys) was generated, allowing for direct comparison of this food additive (sodium benzoate) triggered histone PTM with site-specific natural PTM AcLys analogues (see **Chapter 2**)[76].

Alkyl halides, reactive electrophiles used previously for reductive radical initiation[136], were tested next. Pleasingly, chloro, bromo, and iodo containing sidechains were all tolerated, allowing for the installation of simple sidechains such as iodonorleucine (I-Nle), bromonorleucine (Br-Nle), and bromohomonorleucine (Br-hNle), as well as more complex ribose derivatives with chloro and bromo groups (Br-ribose-Nle and Cl-ribose-Nle), showing the flexibility of this method not only to reactive groups but also more complex structure. In future sections, the installed alkyl halides are used for further ‘on-protein’ chemistry afforded by their unique electrophilic character. In addition, a small boronic acid

containing peptide was successfully installed into proteins (peptide-hAla), showing the potential cooperativity of this method with our recently published protein borylation work[221]. Lastly, a PEG-linked biotin tag was easily installed, providing a useful handle for biotin-streptavidin protein enrichment applications[222].

In conclusion, the current method is mild enough to tolerate introducing a wide range of native, PTM containing, unnatural, and reactive group containing sidechains into proteins. In particular, the ability to install reductively sensitive electrophile containing sidechains is a major advantage over previous techniques[134,135], making this method compatible with a plethora of other bio-orthogonal chemistries relying on alkenes, azides, esters, and halogens.

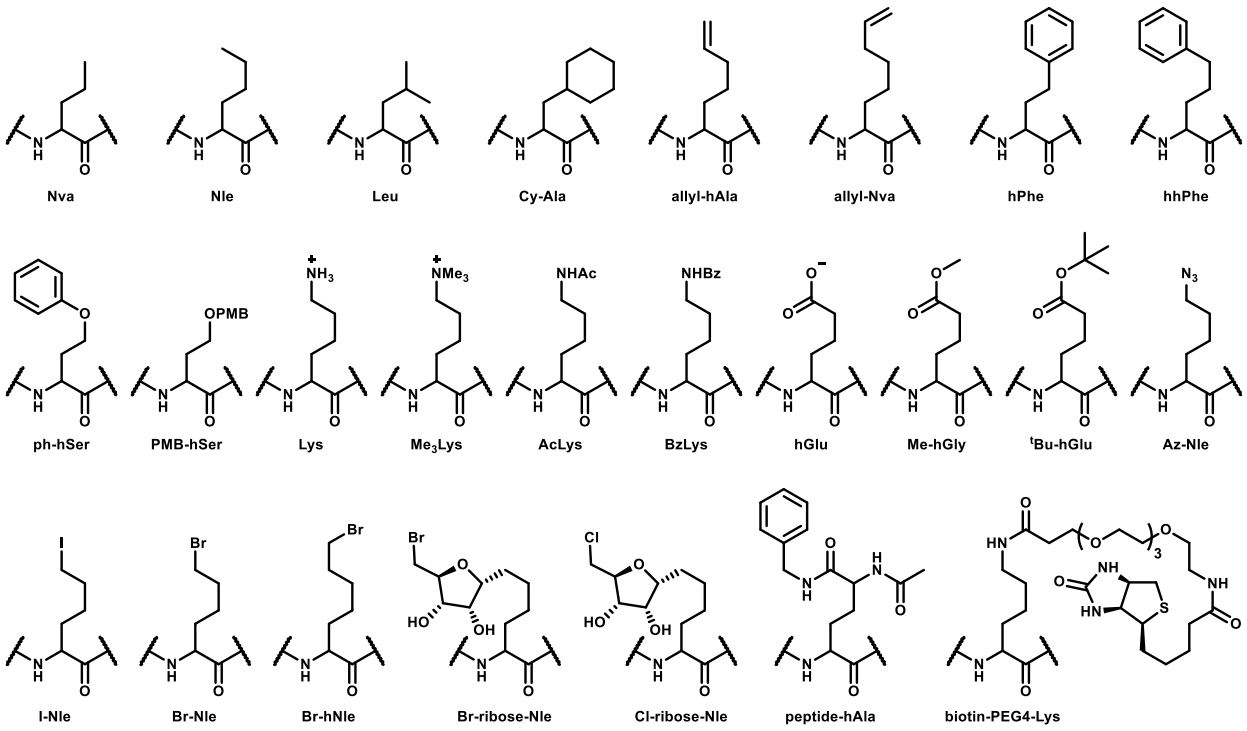


Figure 19: Sidechains installed by light-driven radical Dha modification chemistry.

Characterization of installed protein modifications by HPLC and MS

With the light-driven radical Dha modification reaction developed and optimized, and both substrate and protein scopes explored, it was worth spending time fully characterizing the reaction products. Up to this point, all the screening and scoping reactions were tracked by intact protein LC-MS on aliquots of the crude reaction products (**Figure 20**). This produced a total ion count (TIC) chromatogram, with the intact protein peak appearing around 4.60 minutes. The entirety of this peak was integrated to give the corresponding ion series, ensuring that all species (starting material, product, as well as undesired adducts) were accounted for. The ion series was then deconvoluted, allowing for conversions to be calculated from the relative intensities of the peaks corresponding to the expected mass of each species. This procedure was conducted for all substrate and protein scope products, with complete details containing the TIC chromatogram, ion series and deconvoluted spectra provided in the **Experimental Details** section of this chapter, as well as in the corresponding publication[208].

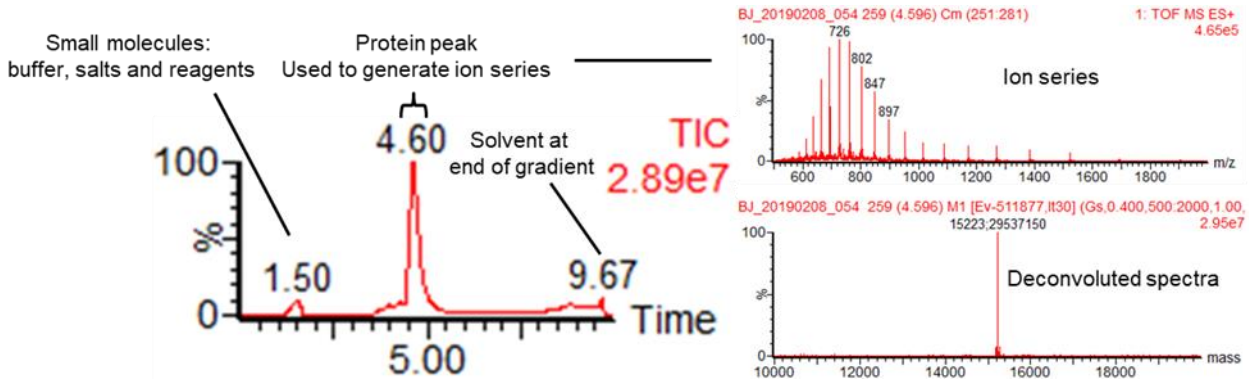


Figure 20: Representative total ion count chromatogram (TIC, left), ion series (top right) and deconvoluted spectra (bottom right) from LC-MS analysis of the crude reaction mixture from the installation of Val on histone H3-Dha9.

While this intact protein LC-MS method is only regarded as semiquantitative[21], we reasoned that the minimal nature of the introduced modification, relative to the rest of the protein, makes chromatographic separation of reaction components (such as protein starting material and product) extremely difficult, disqualifying more accurate means of conversion calculations. However, this principle also suggests that the physical and chemical properties of the protein are not significantly altered post-reaction, imparting the modified protein products with similar ionization qualities to the starting material and allowing for comparison between the relative intensities in the deconvoluted LC-MS spectra to calculate conversions.

To further validate this approach, we conducted a more rigorous HPLC analysis of two model reactions (**Figure 21**), the installation of either bromohomonorleucine (Bhn, Substrate 1), or trimethyllysine (Me_3Lys, Substrate 2) onto human histone eH3.1-Dha9 (e stands for epitope tag, indicating a C-terminal FLAG-HA tag). We reasoned that full HPLC characterization of small molecule reaction components, protein starting material and protein products would allow us to validate the similarity between starting material and product protein species and confirm the difficulty in separating them without a longer gradient and specialized column. Reactions were conducted with individual reaction components and with full reaction mixtures with or without protein or blue light illumination, then run on an HPLC over a longer gradient (30 min) and while collecting UV chromatograms (220 nm). The chromatograms of the individual components identified the peaks associated with buffer, photocatalyst, catechol and small molecule substrates in response to light or other reaction components minus the protein substrate (**Figure 21**).

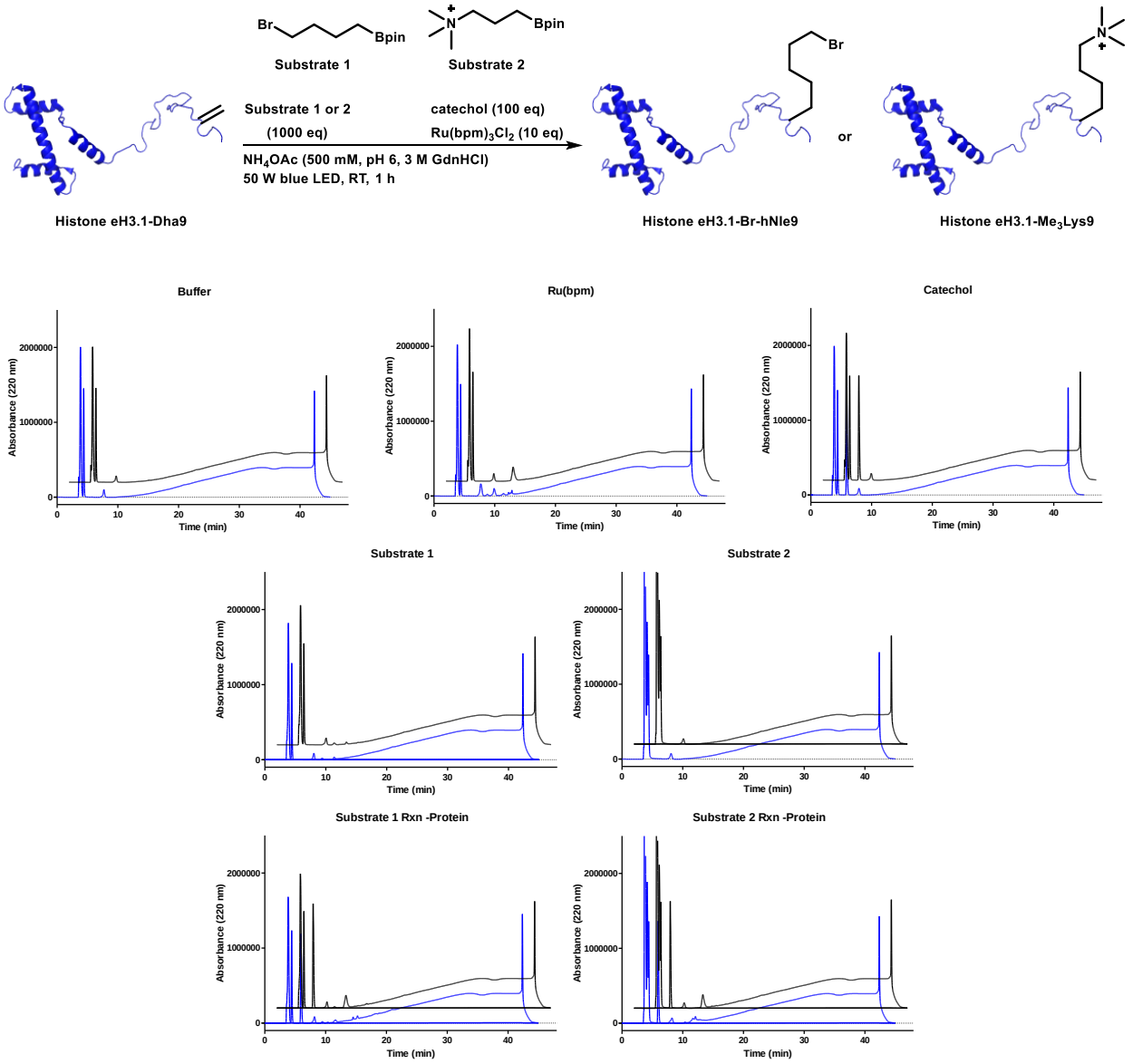


Figure 21: (Top) Reaction details for the installation of Bhn (from Substrate 1) or Me_3Lys (from Substrate 2) onto histone eH3.1-Dha9 used for HPLC analysis. (Bottom) HPLC chromatograms (220 nm) of individual reaction components, as well as full reaction mixtures minus the Dha containing protein, both with (blue lines) and without (black lines) blue light illumination. Note that no peaks appeared in the region expected for protein peaks (~22 min, see below).

Next, a standard curve was established by running the protein starting material, histone eH3.1-Dha9, on the HPLC at three different concentrations, all in triplicate (**Figure 22**). This allowed for the crude reaction protein product peak areas to be compared with the concentration gradient to allow total protein yield to be calculated. Notably, the protein peaks occurred well outside the regions associated with the small molecule reaction components (**Figure 21**).

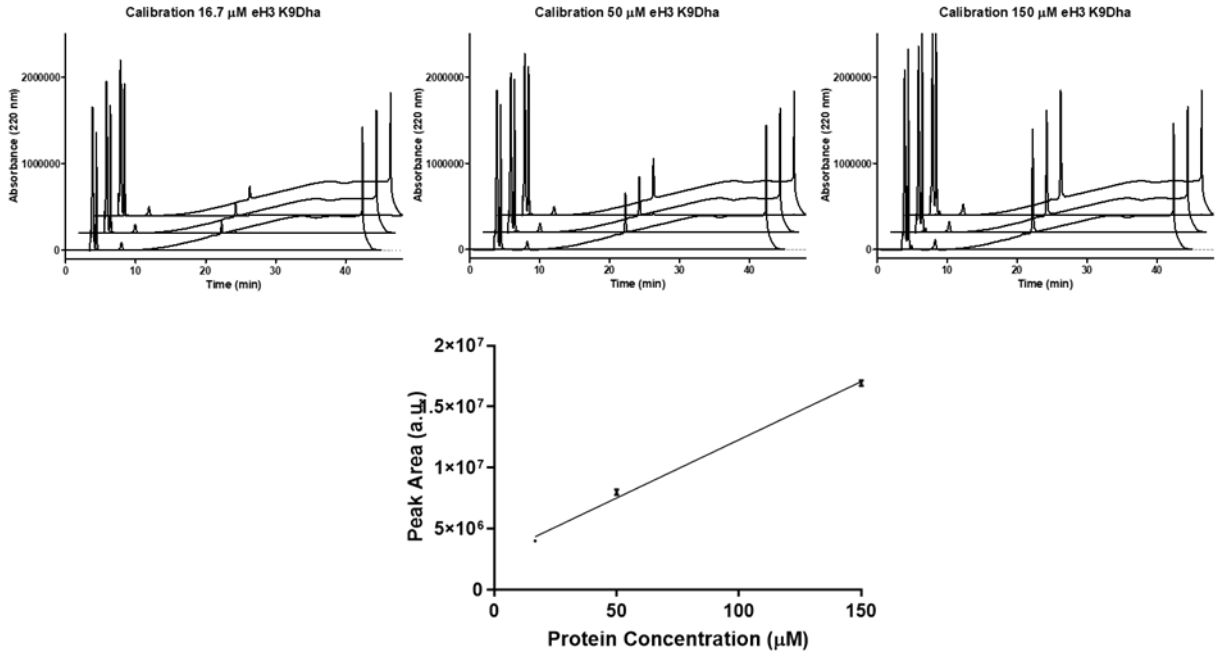


Figure 22: (Top) HPLC chromatograms (220 nm) measured in triplicate for three concentrations of the protein starting material histone eH3.1-Dha9. (Bottom) Standard curve allowing for total protein yield to be calculated based on the protein peak area measured by HPLC.

Lastly, full reactions were performed in triplicate with both Substrate 1 and Substrate 2 and analyzed on both the HPLC (generating 220 nm chromatograms) and intact protein LC-MS (generating a TIC chromatogram, ion series, and deconvoluted spectra) (**Figure 23**). While the deconvoluted spectra showed a ~3:1 ratio of product and starting material Dha for both substrates, both the HPLC and intact protein LC-MS generated chromatograms gave single protein peaks, confirming our expectation that starting material and product protein species could not be chromatographically separated. However, our standard curve (**Figure 22**) allowed for yield determination, > 90% in every case. Furthermore, the consistency seen in the intact protein LC-MS data between the spectra for starting material and products justifies our use of intact protein LC-MS to monitor and semiquanititatively analyze our reactions.

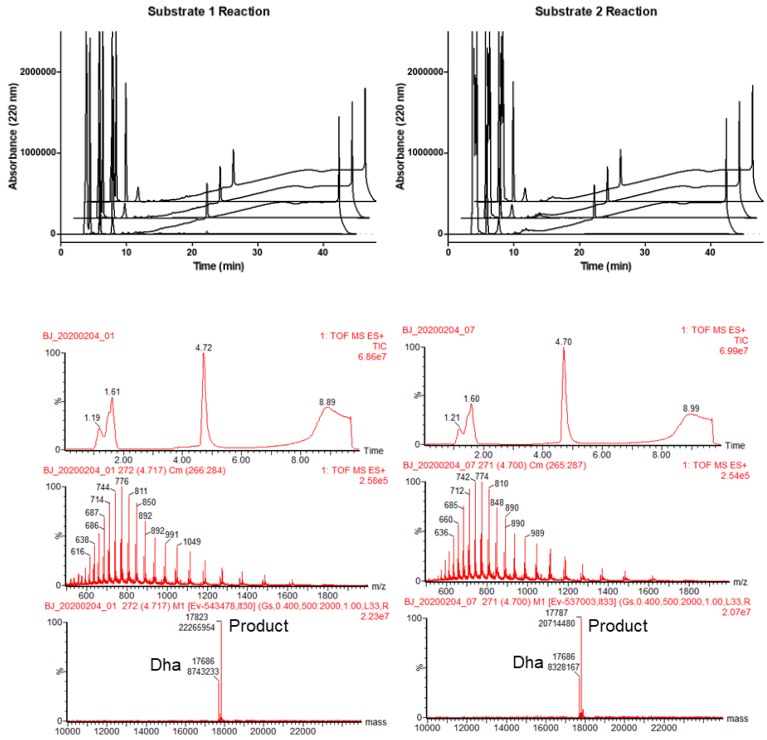
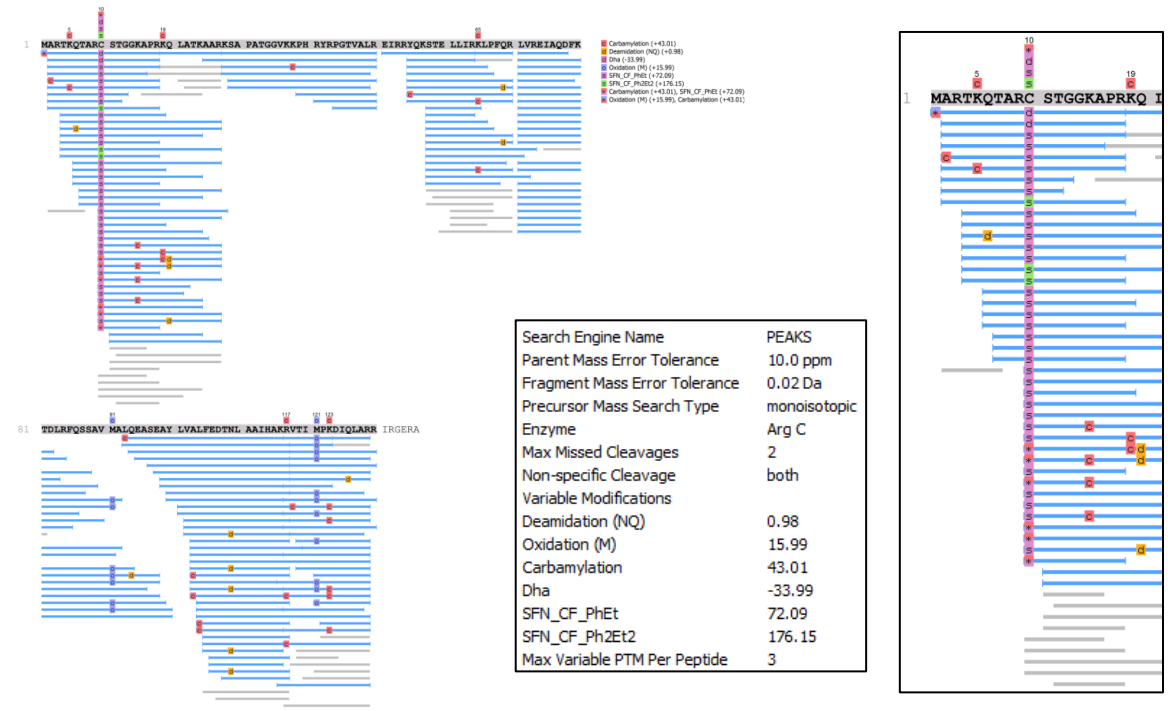
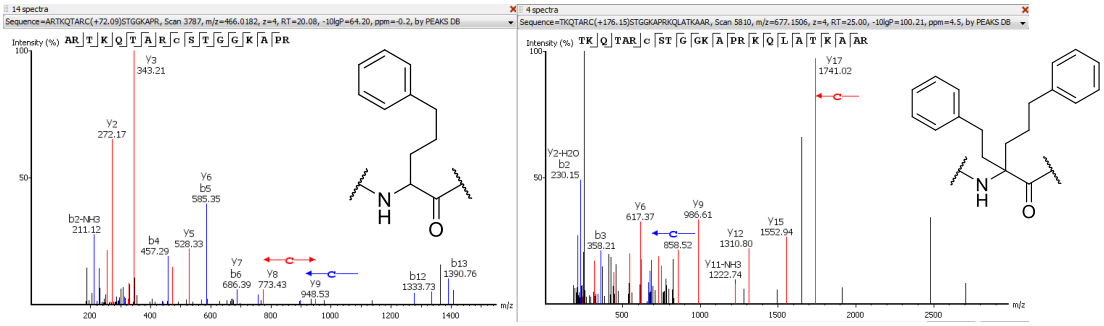


Figure 23: (Top) HPLC chromatograms (220 nm). (Bottom) Intact protein LC-MS chromatograms (TIC), ion series and deconvoluted spectra for the model reactions.

To further characterize the modified protein products, all samples from both the substrate and protein scope were analyzed by LC-MS/MS, which can identify expected modifications, as well as residual Dha and double addition (**Figure 24**), at the site of interest within the protein by tandem mass spectrometry performed on peptides produced after protease digestion of the modified protein of interest[2,21]. All modified proteins gave full sequence coverage and modification masses were easily identified at the expected site. An example is provided in **Figure 24**, identifying the installation of homohomophenylalanine (hhPhe) and corresponding side products as well as some residual starting material Dha. Full LC-MS/MS details are included in the Experimental Details section for this chapter and can be found in the corresponding publication[208]. Special thanks to Dr. Simon Nadal for conducting the LC-MS/MS confirmation experiments, and Dr. Andrew Giltrap for assistance with HPLC experiments.



Single alkylation *Double alkylation*



Dehydroalanine

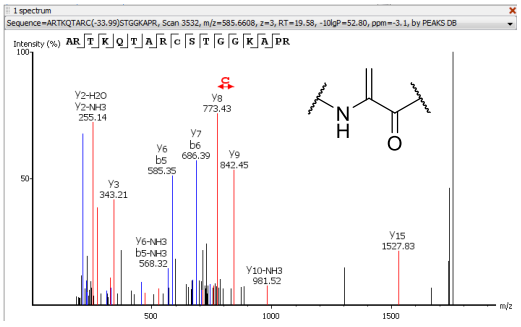


Figure 24: Representative LC-MS/MS results for the installation of hhPhe onto histone H3-Dha9, with full sequence coverage (top) and spectra (bottom) for expected species.

Substrate scope limitations

While a vast range of sidechains can be installed site-specifically into proteins with the light-driven radical Dha modification reaction described above, many sidechains and functional groups proved to be incompatible (**Figure 25**). Perhaps most disappointing was the inability to install phenyl radicals onto Dha, through either the phenyltrifluoroborate or phenylboronic acid 'off-protein' radical precursors. This would have provided access to valuable Phe and Tyr sidechains and their PTMs. In addition, methyl, tertbutyl, allyl, and acyl boronate substrates failed to react.

There are several reasons that may explain lack of reactivity of these various alkylboronate substrates. Perhaps their hydrolysis rates were too slow[210], failing to generate the required boronic acid needed to form the boronic acid catechol ester *in situ*. Additionally, the tested photocatalysts may not have been strong enough to overcome their high redox barrier[217]. Or, alternatively, radical activation may have occurred, but may have resulted in self-termination before Dha addition could occur. In conclusion, the sidechains accessible via the substrates in **Figure 25** remain enticing targets for future Dha modification chemistry.

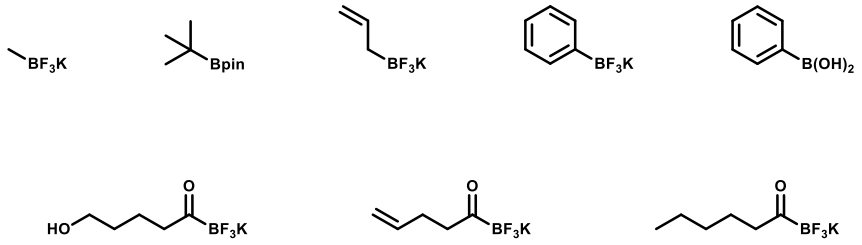


Figure 25: Alkylboronate substrates incompatible with the protein modification method.

‘On-protein’ chemistry afforded by the installation of electrophile containing sidechains

Installation of various electrophile containing sidechains

During the substrate scope investigation described in previous sections, we discovered that the light-driven radical Dha modification reaction was highly tolerant to a variety of ‘off-protein’ radical precursors with electrophilic functional groups. This allowed for the installation of electrophiles into proteins, including esters, azides, and halogens (Cl, Br, and I) that were incompatible with previous methods (**Figure 26A**)[134,135]. Proteins typically lack electrophilic functional groups, and previous protein modification chemistries have utilized nucleophilic residues (e.g. Lys and Cys) to attach ‘off-protein’ electrophilic targets[118,179]. With our ability to site-specifically install electrophilic character into protein sidechains, we reasoned that we could invert this strategy, using ‘off-protein’ nucleophiles to selectively react with ‘on-protein’ electrophilic sidechains.

We focused primarily on sidechains containing alkyl halides, as their incorporation into proteins and subsequent studies of their ‘on-protein’ reactivities has been limited. Previous literature had only described a few methods involving amber codon expression, allowing for alkyl halides to be incorporated into proteins through extension of Lys[223] or Tyr amino acids[224,225], contrasting our method which can install sidechains directly onto the protein backbone with far greater structural flexibility (**Figure 26A**). Initially, the compatibility of our reaction conditions was assessed during installation of various alkyl halide sidechains (**Figure 26B**). Intriguingly, while sidechain bromonorleucine (Br-Nle) could be installed easily using standard conditions (80% conversion), iodonorleucine (I-Nle) exhibited mediocre conversion (~50% conversion). However, the product could be

desalted, degassed, and the reaction repeated up to three times to increase conversion to around 75%. The ability for the alkyl iodide containing sidechain to withstand three rounds of the reaction is further testament to the general tolerance of our method to electrophiles and alkyl halides specifically. Furthermore, two ribose containing sidechains with either bromo or chloro substituents were stored at -20 °C (liquid in the buffer used) for 3 months as a crude reaction mixture and exhibited no degradation or accumulation of unwanted adducts (**Figure 26B**).

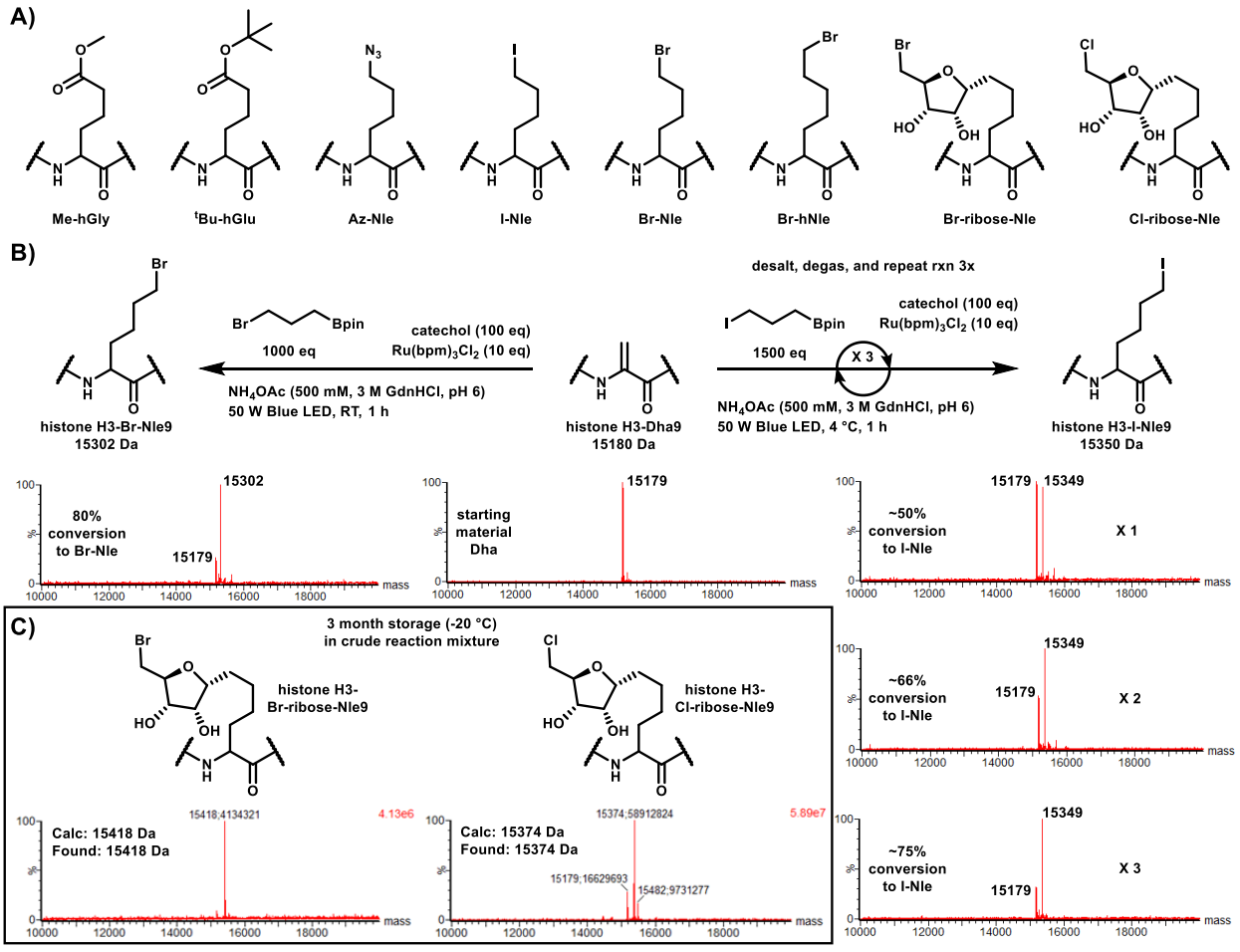


Figure 26: A) Electrophile containing sidechains installed with the light-driven Dha modification reaction. **B)** Installation of Br-Nle and I-Nle details. **C)** Long-term stability testing of two alkyl halide containing sidechains as a crude reaction mixture.

Slow halogen exchange of alkyl halide sidechains

Next, the post-reaction stability of two halogen containing sidechains was investigated. Bromonorleucine (Br-Nle) and iodonorleucine (I-Nle) were installed onto histone H3-Dha9 (see **Figure 26B**), desalted, and incubated in a mildly acidic buffer (pH 6, 37 °C) and monitored over time via intact protein LC-MS (**Figure 27A**). Surprisingly, a new adduct was slowly formed, reaching nearly full conversion after 64 hours. The adduct corresponded to the mass of chloronorleucine (Cl-Nle), at an indistinguishable rate between Br-Nle and I-Nle. Chloride ions were present in the buffer (3 M from GdnHCl), suggesting 'on-protein' halogen exchange passively occurring in the buffer.

We repeated the incubation at a higher pH (pH 11) to promote more deprotonation of potential nucleophiles both within the protein and in the surrounding buffer (**Figure 27B**). As expected, halogen loss for both sidechains Br-Nle and I-Nle was far quicker in these conditions, with most species corresponding to masses that are likely due to halogen exchange, hydroxyl substitution, or elimination after only 16 hours. It is worth noting that these adducts are speculative and conformation of their identities was outside the scope of these initial stability studies. It is possible that the elimination adduct may instead be internal cyclization between nucleophiles in the protein (e.g., Lys) displacing the halogen. However, we could conclude that the Br- and I- containing sidechains were moderately stable near neutral pH, only susceptible to halogen exchange with chloride ions. In contrast, basic pH promoted an accumulation of reaction products where the halogen is lost from the sidechain, hinting that it was possible to control the conditions to promote selective reactions with the alkyl halide.

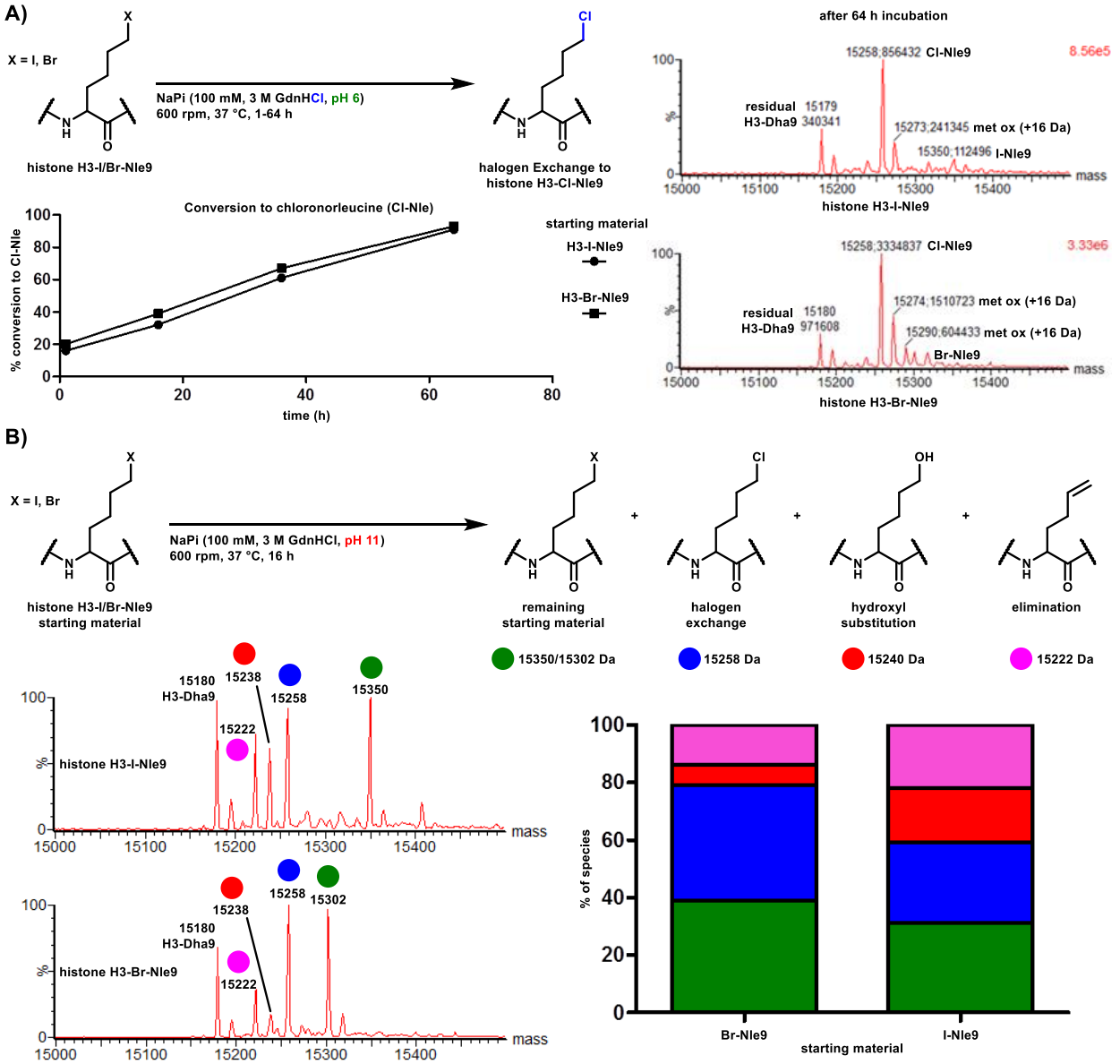


Figure 27: A) Slow halogen exchange of Br-Nle and I-Nle to Cl-Nle at mildly acidic pH. **B)** Quicker degradation of Br-Nle and I-Nle to several species at basic pH. Dha peaks (15180 Da) were disregarded in conversion calculations as the transformation of interest consisted of the Br-Nle and I-Nle starting materials.

Reactivity of alkyl halide sidechains with “off-protein” nucleophiles

Pleased by the apparently stability of the Br-Nle and I-Nle sidechains and encouraged by the unexpected halogen exchange occurring near neutral pH revealed in the previous section (**Figure 27**), we next explored whether ‘off-protein’ nucleophiles could be alkylated by these electrophilic sidechains. This strategy could be, in effect, a reversal of traditional protein alkylation chemistry[118,179], providing a unique electrophilic sidechain that could react selectively with ‘off-protein’ nucleophiles (**Figure 28A**).

To test this strategy, a range of ‘off-protein’ nucleophiles were screened against both Br-Nle and I-Nle sidechains installed on histone H3-Dha9 (**Figure 28B**). Pleasingly, εC-PR_3 and εC-N_3 bonds could be easily introduced in good conversion at neutral pH in only 12 hours with common biological reagents TCEP (25 mM) or sodium azide (200 mM), respectively. The common reducing agent, βME, reacted quickly at slightly elevated pH (25 mM, 4 h, pH 10) to create a εC-SR bond. More difficult were the family of methylamines (mono-, di- and tri-), which due to their high pKa values, required elevated concentrations and pH (0.5 M, pH 11) to promote reactivity. However, while generally exhibiting lower conversions, these reactions allowed for the direct installation of Lys PTMs crucial to histone function (mono-, di-, and trimethyllysine)[61] through εC-NR_3 bond formation, an intriguing example of site-specific *in situ* Lys PTM formation through the addition of simple small molecule reagents. It is worth noting that the reactions often went to near full conversion, and products that were not the desired alkylation adduct were often one of the side products observed previously (**Figure 27A and B**). Reaction details and characterization, done through intact protein LC-MS, can be found in the **Experimental Details** section of this chapter.

In conclusion, alkyl halide containing sidechains such as Br-Nle and I-Nle introduced through light-driven radical Dha modification provide powerful platforms for further 2e- chemistry through their unique and selective electrophilic character. The next chapter will investigate their use as proximity driven protein-protein crosslinking targets.

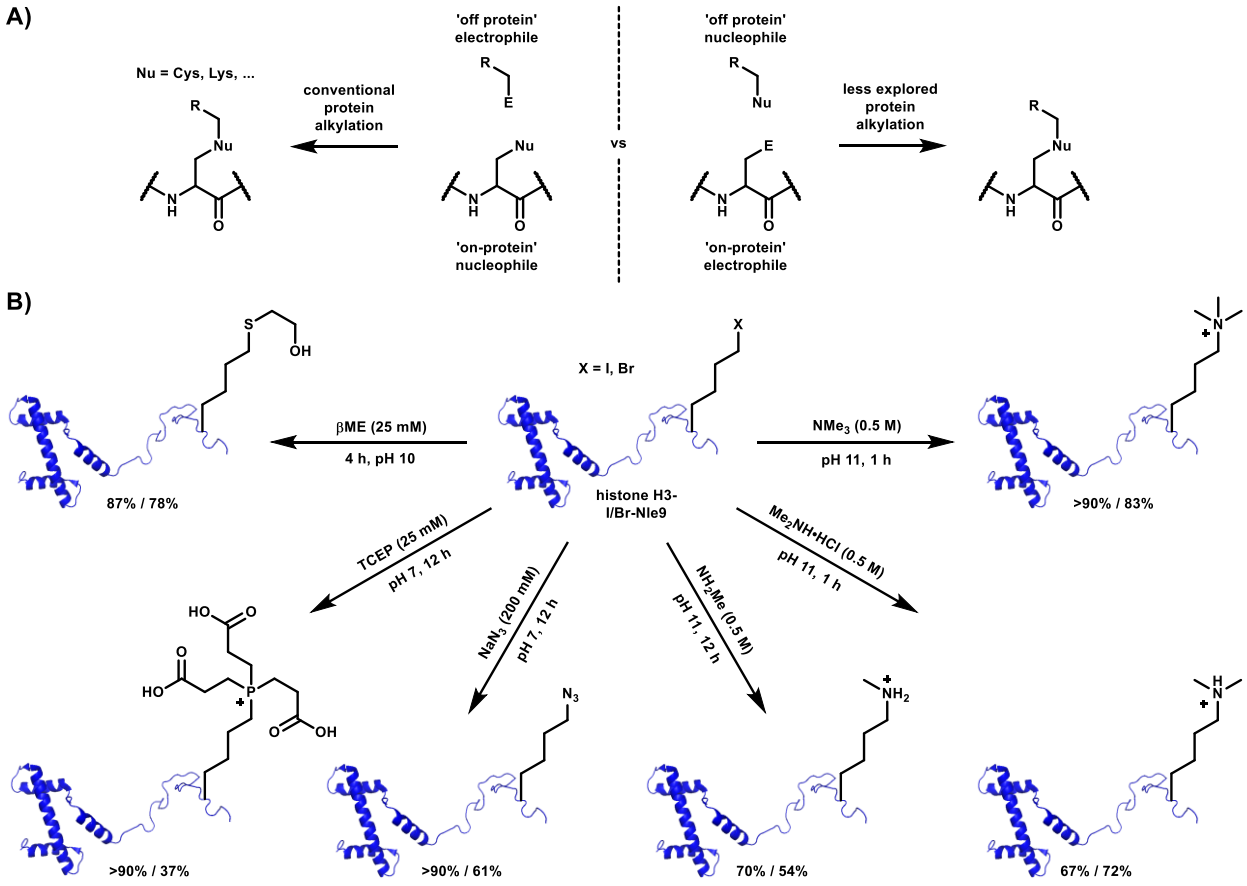


Figure 28: A) Schematic comparing the unselective traditional strategy of reacting 'off-protein' electrophiles with 'on-protein' nucleophiles (left) versus our new selective strategy of reacting installed 'on-protein' electrophiles with 'off-protein' nucleophiles (right). **B)** Summary of investigated reactions involving alkylation of 'off-protein' nucleophiles with 'on-protein' electrophile sidechains Br-Nle and I-Nle. Conversions to the desired adduct are given below (I-Nle left, Br-Nle right).

Conclusion

Summary and advantages brought about by this chapter

Chapter 1 focused on the development and optimization of new chemical methods to site-specifically modify proteins. While the methods were often tested on histones, providing a platform for future applications in upcoming Chapters, they could easily be extended to other proteins for a variety of protein PTM studies, as shown in protein scope investigations[208]. All of the methods either directly utilized previously installed Dha reactive tags to incorporate new sidechains of interest, or utilized said sidechains to investigated new 'on-protein' chemistry, further testifying to the usefulness of Dha as a 'tag-and-modify' target in protein modification strategies[126].

The Chapter consisted of the development and optimization of a light-driven radical Dha modification reaction, which utilized alkylboronate 'off-protein' sidechain radical precursors to form easily oxidizable alkylboronic acid catechol esters *in situ* that were then activated by the mild photocatalysts, $Ru(bpy)_3Cl_2$ or $Ru(bpm)_3Cl_2$. The generated radicals then selectively added to the previously installed SOMOphile Dha, producing the desired sidechain with high conversion and yield. The major advantage of this method was its tolerance toward not only many native and PTM containing sidechains, but also towards reactive and electrophilic functional groups such as azides, esters, and halogens that were not compatible in earlier methods[134,135].

Lastly, the electrophilic sidechains were explored further as 'on-protein' electrophilic targets for 'off-protein' nucleophile alkylation. Various alkyl halide sidechains successfully and selectively reacted with common nucleophilic reagents such as TCEP, βME, sodium azide, and methylamines, successfully creating C-P, C-S and C-N bonds on the protein

sidechain. Of note, the methylamine (mono, di, and tri) nucleophilic additions to I-Nle and Br-Nle residues created methylated Lys PTMs, crucial histone PTMs in epigenetic systems[61]. In addition, it was shown that at near neutral pH, halogen exchange could occur between the installed alkyl halide sidechains and chloride ions present in the buffer. Future Chapters will further exploit these valuable electrophilic sidechains for protein crosslinking (see **Chapter 2**).

Prospects for future work

Throughout the course of this Chapter, a powerful new method was developed for modifying Dha residues in proteins. The resultant installed sidechains, containing reactive residues such as alkyl halides, were further exploited, providing examples of new 'on-protein' chemical transformations. However, there remain many unresolved challenges to Dha centered protein modification methods, which proved to be outside the scope, or unattainable during the time course of this thesis work.

First, there are still many native and PTM containing sidechains that have yet to be accessed through Dha modification reactions on full proteins. Chief among them is the addition of aryl groups to Dha[226,227], which would provide a route to Phe, Tyr, Trp and His installation, as well as their PTMs. While tested, various arylboronates failed to react with our light-driven radical chemical approach. Furthermore, direct introduction of Asp and Asn residues and their PTMs have failed up to this point.

Second, Dha modifications on proteins intrinsically lack diastereoselectivity, resulting in a mixture of $_{D}$- and $_{L}$- sidechains[134,135,208,226,227]. While great efforts were made to characterize these mixtures (see various applications in **Chapter 2**), the methods described here did not solve the diastereomer problem. A recent publication describes a new method that retains the stereocenter of a Cys residue during desulfurization and radical C-C bond formation[228], but many of the potential side products from radical polymerization are not characterized, and it lacks the substrate scope accessible from our methods. Future Dha modification methods should focus on chemical, or even enzymatic[229], means of guiding the formation of the desired diastereomer. Only then can it truly be viewed as an authentic 'post-translational mutagenesis' technique[135].

Lastly, the methods described in this chapter all occurred *in vitro*, as do most Dha modification strategies[127,226,227]. Efforts were undertaken in this thesis to develop an *in vivo* Dha modification reaction (see **Chapter 3**) but have yet to reach fruition. While previous methods were certainly not biocompatible, due to the use of hazardous chemicals such as sodium borohydride[135], the more recent light-driven reactions described here offer more promise. In fact, ruthenium based photocatalysts have been shown to penetrate cell membranes[230], and various biogenic catechol derivatives such as dopamine and $_{L}$-DOPA were shown to be compatible with our light-driven reaction. If future work succeeds in modifying Dha residues in proteins *in vivo*, it will perhaps offer the first general platform for site-specific sidechain alterations in target proteins within living cells.

Statement on collaboration

The work outlined in this thesis chapter was completed with the assistance of many others, who both performed some of the work before I started the project and assisted with various aspects throughout the course of the work. Their contributions are explicitly stated in the text when appropriate and outlined in full here.

Dr. Tom Wright began work on photocatalytic activation of 'off-protein' radical precursors for Dha modifications, and identified alkyl trifluoroborates as promising radical precursors when used with the strongly oxidizing $Ir[dF(CF_3)ppy]_2(bpy)PF_6$ photocatalyst. Professor Charlie Fehl then took over the project and identified catechol as a useful additive that required lower E_{ox} activation by forming boronic acid catechol esters that could be activated with the milder $Ru(bpy)_3Cl_2$ photocatalyst. Charlie, along with collaborators in the group of Professor Richard Compton, Dr. Lifu Chen and Dr. Christopher Batchelor-McAuley, confirmed the role of catechol through Cyclic Voltammetry experiments. Charlie also performed a small initial substrate scope screen and attempted early protein-protein crosslinking experiments with alkyl halide modified histones.

When I took over the project, I was kindly provided several small molecule radical precursors that were synthesized by Dr. Patrick Isenegger, Dr. Andrew Giltrap, Adeline Poh, Dr. Tim Mollner, and Dr. Oluwatobi Arisa. Various proteins were provided by Dr. Tim Mollner, Dr. Benjamin Bower, and Dr. Patrick Isenegger for substrate and enzymatic studies. Protein LC-MS/MS analysis for both protein-protein crosslinking and protein modification validation were performed by Dr. Simon Nadal and Professor Shabaz Mohammed. The KDM4A enzyme was provided by Grace Roper and Professor Akane Kawamura, who also assisted with Zinc ejection experiments (discussed next chapter).

Experimental Details

General Methods

General experimental procedures

Unless otherwise noted, chemical reagents, media, and *Escherichia coli* cell stocks were obtained from commercial suppliers (Sigma-Aldrich, Fluorochem, Carbosynth, VWR, Alfa Aesar, Fisher Scientific) and used without further purification. Sonication was performed using a Fisher Scientific Model 505 Sonic Dismembrator. Proteins were purified using an Äkta FPLC System UPC-900 (GE Healthcare, UK). Protein crystal structures were analyzed and displayed using MacPyMOL v. 1.3 (Schrodinger, Inc.). Synthetic gene fragments (i.e. for human histone eH3.1 constructs) were obtained from GeneArt Gene Synthesis (Thermo-Fisher). Nucleotide sequences were confirmed by the Source Bioscience DNA Sanger sequencing services based at Oxford University.

Intact Protein LC-MS

Intact protein mass spectrometry was performed on a Waters Xevo G2-S QTof coupled to Water Acquity UPLC. Separation was achieved using a Thermo Proswift (250 mm x 4.6 mm x 5 µm) column with water + 0.1% formic acid (solvent A) and acetonitrile + 0.1% formic acid (solvent B) as the eluent system over a 10-minute linear gradient. Nitrogen was used as the desolvation gas (600 L/h) for positive electrospray ionization. Voltages used were capillary: 3000 V, cone: 160 V. Lock-spray analysis ensured continual calibration against a leucine enkephalin standard solution.

Raw spectra containing multiple charged ion series were deconvoluted using MassLynx (Waters) and its maximum entropy (MaxEnt1) deconvolution algorithm (Resolution: 1.00 Da/channel, Width at half height: ion series/protein dependent, Minimun intensity ratios: 33% Left and Right). Spectra were deconvoluted between 10000 and 20000 Da for *Xenopus laevis* histone H3, between 10000 and 25000 Da for human histone eH3.1, between 5000 and 15000 Da for *Xenopus laevis* histone H4, between 10000 and 30000 Da for NPβ, between 30000 and 50000 Da for AcrA, and between 30000 and 40000 Da for PanC. Any reaction conversions were calculated from relative peak intensities in the deconvoluted spectra. On histones, ~10% baseline methionine oxidation often occurred during production, storage, and use, and these “+16 Da adducts” were combined into the total sums for starting material/products when appropriate.

Tandem Mass Spectrometry

For all samples, modifications were searched as a modification on cysteine. Presence of carbamidomethylation on cysteine suggests presence of non-reduced cysteine in the sample (did not undergo alkylation-elimination to Dha in the first reaction). Presence of Dha as a modification on cysteine indicates that the respective Dha modification reaction did not go to full conversion.

ArgC in-solution digest:

Variant 1: Denatured protein samples, no alkylation

Approximately 10 µg (20 µL) of desalted & denatured modified protein sample were taken in 50 mM TEAB to a total volume of 100 µL and reduced with 10 mM TCEP for 30 min at

RT. Samples were digested with Arg-C (1:20 w/w) in activation buffer (50 mM TEAB, 0.2 mM EDTA, 5 mM TCEP) for 3 h at 37 °C. The reaction was stopped by addition of 10% formic acid to a final concentration of 0.5%. Samples were desalted by C18 (Oasis HLB 10 mg) and dried in a speed-vac before being resuspended in 5% formic acid and 5% DMSO.

Variant 2: with denaturation, with alkylation

Approximately 10 µg of modified protein sample were taken in 8 M urea in 50 mM TEAB containing 20 mM methylamine to a total volume of 100 µL and denatured for 30 min at RT. Samples were reduced with 10 mM TCEP for 30 min at RT and alkylated with 50 mM chloroacetamide for 30 min at RT in the dark. Samples were diluted to 1 M urea and digested with Arg-C (1:20 w/w) in activation buffer (50 mM TEAB, 0.2 mM EDTA, 5 mM TCEP) for 4 h – O/N at 37 °C. The reaction was stopped by addition of 10% formic acid to a final concentration of 0.5%. Samples were desalted by C18 (Oasis HLB 10 mg cartridge) and dried in a speed-vac before being resuspended in 5% formic acid and 5% DMSO.

LysC in-solution digest

Approximately 10 µg (20 µL) of desalted & denatured modified protein sample were taken in 8 M urea in 100 mM TEAB to a total volume of 100 µL. Samples were reduced with 10 mM TCEP for 30 min at RT and alkylated with 50 mM chloroacetamide for 30 min at RT in the dark. The solution was diluted to 6 M urea with 50 mM TEAB and digested with

LysC 1:20 (w/w) over night at 37 °C. Samples were desalted by C18 (Oasis HLB 10 mg) and dried in a speed-vac before being resuspended in 5% formic acid and 5% DMSO.

Tryptic or AspN or Elastase in-solution digest

Approximately 10 µg (20 µL) of desalted & denatured modified protein sample were taken in 8 M urea in 100 mM TEAB to a total volume of 100 µL. Samples were reduced with 10 mM TCEP for 30 min at RT and alkylated with 50 mM chloroacetamide for 30 min at RT in the dark. The solution was diluted to 1 M urea with 50 mM TEAB and digested with AspN or Trypsin or Elastase 1:20 (w/w) 4h to overnight at 37 °C. Samples were desalted by C18 (Oasis HLB 10 mg) and dried in a speed-vac before being resuspended in 5% formic acid and 5% DMSO.

Data Acquisition

Standard data acquisition – Q Exactive

Resulting peptides were separated by nano-flow reversed-phase liquid chromatography Ultimate 3000 UHPLC system (Thermo Fisher Scientific) coupled to a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fischer Scientific). The peptides were loaded on a C18 PepMap100 precolumn (inner diameter 300 µm x 5 mm, 3 µm C18 beads; Thermo Fisher Scientific) and separated on an in-house packed analytical column (75 µm inner diameter x 50cm packed with ReproSil-Pur 120 C18-AQ, 1.9 µm, 120 Å, Dr. Maisch GmbH). The raw data were acquired on the mass spectrometer in data-dependent mode. Automatically switching from MS to higher energy collision induced dissociation MS/MS. Full-scan spectra were acquired in the Orbitrap [scan range 350-

2000 m/z, resolution 70000, automatic gain control (AGC) target 3 x 106, maximum injection time 50 ms]. After the MS scan, the top 10 most intense peaks were selected for HCD fragmentation at 30% of normalized collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500; AGC target 5 x 104; maximum injection time 120 ms), with first fixed mass at 180 m/z.

Data analysis

Searches were performed with Peaks version 8.5 (Bioinformatics Solutions Inc.) for identification and de-novo analysis. The raw MS file was searched against the given protein sequence and a list of contaminants (generated from MaxQuant contaminants database). Samples were additionally searched with MaxQuant against the UniProt human database for confirming the purity of the sample. Precursor mass tolerance was set to 10 ppm. Fragment mass tolerances for HCD was set to 0.02 Da. The corresponding protease was selected with a maximum number of 3 missed cleavages and non-specific cleavage at one end of the peptide. For Elastase, an unspecific search was used. Oxidation (Methionine), Deamination (Asparagine, Glutamine), Carbamidomethylation (Cysteine – except for ArgC variant 1 digest), Carbamylation (lysine, peptide N-term), Amidation (C-terminus) and dehydroalanine (Cysteine, -33.9887) were set as variable modifications, as well as the sample-specific modifications in the following table below. A maximum of 4 variable modifications was set. An FDR of 1% on peptide level and de-novo ALC of 80 was applied. All spectra and identifications were manually validated. For analysis of isotopic pattern, manual analysis with XCalibur Qual Browser 4.0 was performed.

HPLC analysis of modified proteins

HPLC analysis was performed on a Shimadzu 2020 LC-MS instrument with an LCM20AD pump, SPD-20A UV/Vis operating at 220 nm and 280 nm using a Phenomenex Jupiter C-4 (5 mm, 300 Å) 4.6 x 250 mm column operating at a flow rate of 1 mLmin-1. The analysis was performed using a mobile phase of 0.1% TFA in water (Solvent A) and 0.1% TFA in MeCN (Solvent B) using a linear gradient as follows: 0% B for 4 min, 0 to 100% B over 26 min. Chromatograms recorded at 220 nm were analyzed using the Shimadzu LabSolutions software.

Circular Dichroism

CD spectra were recorded on a Chirascan device at 25 ºC. Experiments were performed using 300 µL of protein solution in a 1 mm cuvette in buffer. Spectra were collected from 200 to 260 nm (0.5 nm step, 1 s per point) as a smoothened average of three scans. Buffer solution was used as a blank sample.

Protein Production

Histones H3 and H4 (human and Xenopus laevis)

Histone expression and purification was carried out essentially as previously described[135]. Bacterial expression plasmids encoding all canonical *Xenopus laevis* histones were kindly donated by Prof. Rob Klose, University of Oxford in the pET3 production vector. The gene for the WT human histone eH3.1 (C-terminal FLAG-HA tag, C96A and C110A) was ordered from Thermofischer (GeneArt service) and cloned into the pET3d expression plasmid at the NcoI and BamHI restriction enzyme sites. Quickchange mutagenesis was performed per the manufacturer's instructions (QuikChange II Site-Directed Mutagenesis Kit, Agilent) to create the desired cysteine mutants. The *Escherichia coli* strain BL21(DE3)pLysS was transformed as appropriate and selected on chloramphenicol and ampicillin. Single colonies were used to inoculate 10 mL starter cultures in LB broth with the same antibiotics. Flasks containing 500 mL 2xTY media were inoculated with 1% v/v of starters, grown at 37 °C until OD_{600} = 0.4-0.8. The expression of histones was induced with 0.5 mM IPTG and allowed to proceed for 2 h before harvest and resuspension into 5-fold volume/weight "wash buffer" (50 mM Tris, pH 7.5, 100 mM NaCl with a protease inhibitor cocktail). Suspensions were flash-frozen and stored at -80 °C until lysis. Lysis proceeded via sonication in the presence of 1 mg DNase for 5 x 30 s bursts at 40% amplitude. The sonicate was centrifuged for 20 min at 20k rpm at 4 °C. The supernatant was discarded, and the pellet resuspended in 40 mL ["wash buffer" + 1% Triton-X detergent]. Sonication was repeated once at 40% amplitude, 30 s, and the suspension centrifuged at 20k rpm for 10 min. The pellet was washed twice more in this fashion, then once with the non-Triton containing "wash buffer." 1 mL of DMSO was added to the pellet

and crudely mixed with a spatula to aid histone desolvation for 10 min. 10 mL “unfolding buffer” (7 M GdnHCl, 20 mM Tris, pH 7.5, 10 mM DTT) was added and shaken for 1 h at RT, then the mixture was centrifuged for 10 min at 20k rpm at room temperature. The supernatant was loaded onto an S200 size exclusion column (GE Healthcare) pre-equilibrated with “SAU-100” buffer (7 M urea, 20 mM NaOAc, pH 5.2, 100 mM NaCl, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine). Protein was eluted with SAU-100, analyzed by SDS-PAGE, and histone fractions were pooled and concentrated to 1-4 mL. Cation exchange chromatography was used to further purify histones (HiTrap SP 5 mL) using a linear gradient of 0-100% SAU-1000 buffer (“SAU-100” with 1000 mM NaCl final concentration). Pure fractions were pooled, dialyzed against water (with 2 mM β-mercaptoethanol) and lyophilized. The sequences of the WT histones used to create further mutants for expression and purification are given below. Note that “WT” in the setting of this thesis refers to the histone sequences with Cys mutated to Ala to allow Cys based chemistry. Dha formation was also carried out as previously described[135,208].

Xenopus laevis histone H3 WT (C110A)

ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKS
TELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRV
TIMPKDIQLARRIRGERA

Human histone eH3.1-WT sequence (C96A, C110A, C-terminal FLAG-HA epitope tags)

ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKS
TELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEAAEAYLVGLFEDTNLAAIHAKRV
TIMPKDIQLARRIRGERAGGDYKDDDDKSAAGGYPYDVPDYA

Other proteins

AcrA, PanC, and NPβ WT and Cys mutants were kindly provided by Dr. Benjamin Bower, Dr. Patrick Isenegger, and Dr. Tim Mollner, and were expressed and purified as previously described, as was Dha formation[133,135,221].

Protein modification characterization details

General procedure for light-driven radical Dha modification

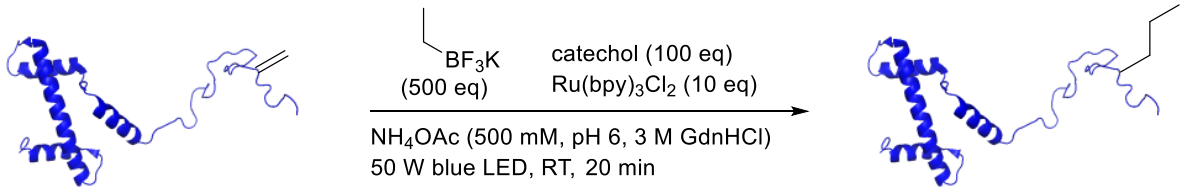
All solutions were degassed for at least 8 h in a glovebox (<6 ppm O_2). Clear glass vials (Chromacol 300 µL fixed insert vial, clear, screw top, Thermo Scientific for < 100 µL, and 2 mL CLR RAM VIAL 9MM THD, 32009-1232, Novetech for ≥ 100 µL) with gas-tight caps (Cat. No VWRI548-3298, VWR) were used. Standard reactions consisted of mixing the Dha-containing protein of choice into the desired reaction buffer in the glovebox, followed by the sequential addition of photocatalyst, catechol, and chemical substrates from 100X stocks prepared fresh in the glovebox. Most reaction optimization and chemical substrate screening reactions were performed our model protein substrate X.l. Histone H3-Dha9 (1 mg/mL) in denaturing buffer (500 mM NH4OAc, 5 M guanidinium chloride, pH 6.0) at a final concentration of 1 mg/mL (66 µM) in volumes of 50-200 µL. All reagents were first ported into a glovebox where subsequent stock solutions and reactions were prepared. All reagents were water soluble at their final concentrations and required no cosolvents unless explicitly noted. The reactions were then mixed thoroughly by pipette, capped, and removed from the glovebox for irradiation. A photobox[208] was used for up to 7 reactions at a time with blue LED (50 W). Short reaction times (<20 min) did not result in significant temperature increases but longer reaction times (>20 min) could have their temperatures controlled by submerging the reaction vials in a glass beaker filled with water at the desired temperature. Irradiation proceeded for the desired time, afterwards which an aliquot of the reaction was diluted 25-fold for mass spectrometric analysis (2 µL in 48 µL water + 0.1% formic acid) and conversions calculated relative to total ion counts of starting material, single, double addition, and any side reactions observed. Protein recovery was

generally above 85% using PD SpinTrap G-25 (GE Healthcare) desalting columns and tracking overall protein absorbance, though this analysis was not performed for all conditions and substrates. The modified proteins could be stored in the freezer as a crude reaction mixture for several months without any degradation or appearance of new adducts, and in several cases, incomplete reactions could be continued simply by degassing the reaction mixture again and continuing with irradiation.

Below, intact protein LC-MS results showing the TIC chromatograms, ion series, and deconvoluted spectra, as well as LC-MS/MS results, are given for every substrate and protein scoping reaction. Furthermore, reaction details are given for the ‘on-protein’ electrophile reactions. Details for light-driven radical Dha modification reactions can also be found in the corresponding publication[208].

Substrate scope for light-driven radical Dha modification

Installation of norvaline (Nva)

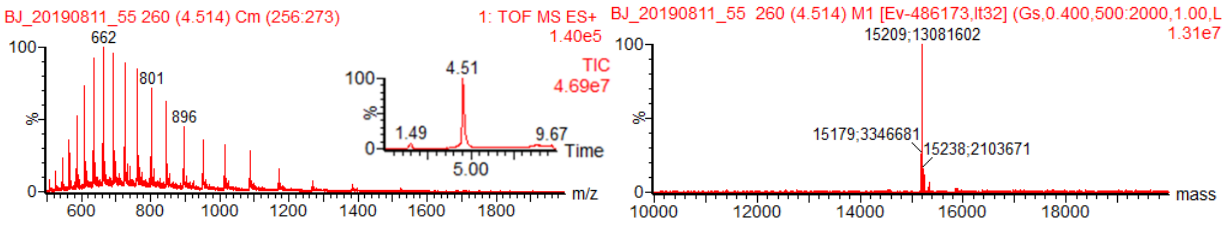


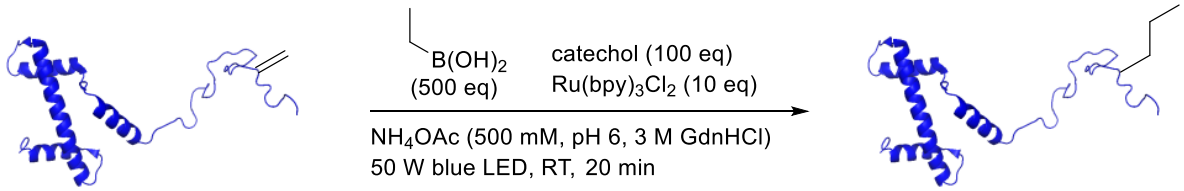
histone H3-Dha9 — **82% Conversion (14% double addition)** — histone H3-Nva9

Calculated mass: 15210 Da. Observed mass: 15209 Da

Calculated double addition mass: 15238 Da. Observed mass: 15238 Da

Calculated Dha mass: 15180 Da. Observed mass: 15179 Da



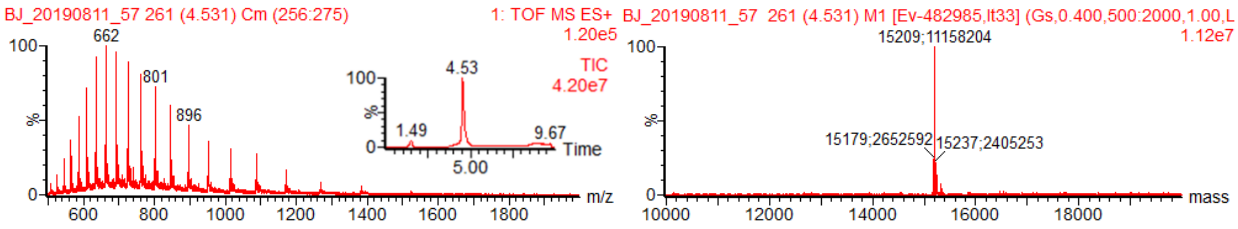


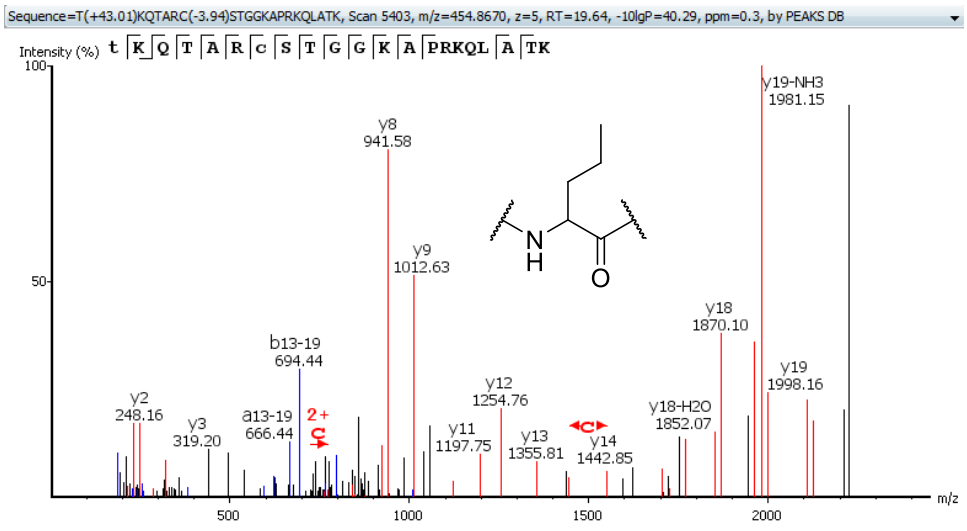
histone H3-Dha9 — **84% Conversion (18% double addition)** — histone H3-Nva9

Calculated mass: 15210 Da. Observed mass: 15209 Da

Calculated double addition mass: 15238 Da. Observed mass: 15237 Da

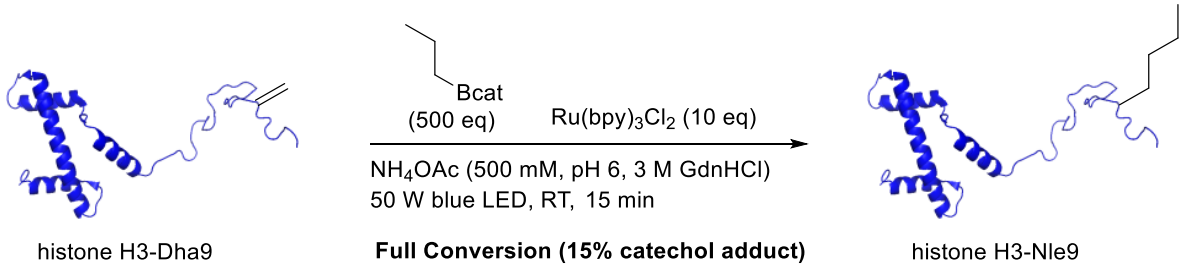
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





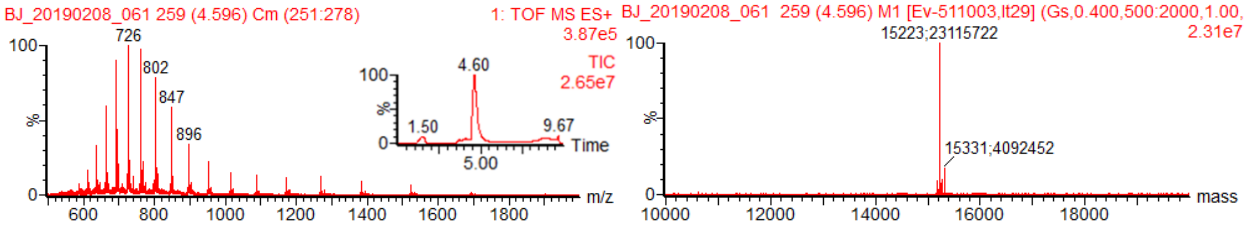
LC-MS/MS of histone H3-Nva9: ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

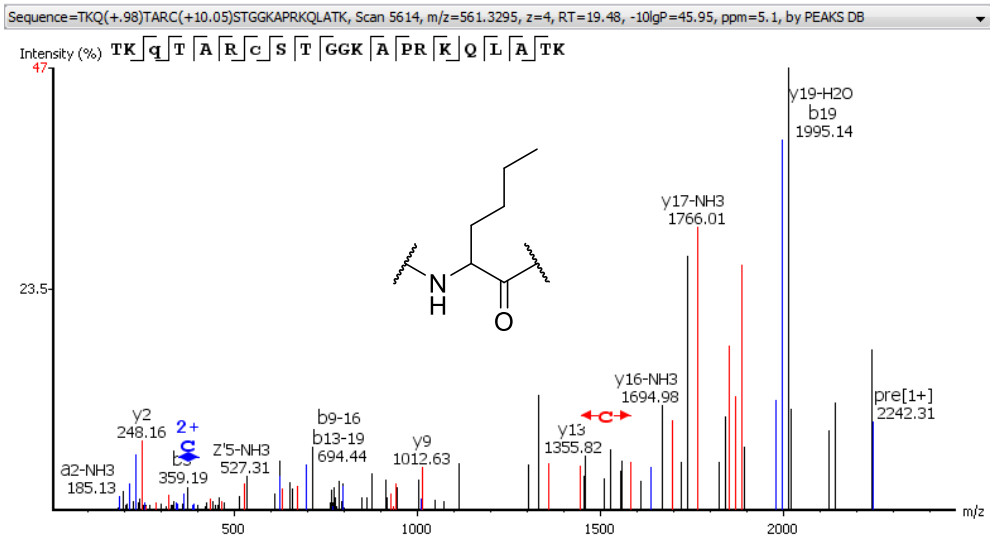
Installation of norleucine (Nle)



Calculated mass: 15224 Da. Observed mass: 15223 Da

Calculated catechol mass: 15311 Da. Observed mass: 15331 Da

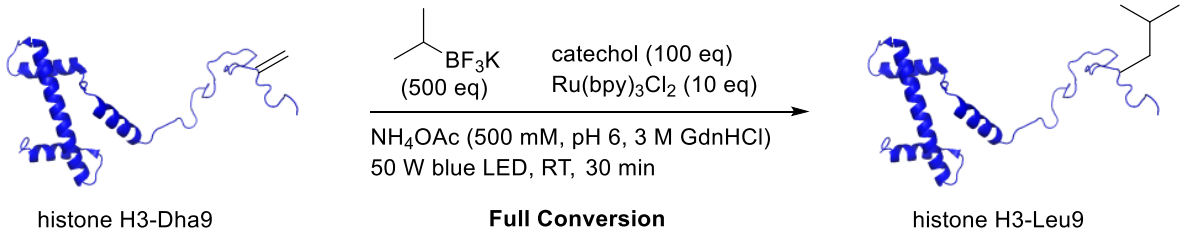




LC-MS/MS of histone H3-Nle9: ArgC digest leads to 99% sequence coverage (1% FDR).

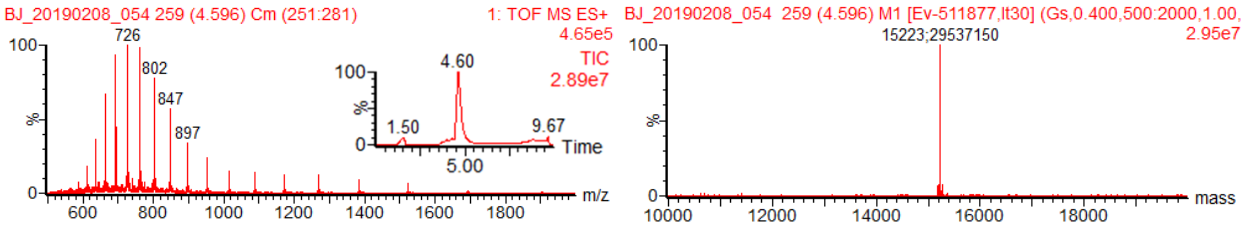
Q Exactive (HCD)

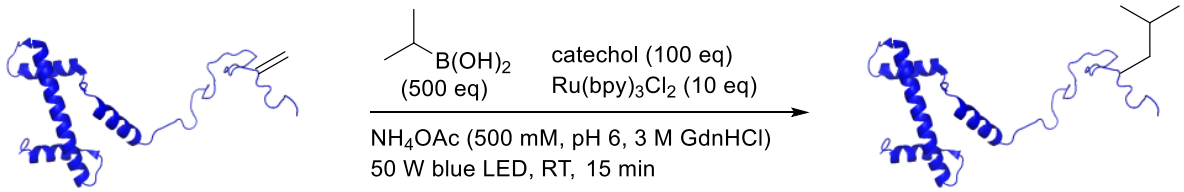
Installation of leucine (Leu)



histone H3-Dha9 — **Full Conversion** — histone H3-Leu9

Calculated mass: 15224 Da. Observed mass: 15223 Da



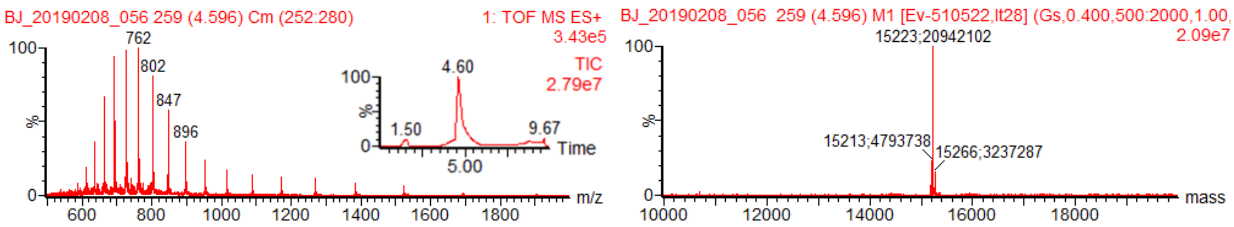


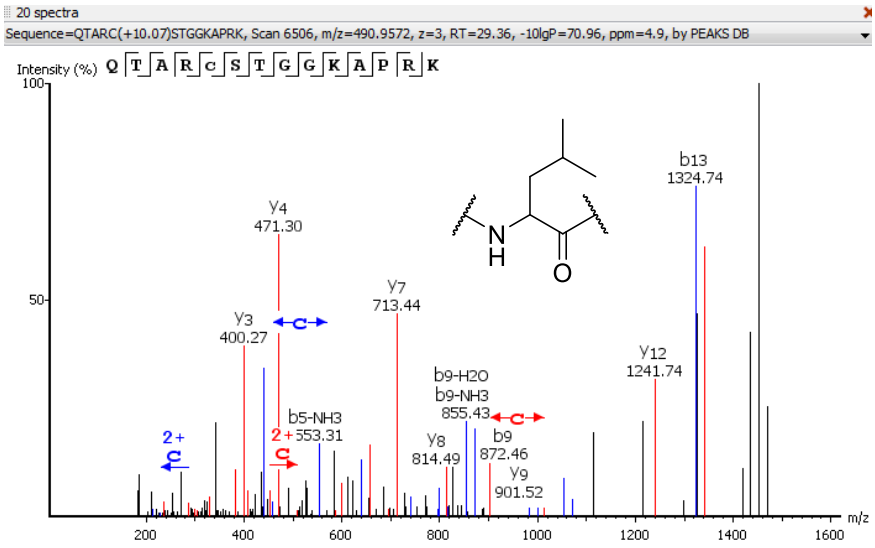
histone H3-Dha9 — **83% Conversion (13% double addition)** — histone H3-Leu9

Calculated mass: 15224 Da. Observed mass: 15223 Da

Calculated Dha + Met Ox mass: 15212 Da. Observed mass: 15213 Da

Calculated Double Addition mass: 15266 Da. Observed mass: 15266 Da

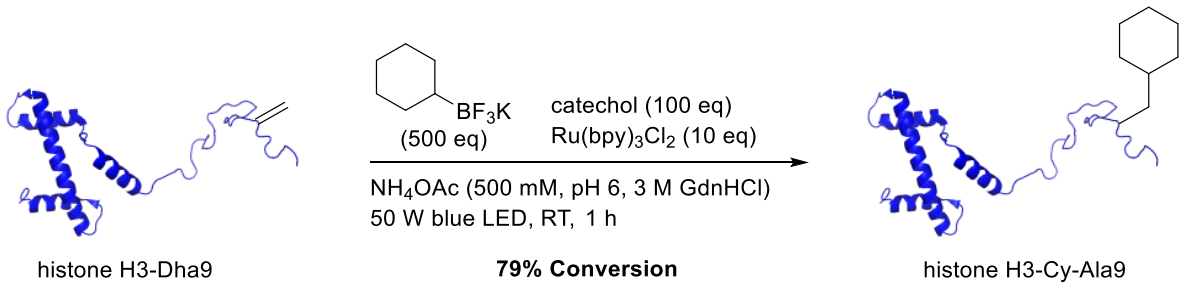




LC-MS/MS of histone H3-Leu9: LysC digest yields 94% sequence coverage (1% FDR).

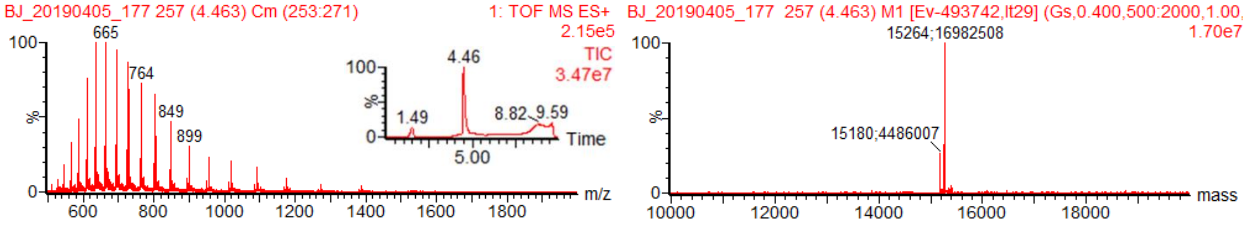
LC-MS/MS indicates full conversion to the desired product. Q Exactive (HCD).

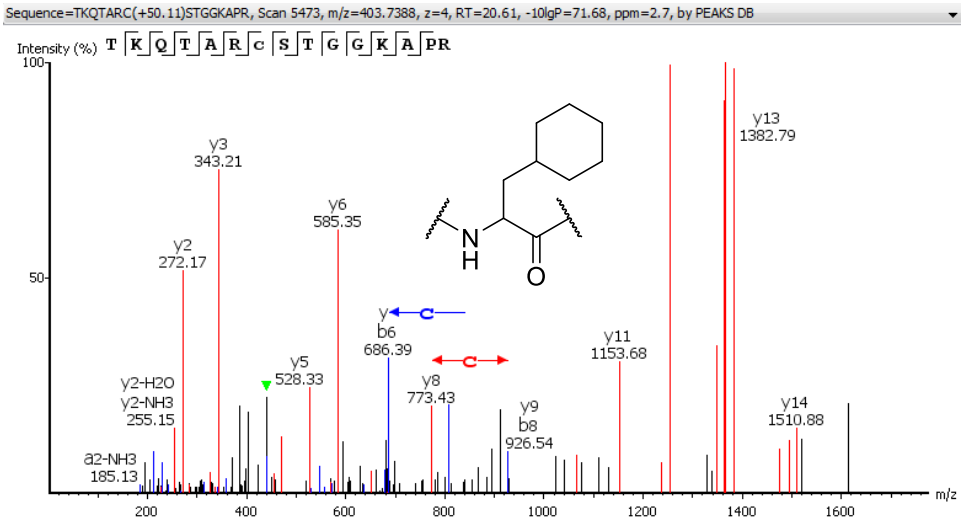
Installation of cyclohexylalanine (Cy-Ala)



Calculated mass: 15264 Da. Observed mass: 15264 Da

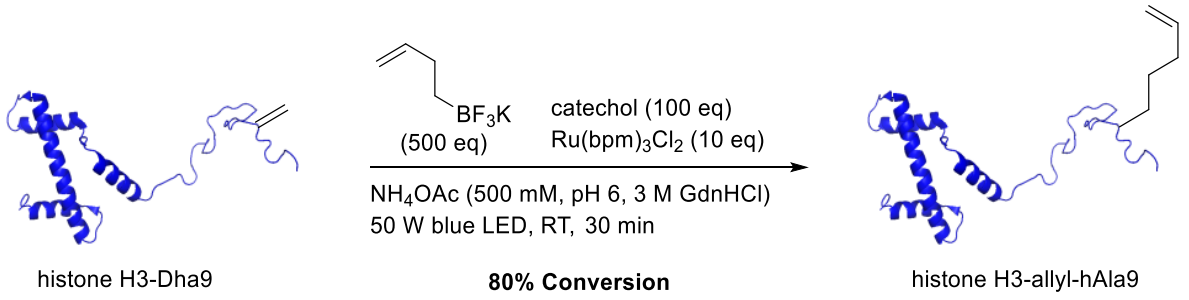
Calculated Dha mass: 15180 Da. Observed mass: 15180 Da





LC-MS/MS of histone H3-Cy-Ala9: ArgC digest leads to 97% sequence coverage (1% FDR). Q Exactive (HCD).

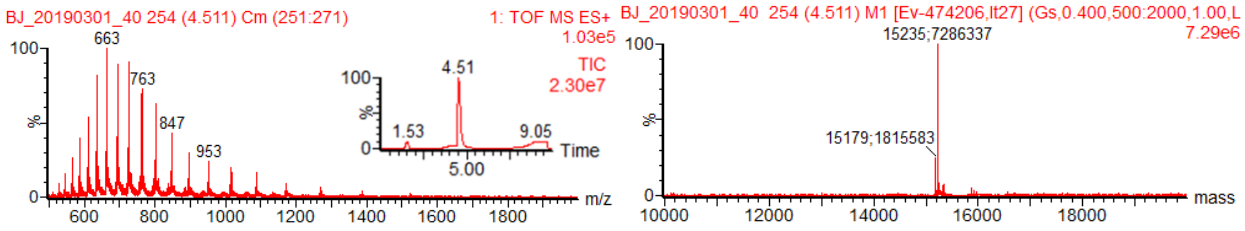
Installation of allylhomoalanine (allyl-hAla)

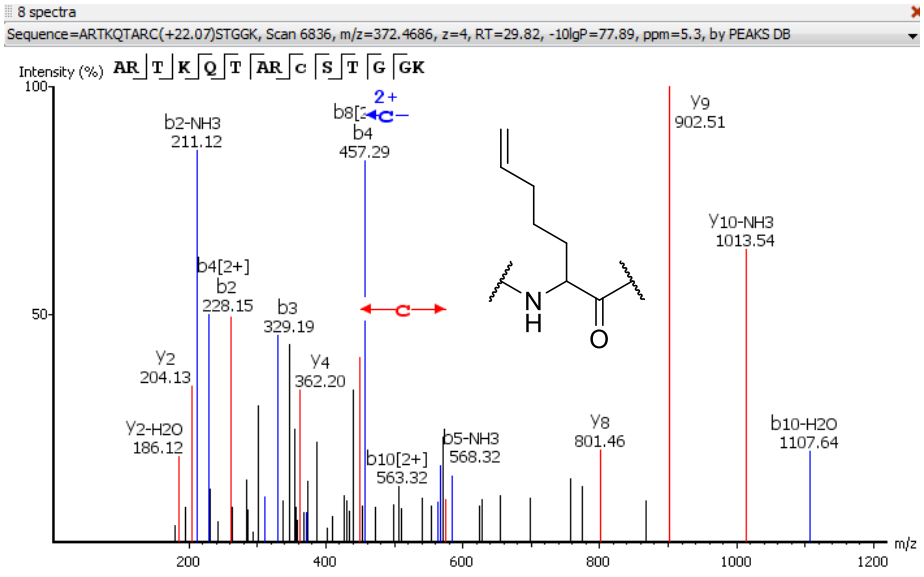


histone H3-Dha9 **80% Conversion** histone H3-allyl-hAla9

Calculated mass: 15236 Da. Observed mass: 15235 Da

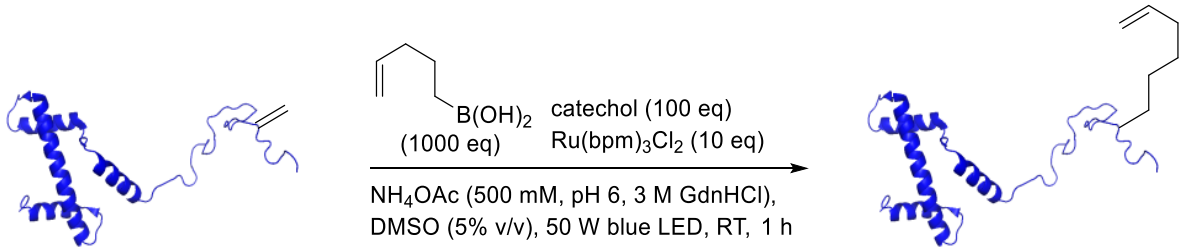
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS analysis of histone H3-allyl-hAla9: Q Exactive (HCD).

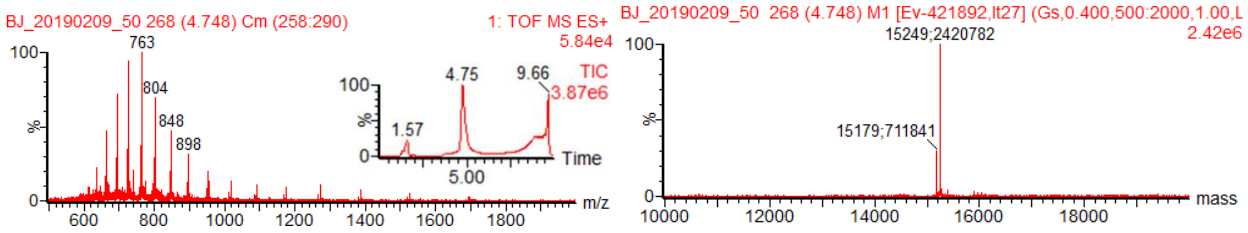
Installation of allylnorvaline (allyl-Nva)

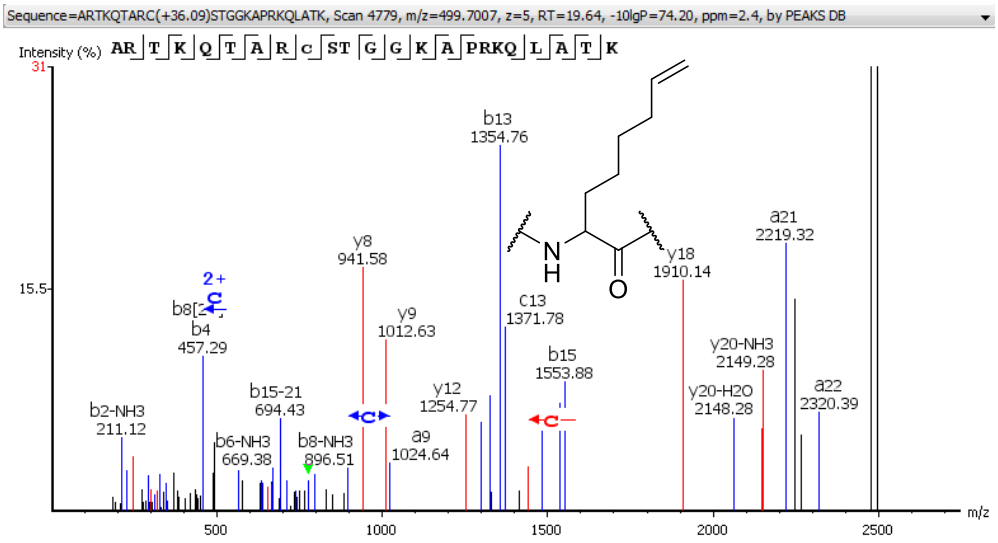


histone H3-Dha9 **77% Conversion** histone H3-allyl-Nva9

Calculated mass: 15250 Da. Observed mass: 15249 Da

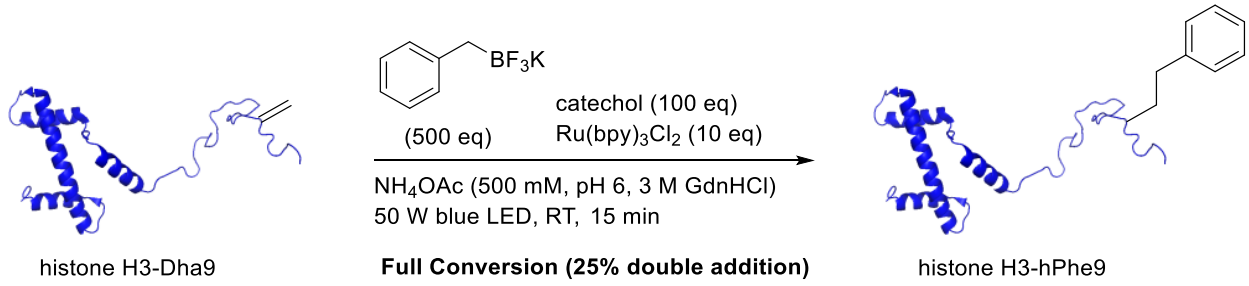
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS of histone H3-allyl-Nva9: ArgC digest leads to 96% sequence coverage (1% FDR). Q Exactive (HCD)

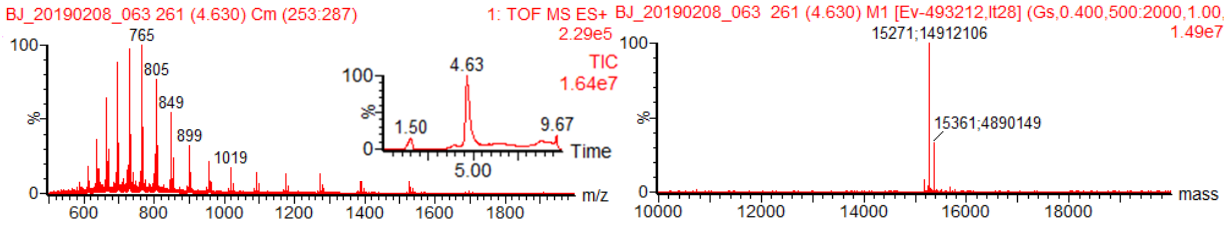
Installation of homophenylalanine (hPhe)

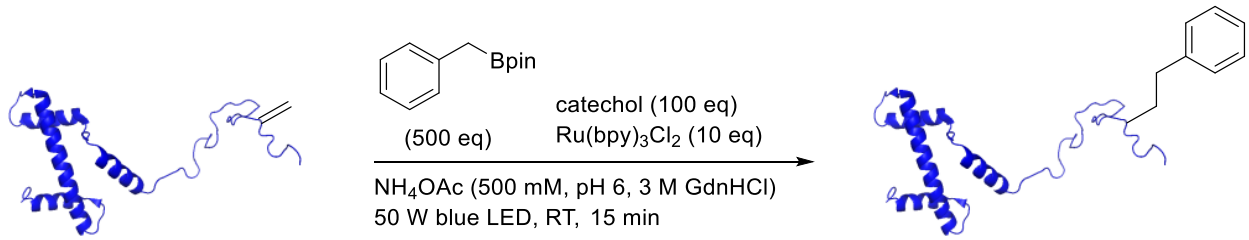


histone H3-Dha9 **Full Conversion (25% double addition)** histone H3-hPhe9

Calculated mass: 15272 Da. Observed mass: 15271 Da

Calculated double addition mass: 15362 Da. Observed mass: 15361 Da



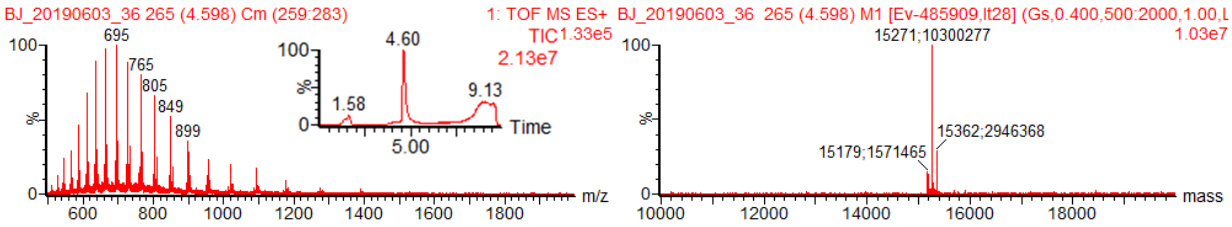


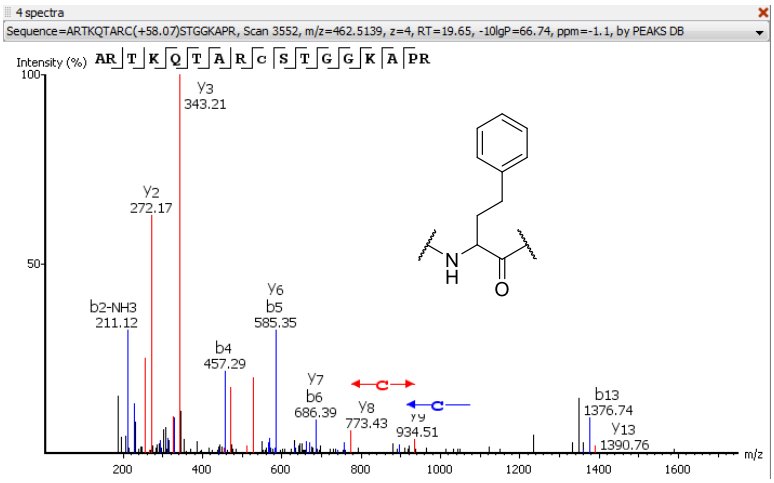
histone H3-Dha9 **89% Conversion (22% double addition)** histone H3-hPhe9

Calculated mass: 15272 Da. Observed mass: 15271 Da

Calculated double addition mass: 15362 Da. Observed mass: 15362 Da

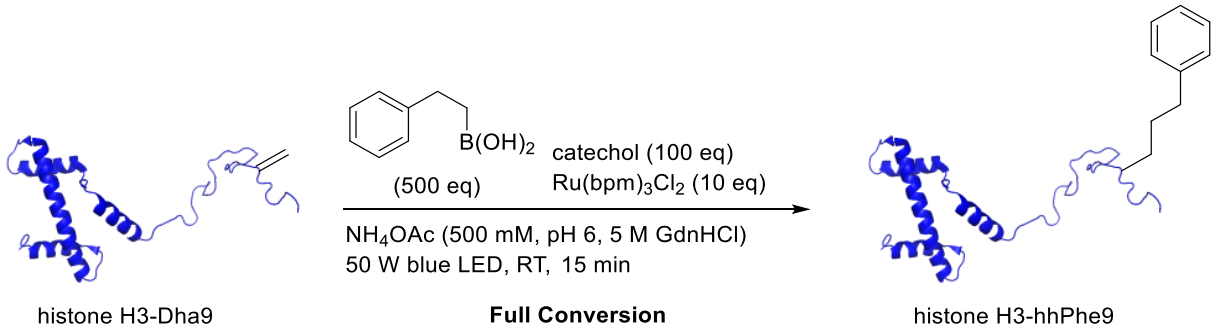
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS of histone H3-hPhe9: ArgC digest leads to 96% coverage (1% FDR). Q Exactive (HCD).

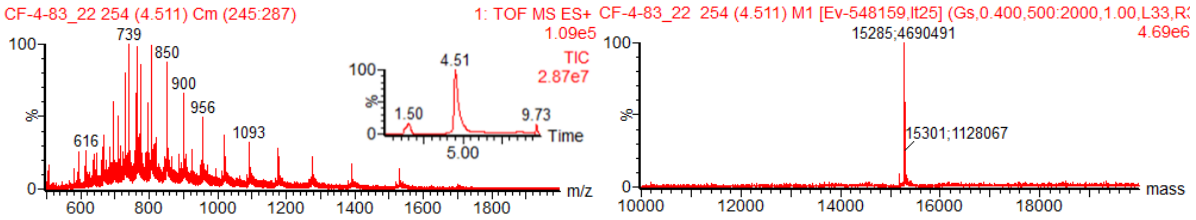
Installation of homohomophenylalanine (hhPhe)

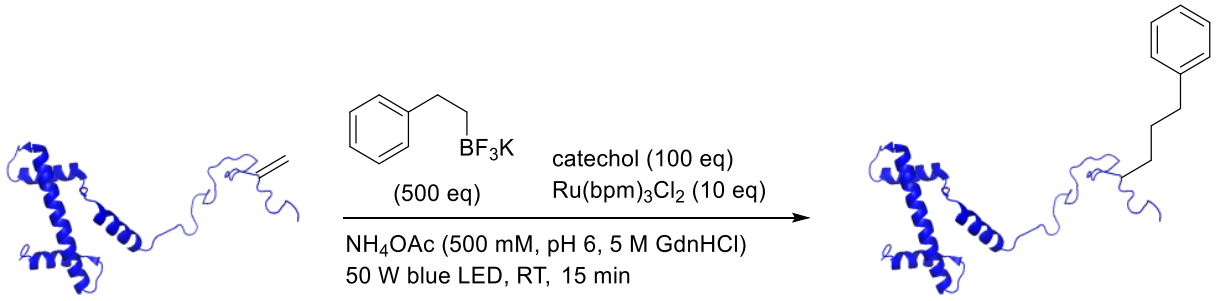


histone H3-Dha9 **Full Conversion** histone H3-hhPhe9

Calculated mass: 15286 Da. Observed mass: 15285 Da

Calculated Product + Met Ox mass: 15302 Da. Observed mass: 15301 Da

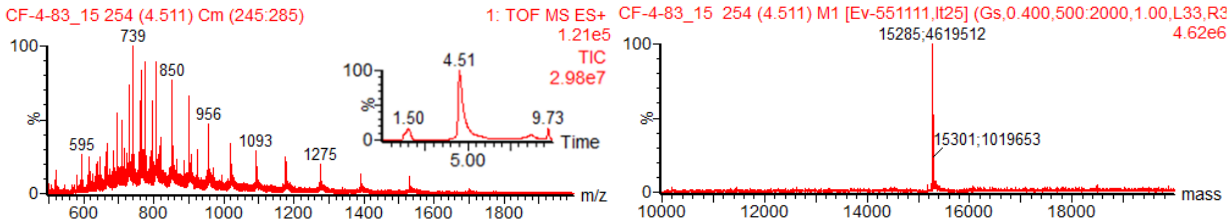


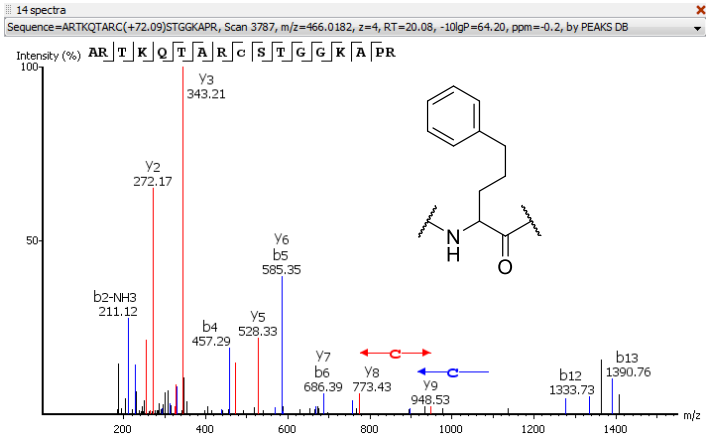


histone H3-Dha9 **Full Conversion** histone H3-hhPhe9

Calculated mass: 15286 Da. Observed mass: 15285 Da

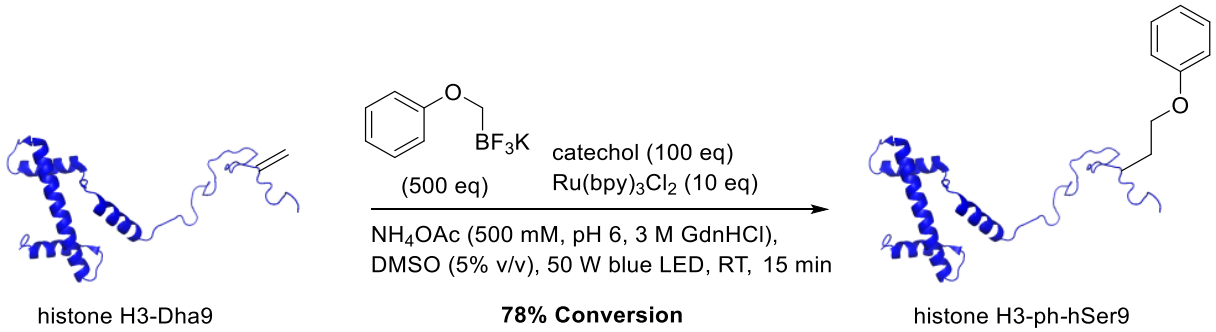
Calculated Product + Met Ox mass: 15302 Da. Observed mass: 15301 Da





LC-MS/MS of histone H3-hhPhe9: ArgC digest leads to 96% coverage (1% FDR). Q Exactive (HCD).

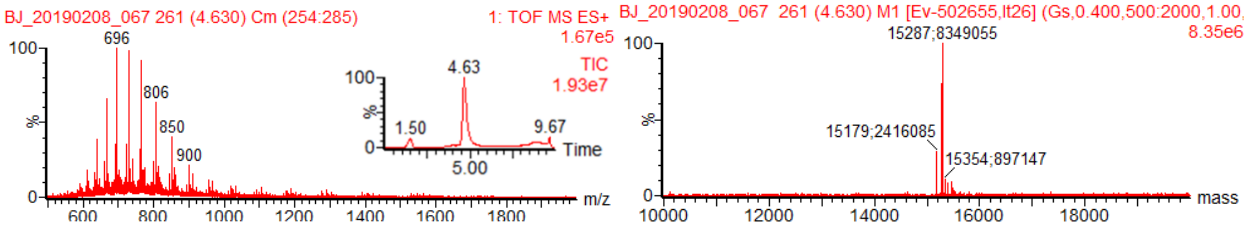
Installation of phenylhomoserine (ph-hSer)

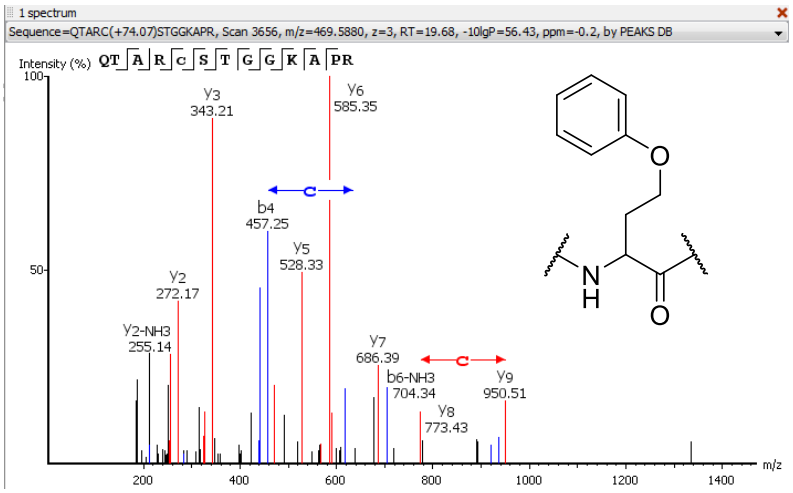


Calculated mass: 15288 Da. Observed mass: 15287 Da

Calculated Dha mass: 15180 Da. Observed mass: 15179 Da

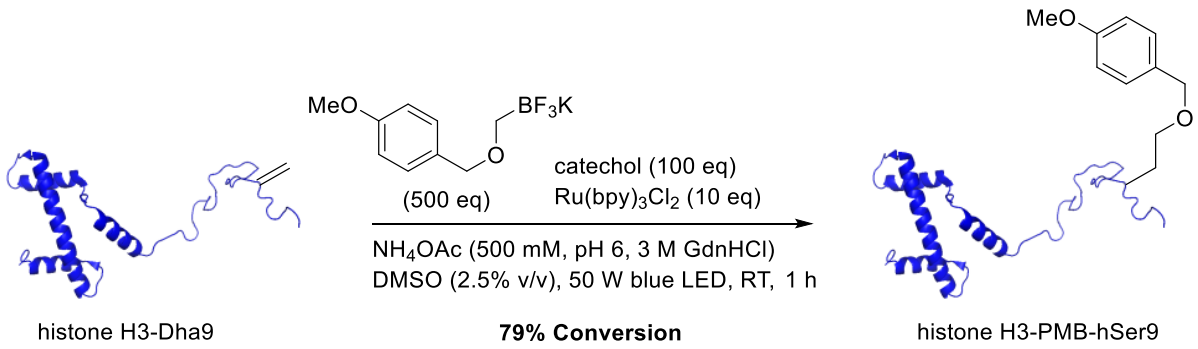
15354 Da mass unknown





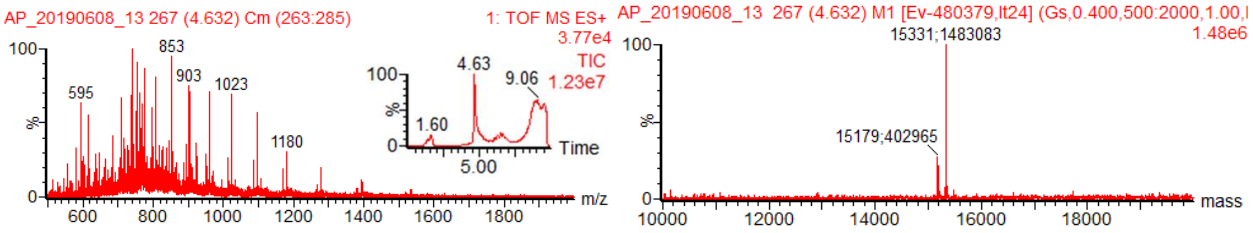
LC-MS/MS of histone H3-ph-hSer9: ArgC digest leads to 92% sequence coverage (1% FDR). Q Exactive (HCD).

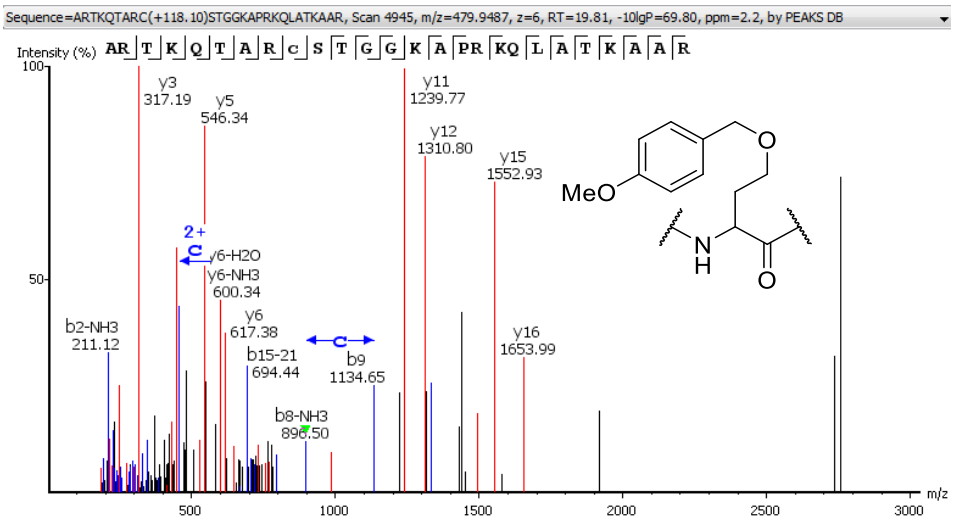
Installation of PMBhomoserine (PMB-hSer)



Calculated mass: 15332 Da. Observed mass: 15331 Da

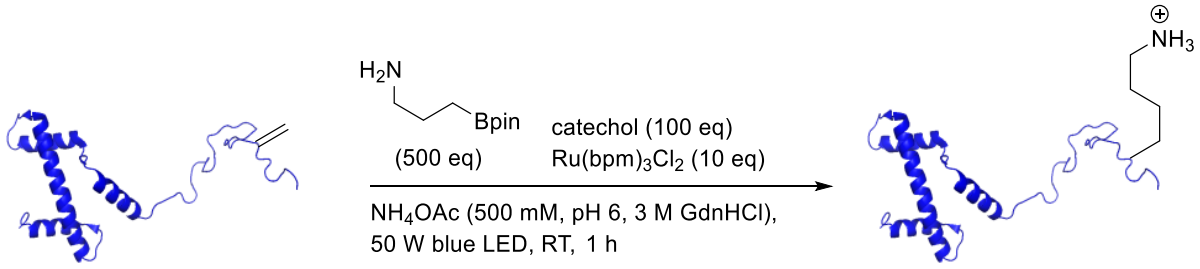
Calculated double addition mass: 15180 Da. Observed mass: 15179 Da





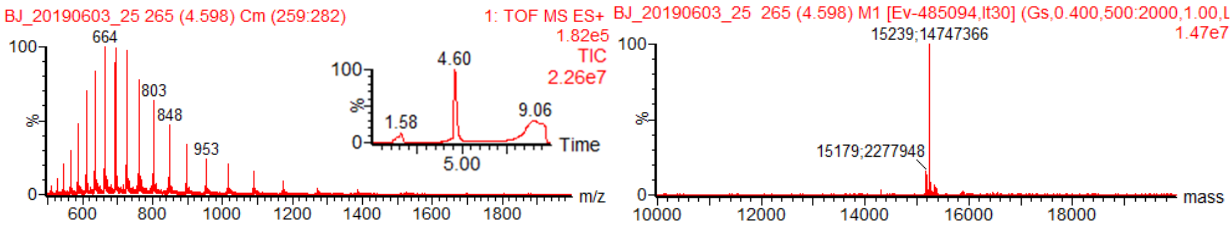
LC-MS/MS of histone H3-PMB-hSer9: ArgC digest leads to 99% sequence coverage (1% FDR). Q Exactive (HCD).

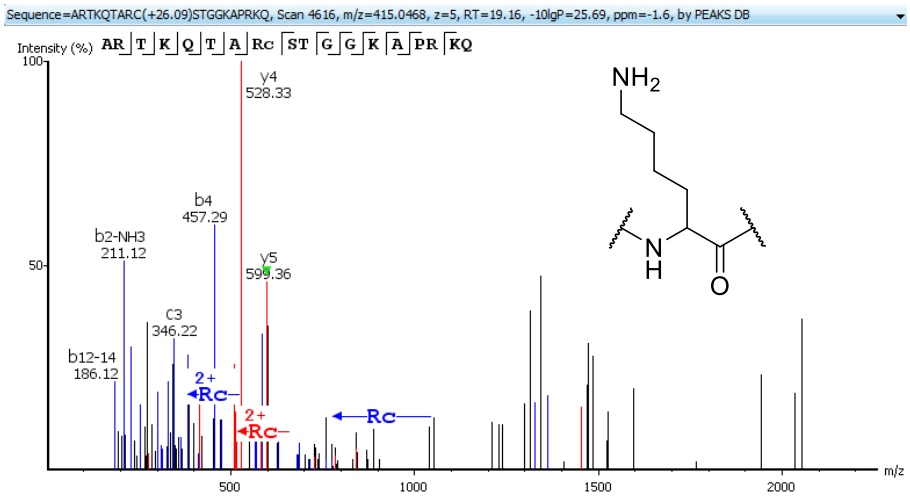
Installation of lysine (Lys)



Calculated mass: 15240 Da. Observed mass: 15239 Da

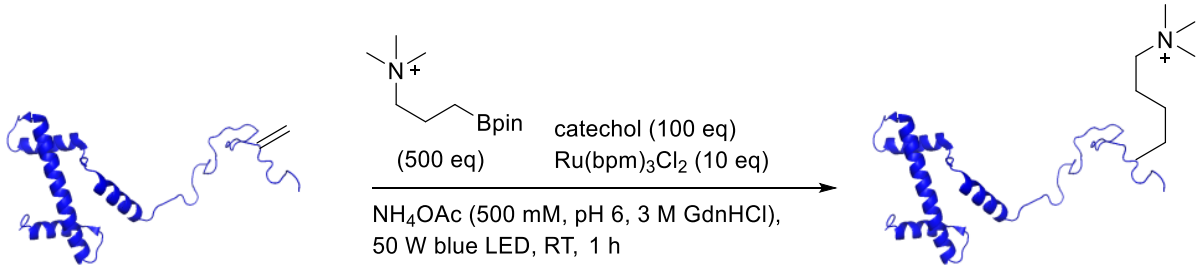
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS of histone H3-Lys9: Elastase leads to 93% sequence coverage (1% FDR). Q Exactive (HCD)

Installation of trimethyllysine (Me_3Lys)

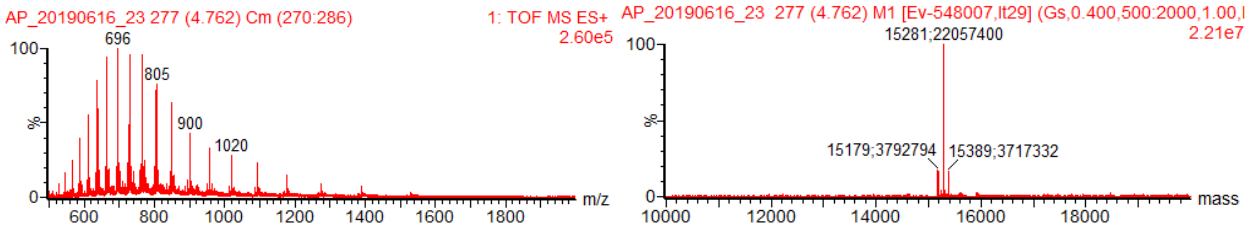


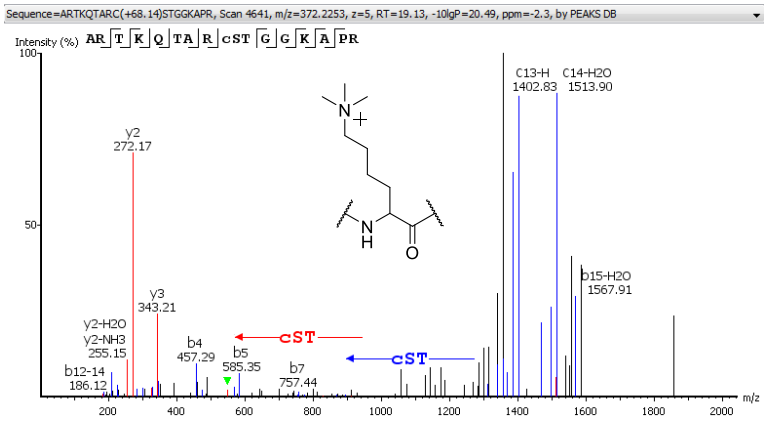
histone H3-Dha9 **87% Conversion (14% catechol adduct)** histone H3-Me_3Lys9

Calculated mass: 15282 Da. Observed mass: 15281 Da

Calculated catechol adduct: 15390 Da. Observed mass: 15389 Da

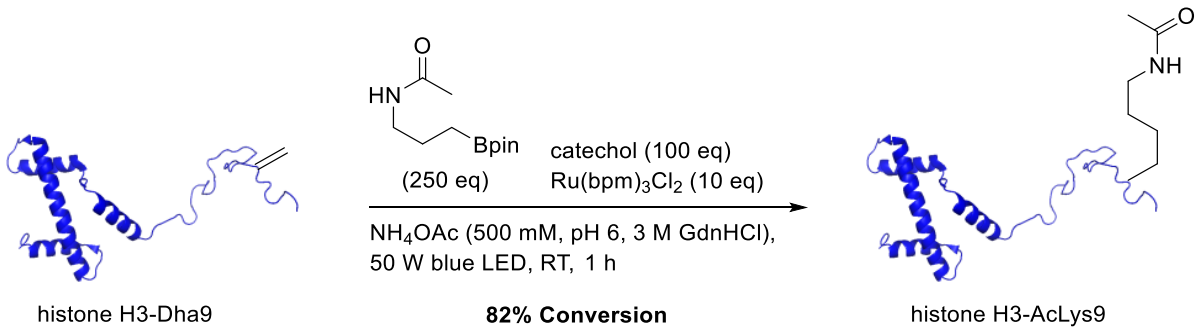
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS of histone H3-Me_3Lys9:. Elastase digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

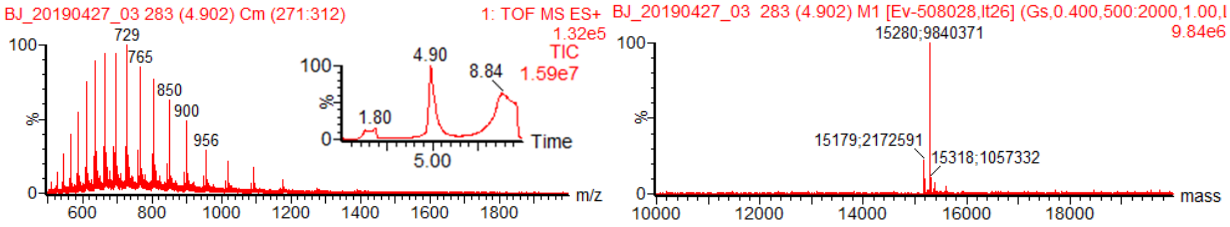
Installation of acetyllysine (AcLys)

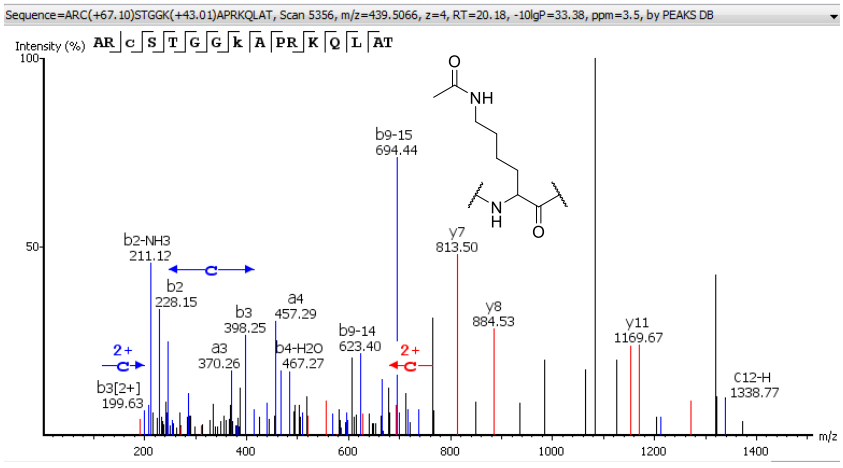


Calculated mass: 15281 Da. Observed mass: 15280 Da

Calculated Dha mass: 15180 Da. Observed mass: 15179 Da

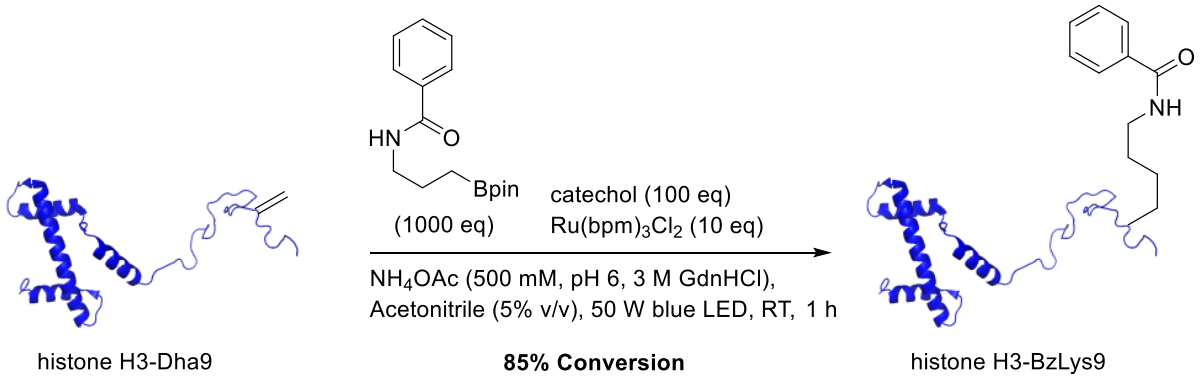
15318 Da mass unknown





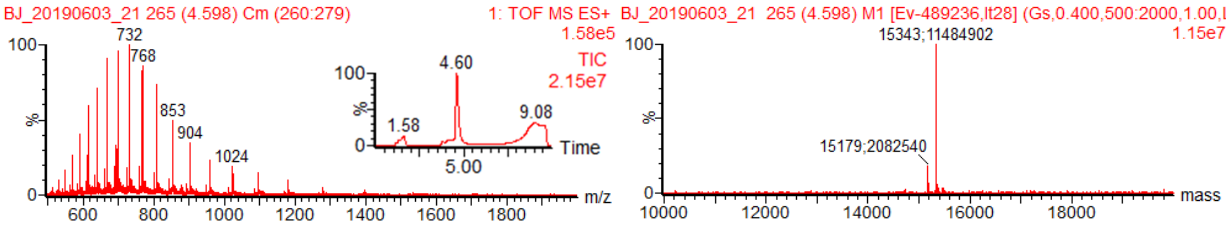
LC-MS/MS of histone H3-AcLys9: Elastase digest leads to 93% sequence coverage (1% FDR). Q Exactive (HCD)

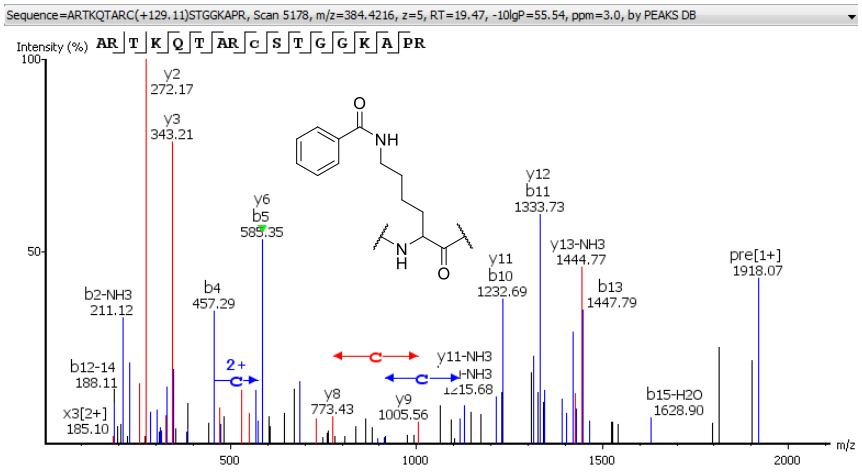
Installation of benzoyllysine (BzLys)



Calculated mass: 15343 Da. Observed mass: 15343 Da

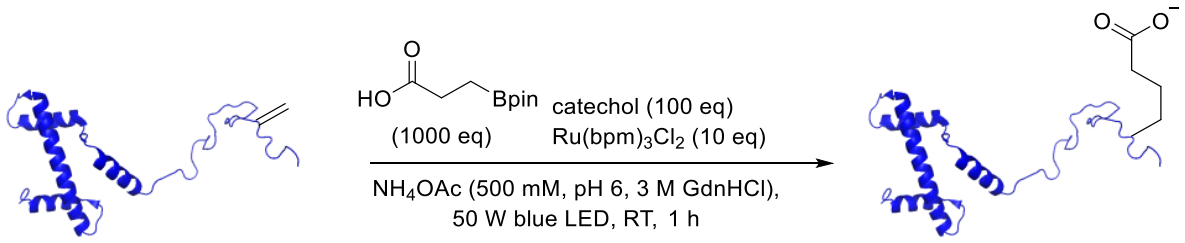
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS of histone H3-BzLys9: ArgC digest 99 % sequence coverage (1% FDR). Q Exactive (HCD)

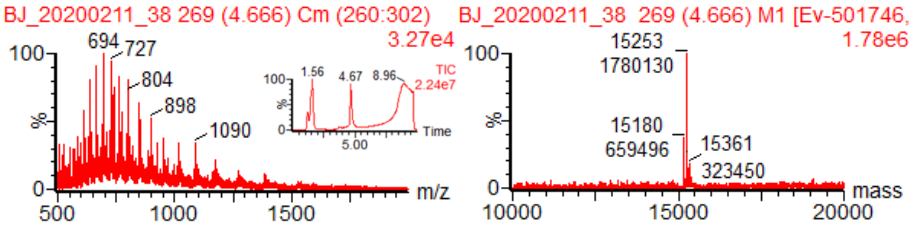
Installation of homoglutamic acid (hGlu)



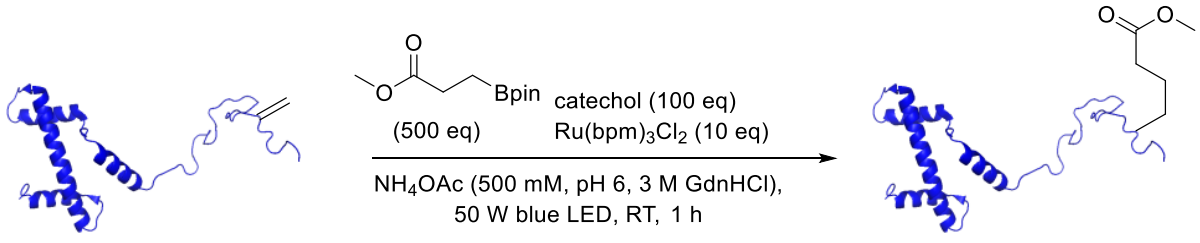
histone H3-Dha9 **76% Conversion (15% catechol adduct)** histone H3-hGlu9

Calculated mass: 15253 Da. Observed mass: 15253 Da

Calculated Dha mass: 15180 Da. Observed Dha mass: 15180 Da



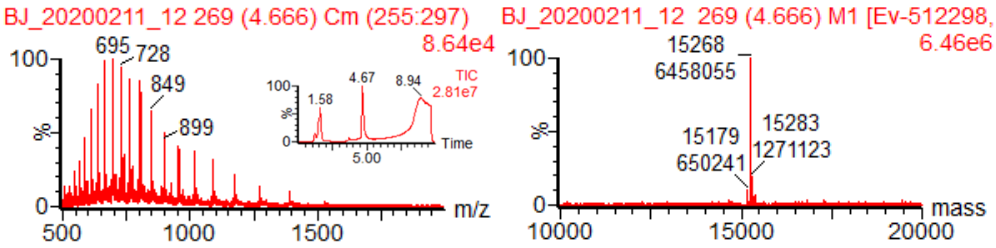
Installation of methylhomoglutamic acid (Me-hGlu)



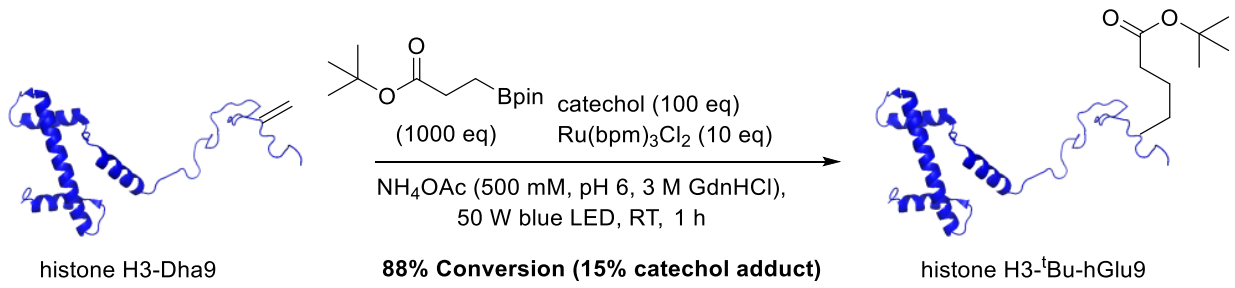
histone H3-Dha9 **92% Conversion** histone H3-Me-hGlu9

Calculated mass: 15268 Da. Observed mass: 15268 Da

Calculated Dha mass: 15180 Da. Observed Dha mass: 15179 Da



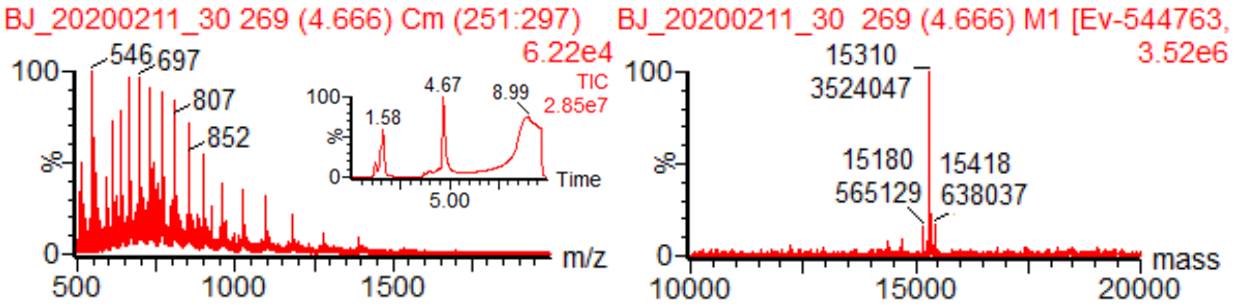
Installation of tertbutylhomoglutamic acid (^{t}Bu-hGlu)



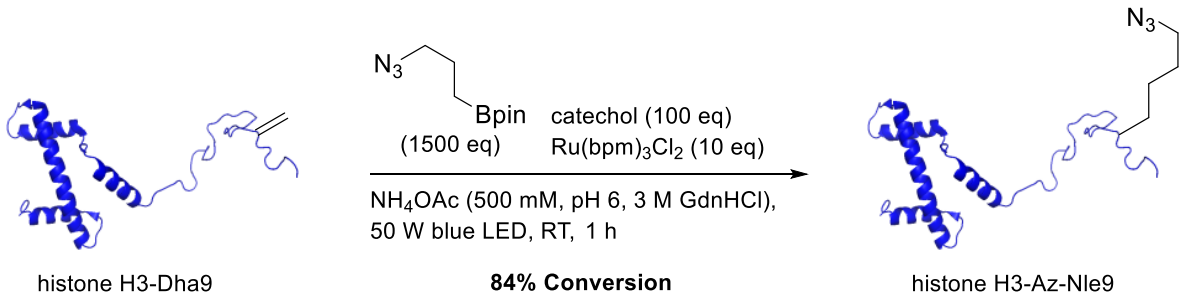
Calculated mass: 15310 Da. Observed mass: 15310 Da

Calculated Dha mass: 15180 Da. Observed Dha mass: 15180 Da

Calculated catechol adduct: 15418 Da. Observed catechol adduct: 15418 Da



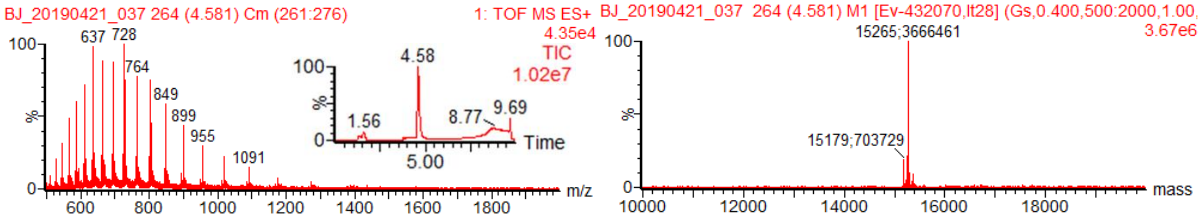
Installation of azidonorleucine (Az-Nle)

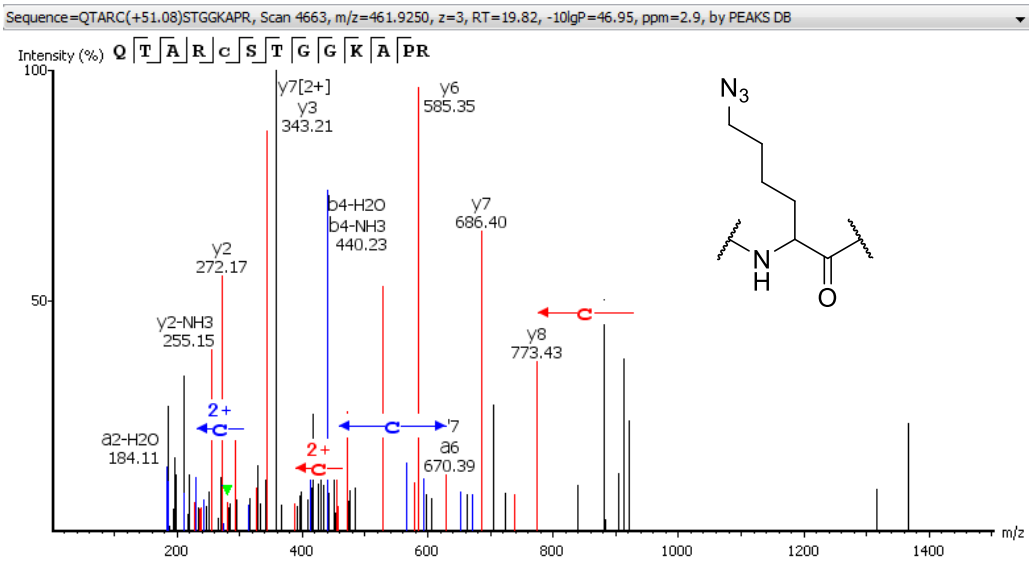


histone H3-Dha9 **84% Conversion** histone H3-Az-Nle9

Calculated mass: 15265 Da. Observed mass: 15265 Da

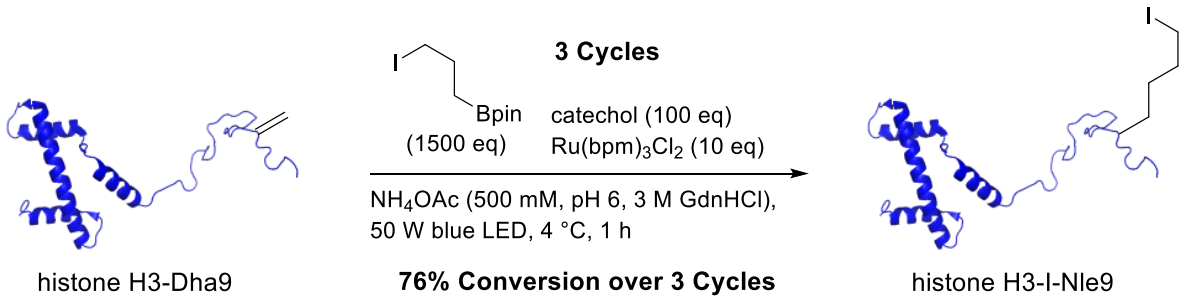
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da





LC-MS/MS of histone H3-Az-Nle9: ArgC digest leads to 93% sequence coverage (1% FDR). Q Exactive (HCD)

Installation of iodonorleucine (I-Nle)

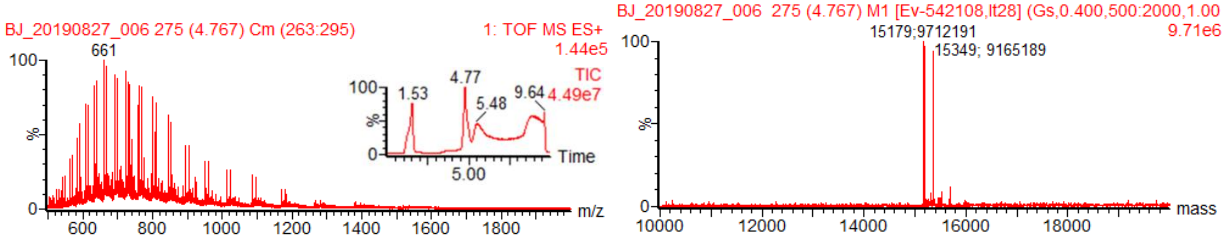


histone H3-Dha9 **76% Conversion over 3 Cycles** histone H3-I-Nle9

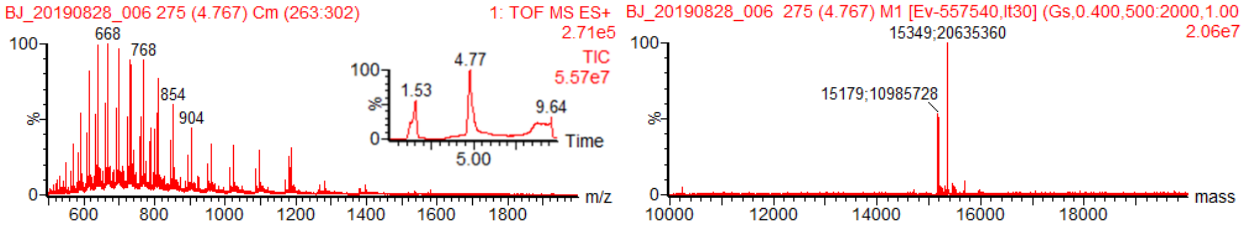
Calculated mass: 15350 Da. Observed mass: 15349 Da

Calculated Dha mass: 15180 Da. Observed mass: 15179 Da

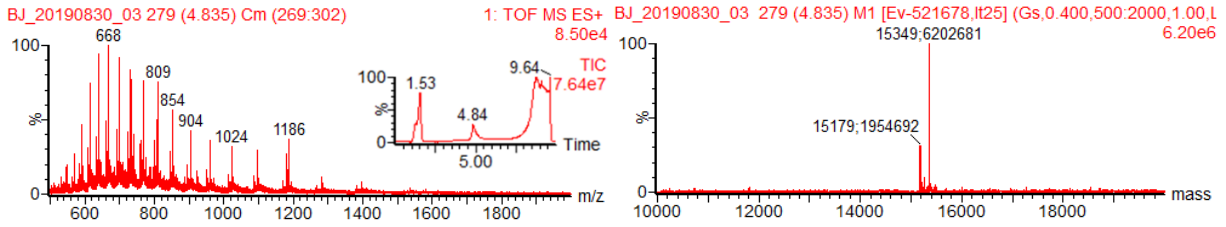
1st reaction cycle

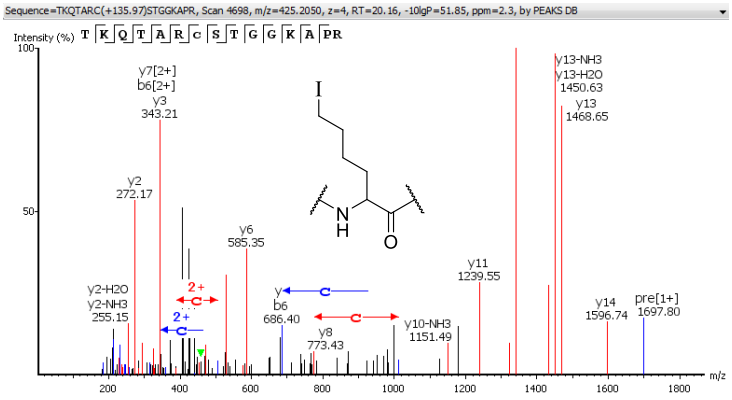


2nd reaction cycle



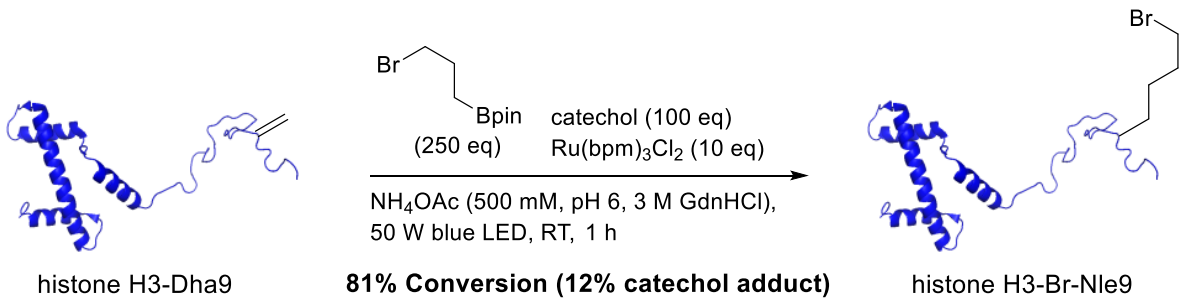
3rd reaction cycle





LC-MS/MS of histone H3-I-Nle9: ArgC digest leads to 99% sequence coverage (1% FDR). Q Exactive (HCD).

Installation of bromonorleucine (Br-Nle)

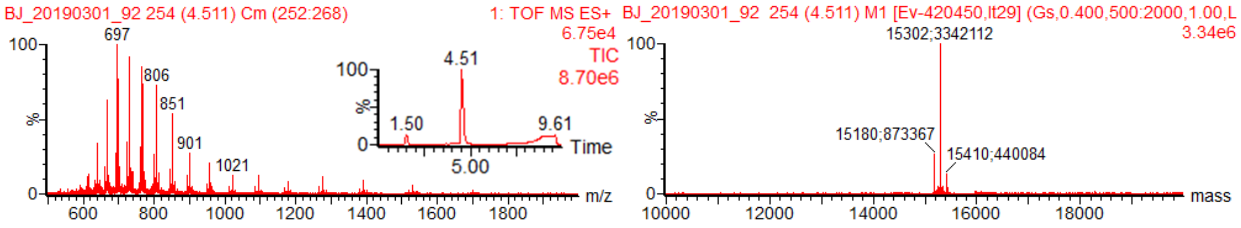


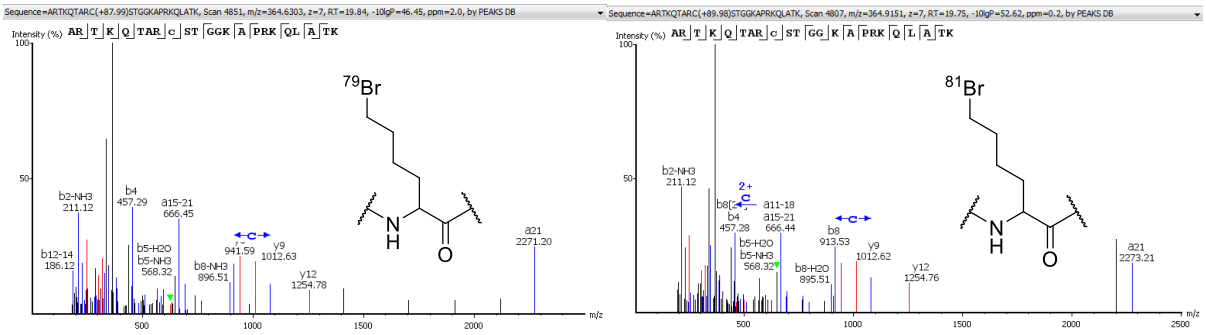
histone H3-Dha9 **81% Conversion (12% catechol adduct)** histone H3-Br-Nle9

Calculated mass: 15302 Da. Observed mass: 15302 Da

Calculated Dha mass: 15180 Da. Observed mass: 15180 Da

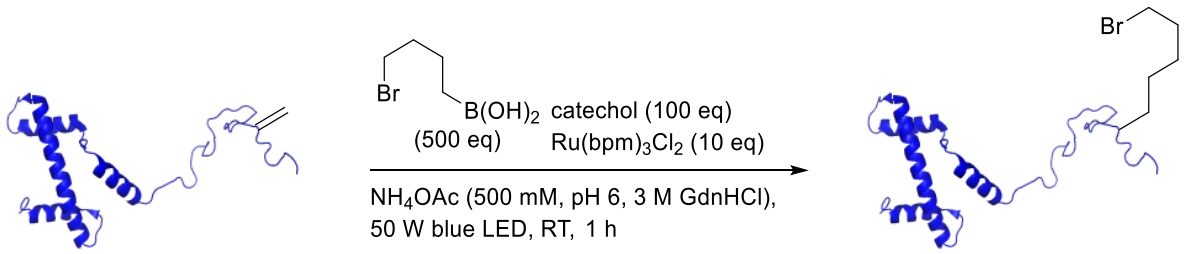
Calculated catechol mass: 15410 Da. Observed mass: 15410 Da





LC-MS/MS of histone H3-Br-Nle9: ArgC digest leads to 99% sequence coverage (1% FDR). Q Exactive (HCD).

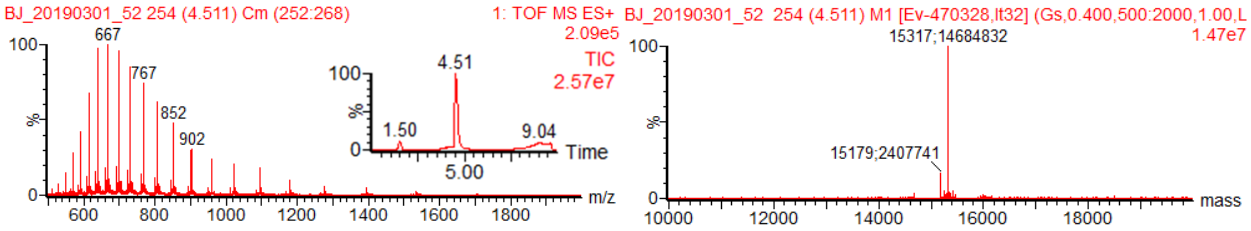
Installation of bromohomonorleucine (Br-hNle)

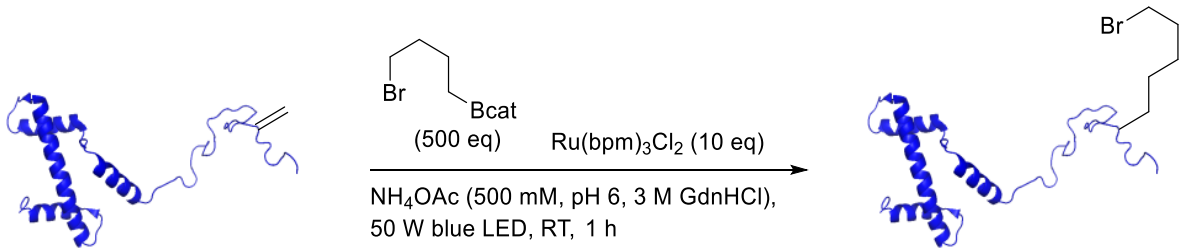


histone H3-Dha9 **86% Conversion** histone H3-Br-hNle9

Calculated mass: 15316 Da. Observed mass: 15317 Da

Calculated Dha mass: 15180 Da. Observed mass: 15179 Da

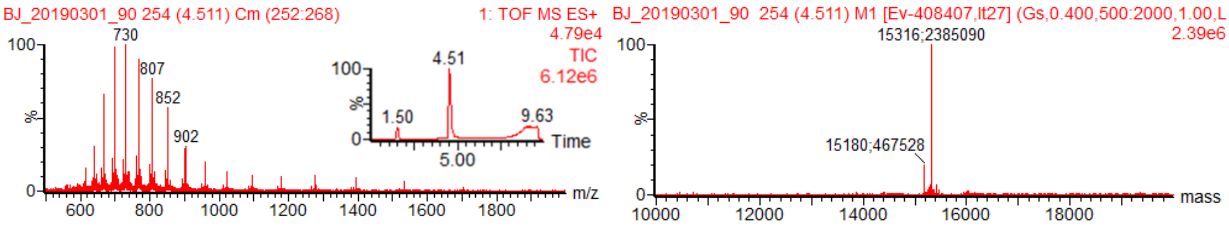


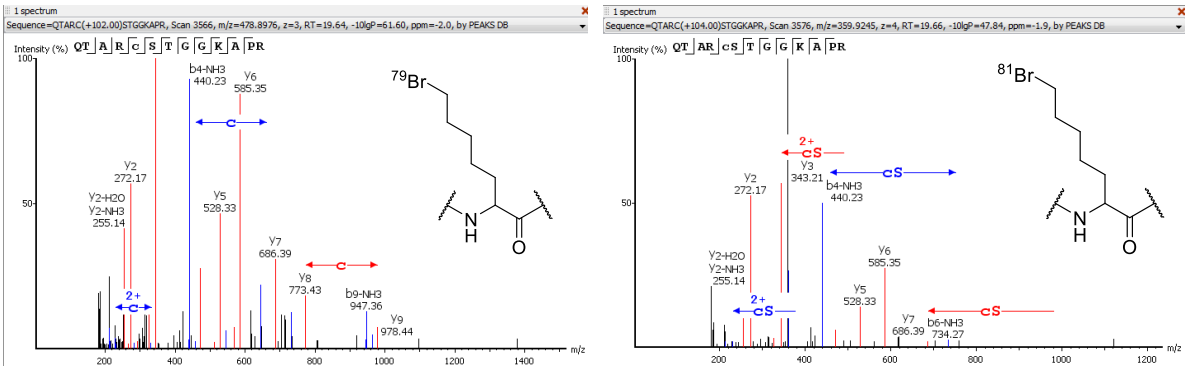


histone H3-Dha9 **84% Conversion** histone H3-Br-hNle9

Calculated mass: 15316 Da. Observed mass: 15316 Da

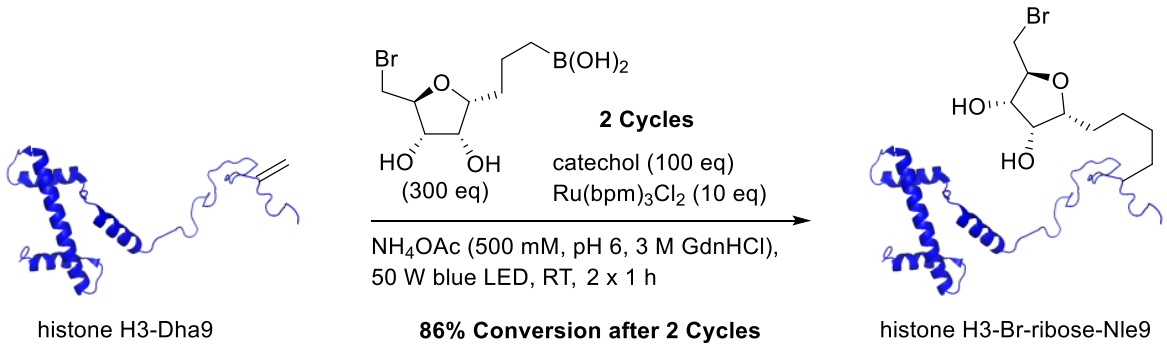
Calculated Dha mass: 15180 Da. Observed mass: 15180 Da





LC-MS/MS of histone H3-Br-hNle9: ArgC digest leads to 95% sequence coverage (1% FDR). Q Exactive (HCD).

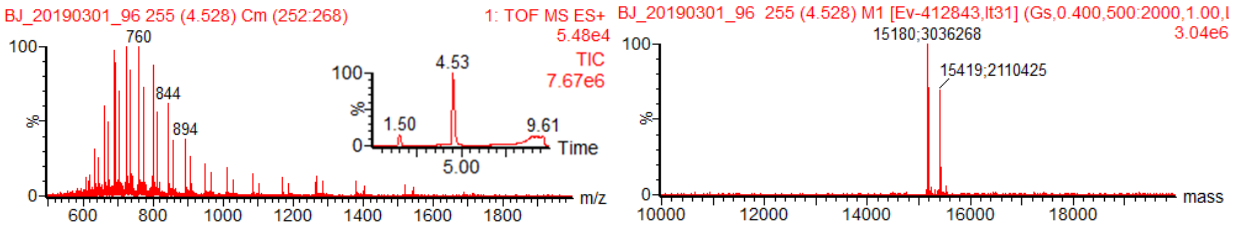
Installation of bromoribosenorleucine (Br-ribose-Nle)



histone H3-Dha9 **86% Conversion after 2 Cycles** histone H3-Br-ribose-Nle9

Calculated mass: 15418 Da. Observed mass: 15419 Da

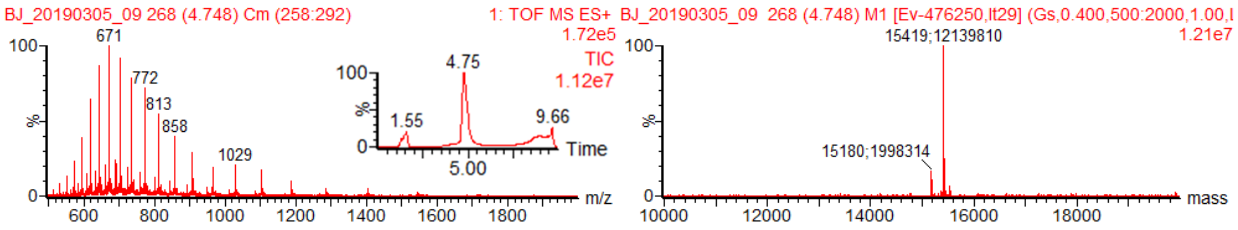
Calculated Dha mass: 15180 Da. Observed mass: 15180 Da

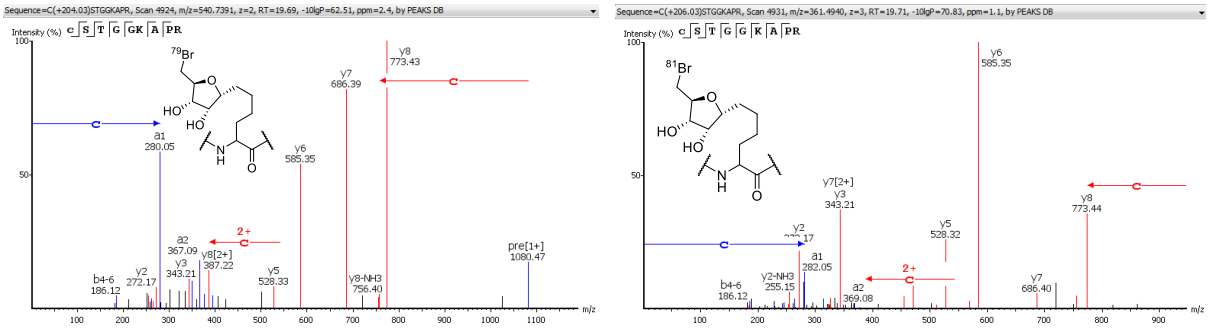


ESI-MS for the modified histone H3 after an additional 1 h reaction is shown below:

Calculated mass: 15418 Da. Observed mass: 15419 Da

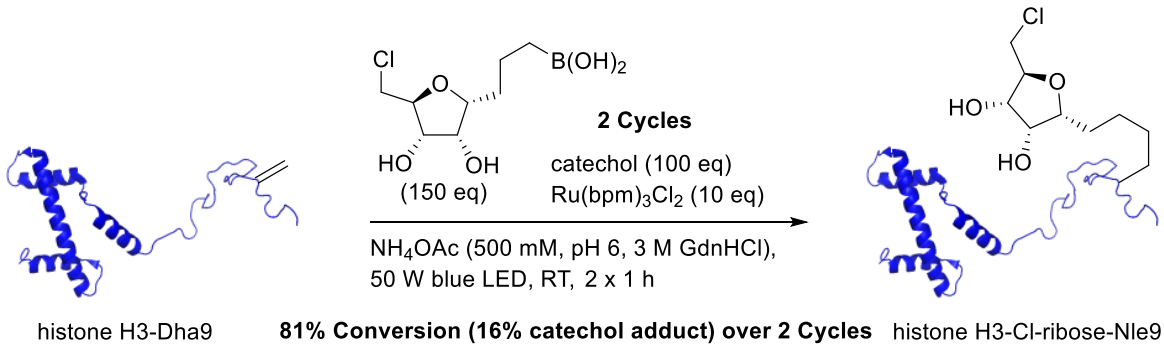
Calculated Dha mass: 15180 Da. Observed mass: 15180 Da





LC-MS/MS of histone H3-Br-ribose-Nle9: ArgC digest leads to 99% sequence coverage (1% FDR). Q Exactive (HCD).

Installation of chlororibosenorleucine (Cl-ribose-Nle)

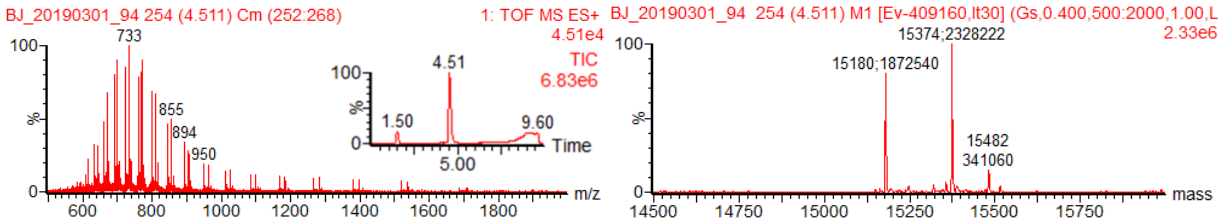


histone H3-Dha9 **81% Conversion (16% catechol adduct) over 2 Cycles** histone H3-Cl-ribose-Nle9

Calculated mass: 15374 Da. Observed mass: 15374 Da

Calculated Dha mass: 15180 Da. Observed mass: 15180 Da

Calculated catechol mass: 15482 Da. Observed catechol mass: 15482 Da

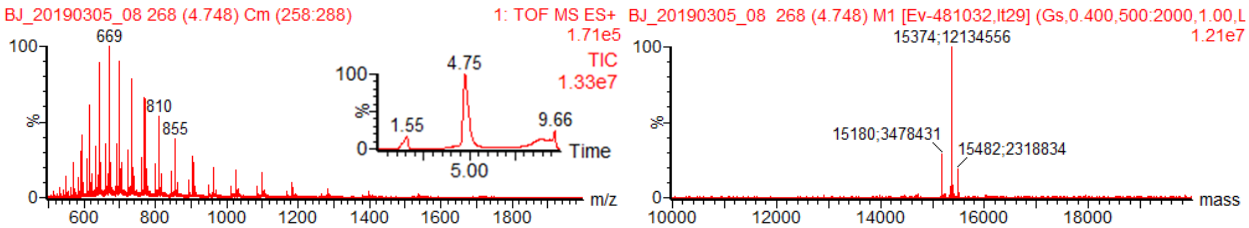


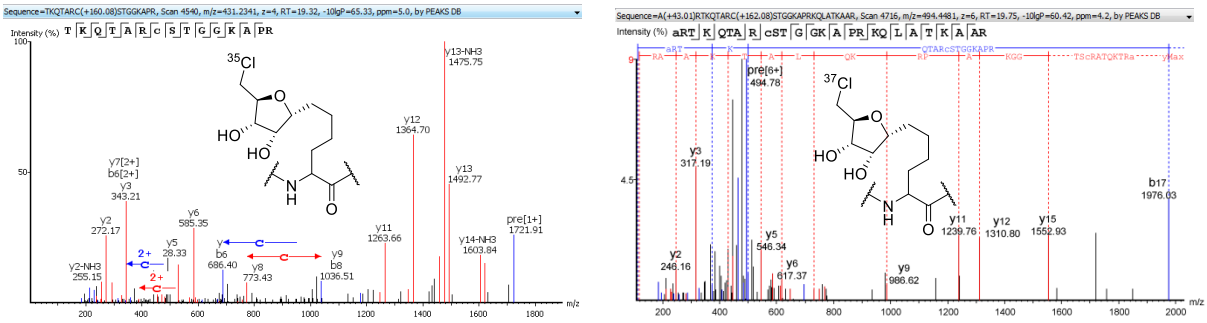
ESI-MS for the modified histone H3 after an additional 1 h reaction is shown below:

Calculated mass: 15374 Da. Observed mass: 15374 Da

Calculated Dha mass: 15180 Da. Observed mass: 15180 Da

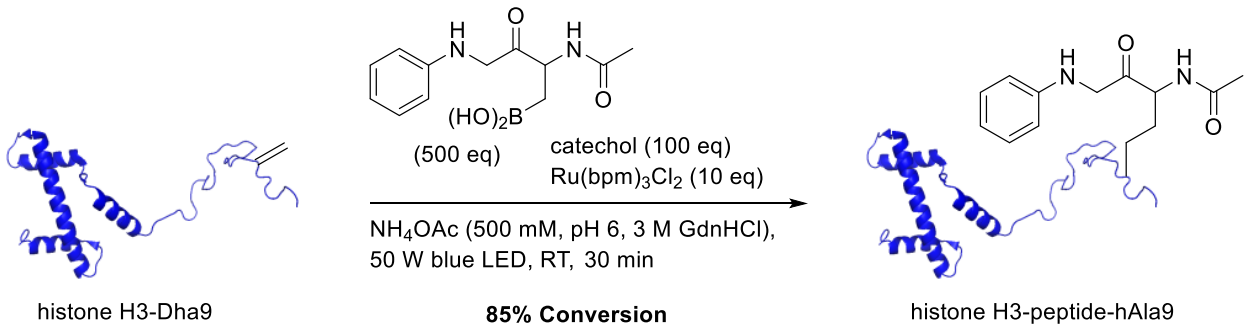
Calculated catechol mass: 15482 Da. Observed catechol mass: 15482 Da





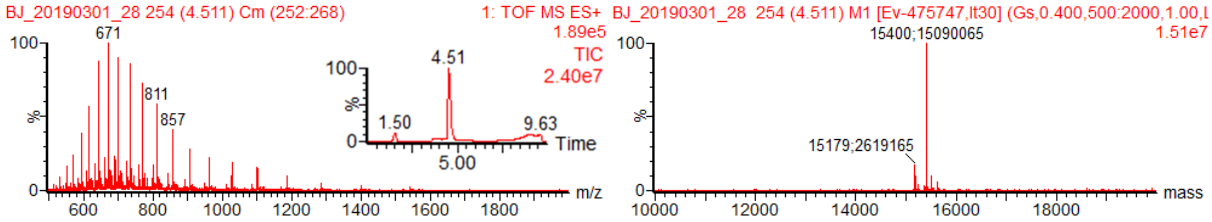
LC-MS/MS of histone H3-Cl-ribose-Nle9: ArgC digest leads to 86% sequence coverage (1% FDR). Q Exactive (HCD).

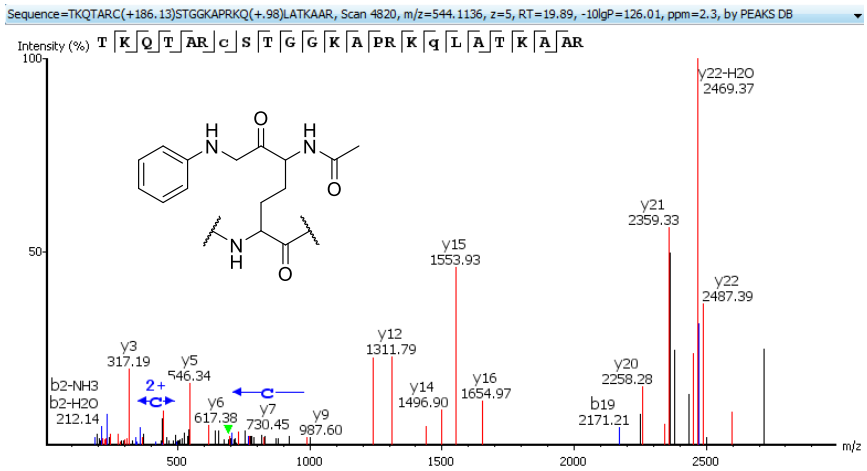
Installation of peptidehomoalanine (peptide-hAla)



Calculated mass: 15400 Da. Observed mass: 15400 Da

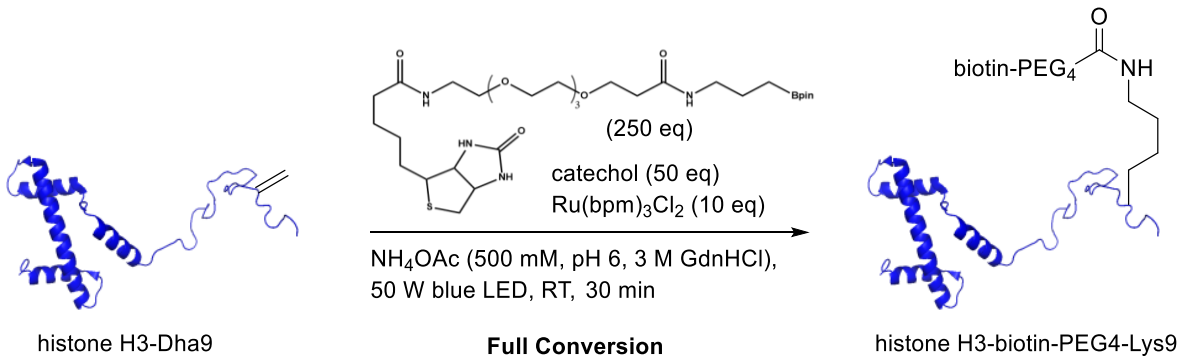
Calculated Dha mass: 15180 Da. Observed mass: 15179 Da



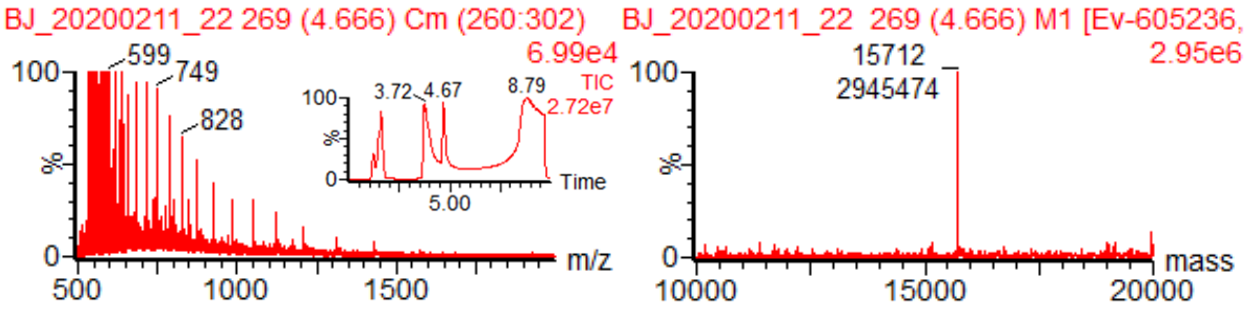


LC-MS/MS of histone H3-peptide-hAla9: ArgC digest leads to 99% sequence coverage (1% FDR). Q Exactive HCD.

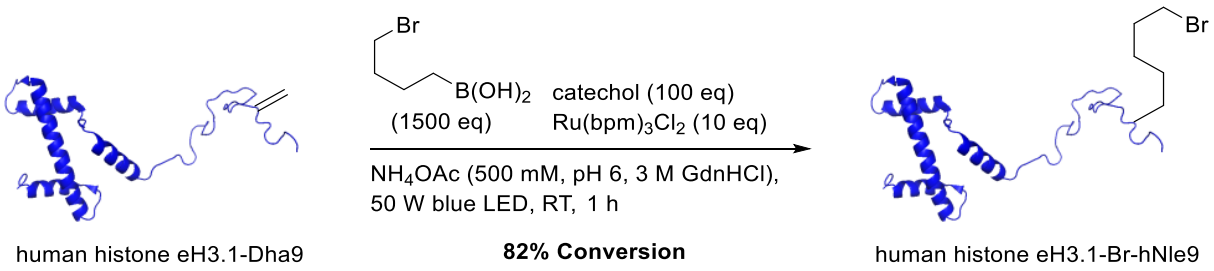
Installation of biotinPEG4lysine (biotin-PEG4-Lys)



Calculated mass: 15712 Da. Observed mass: 15712 Da



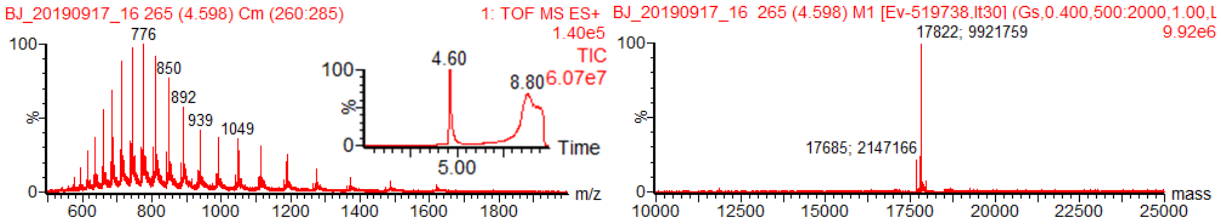
Large scale (5 mg) installation of bromohomonorleucine (Br-hNle)

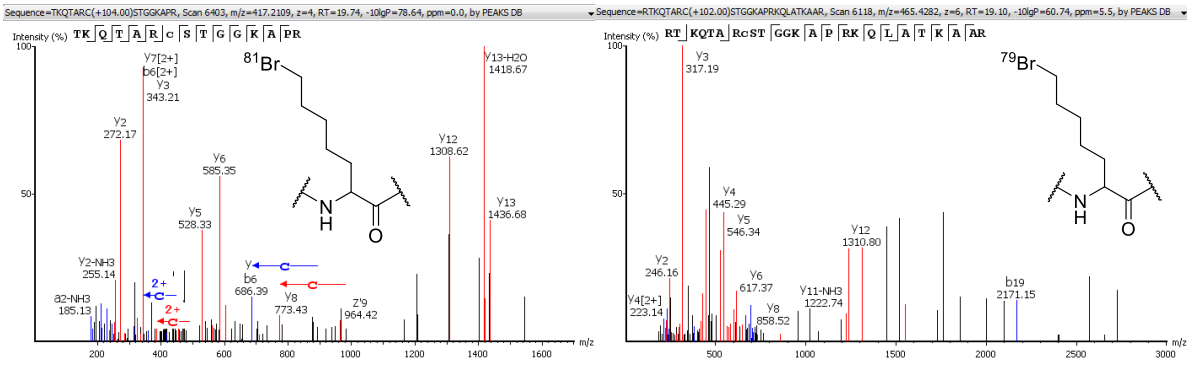


human histone eH3.1-Dha9 **82% Conversion** human histone eH3.1-Br-hNle9

Calculated mass: 17822 Da. Observed mass: 17822 Da

Calculated Dha mass: 17686 Da. Observed mass: 17685 Da

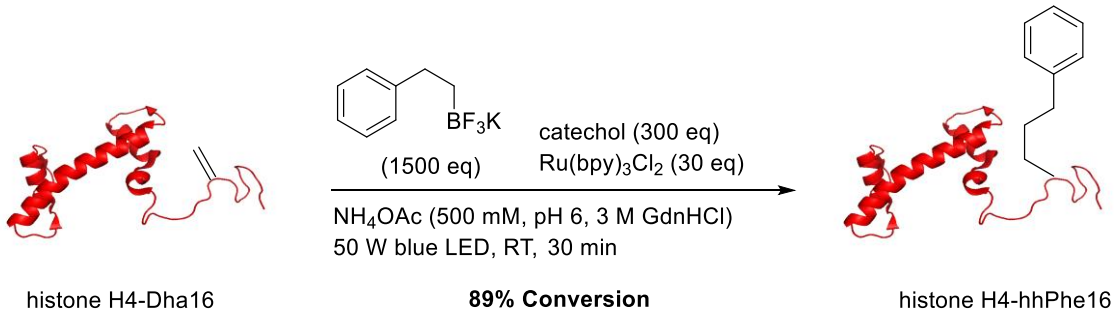




LC-MS/MS for histone eH3-Br-hNle9. ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

Protein scope for light-driven radical Dha modification

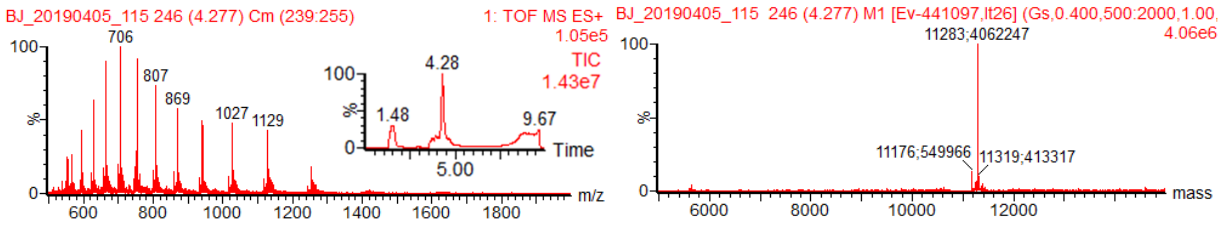
Installation of hhPhe on histone H4

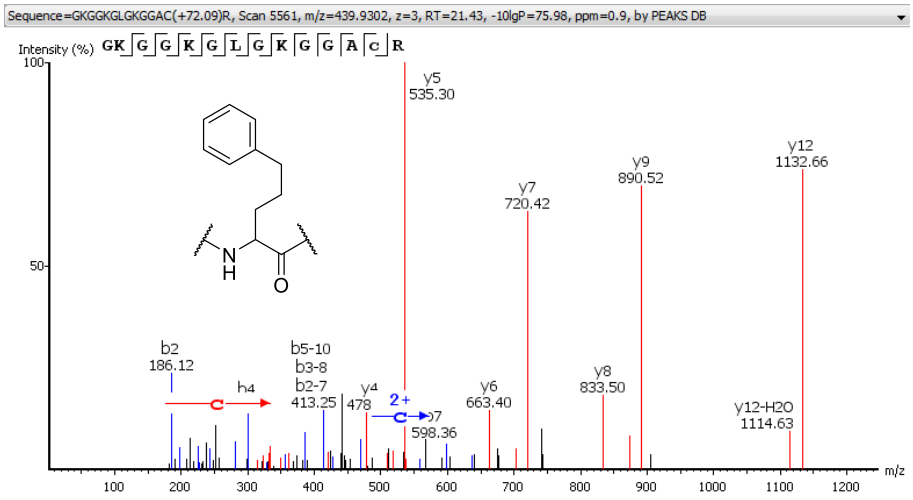


histone H4-Dha16 **89% Conversion** histone H4-hhPhe16

Calculated mass: 11283 Da. Observed mass: 11283 Da (11298 Da + Met Ox)

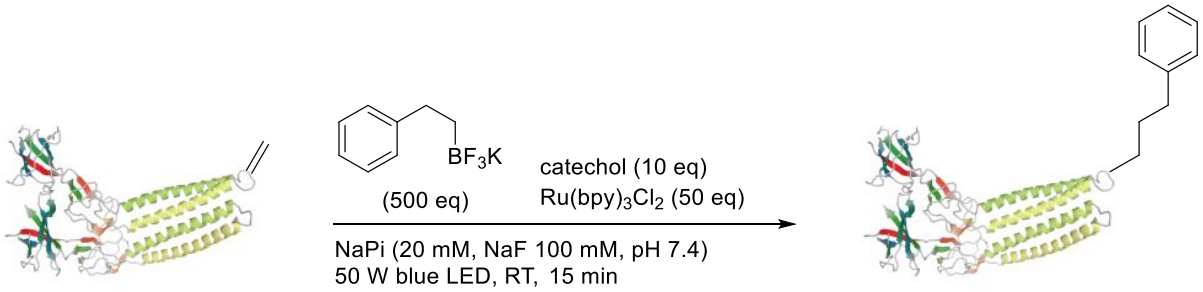
Calculated Dha mass: 11177 Da. Observed mass: 11176 Da





LC-MS/MS of histone H4-hhPhe16: 99% sequence coverage with ArgC digest (1% FDR). Q Exactive (HCD).

Installation of hhPhe on AcrA

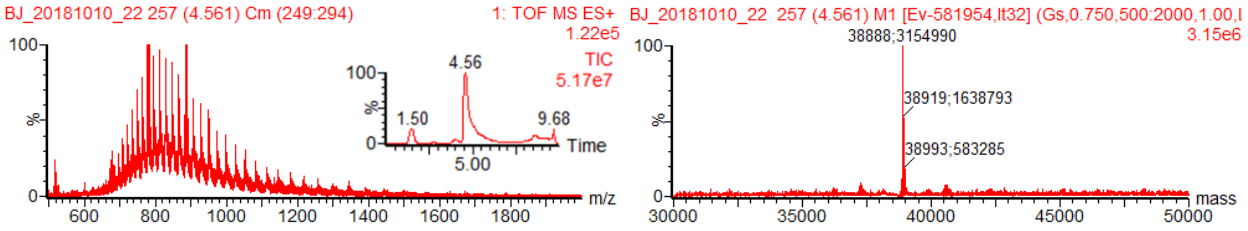


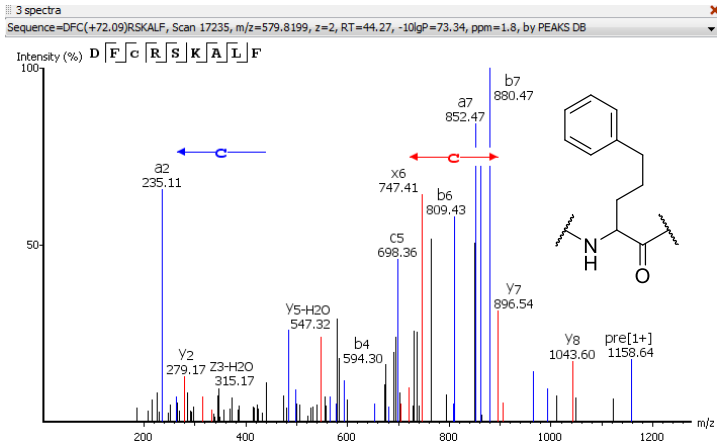
AcrA-Dha123 **Full Conversion (16% double addition)** AcrA-hhPhe123

Calculated mass: 38889 Da. Observed mass: 38888 Da (+ double addition 38992 Da)

Calculated initial hydrolyzed MDBP from Dha formation mass: 38915 Da

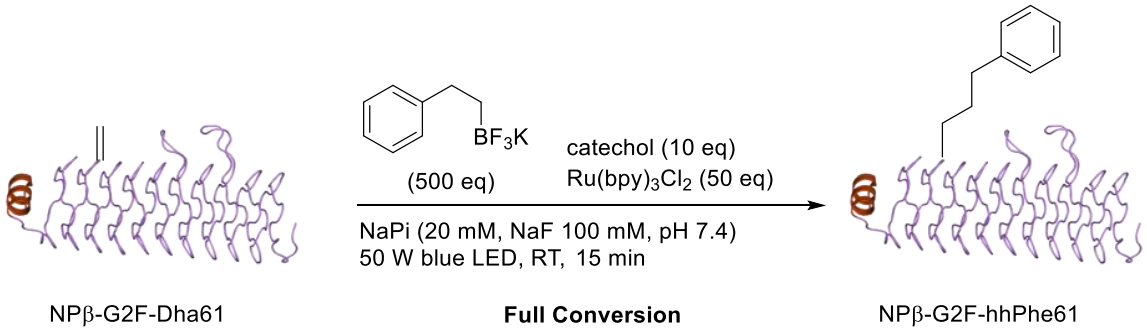
Observed mass: 38919 Da





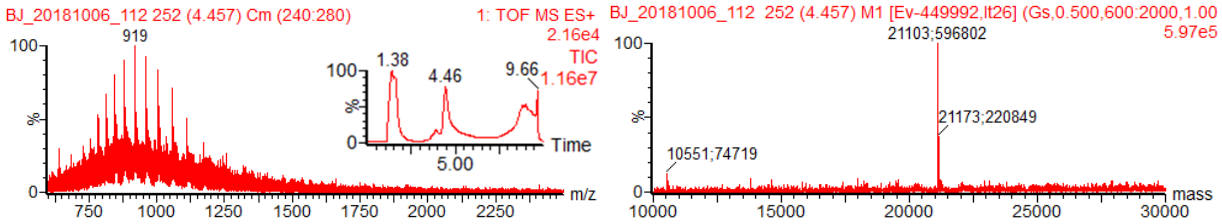
LC-MS/MS analysis of AcrA-hhPhe123: AspN digest leads to 99% sequence coverage (1% FDR). Q Exactive (HCD).

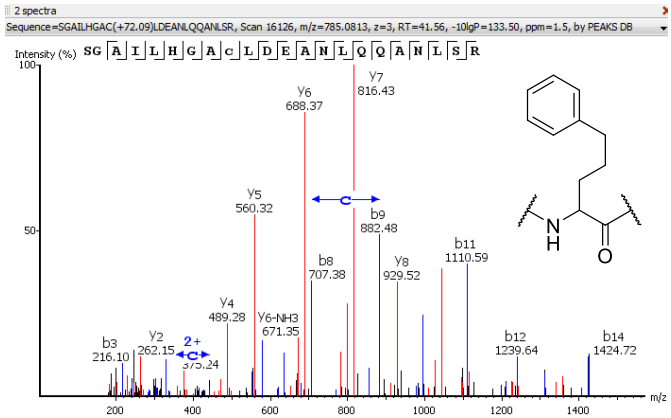
Installation of hhPhe on NPβ



NPβ-G2F-Dha61 **Full Conversion** NPβ-G2F-hhPhe61

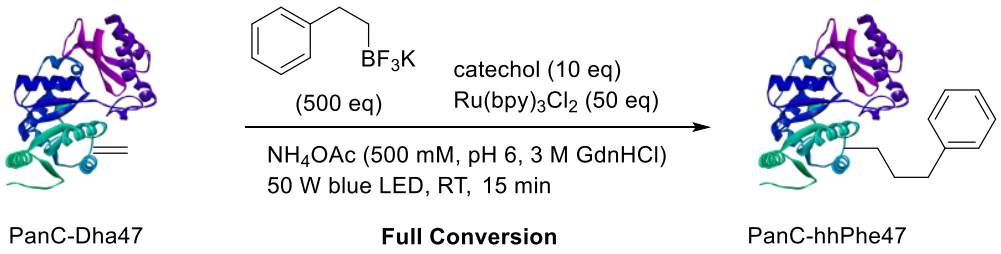
Calculated mass: 21103 Da. Observed mass: 21103 Da (+70 Da mass unknown)





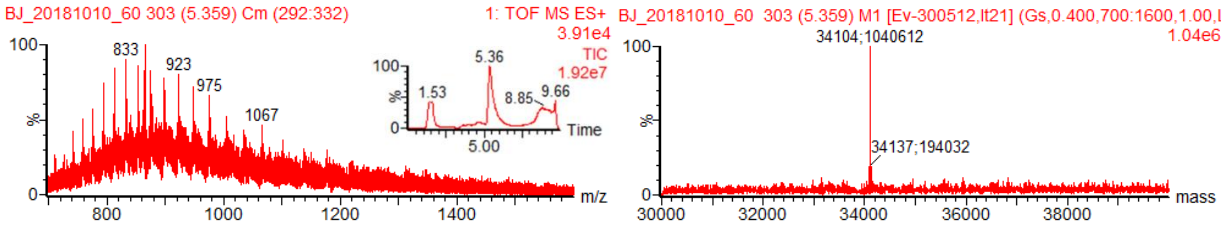
LC-MS/MS of NPβ-G2F-hhPhe: Tryptic digest leads to 96% sequence coverage (1% FDR). Q Exactive (HCD).

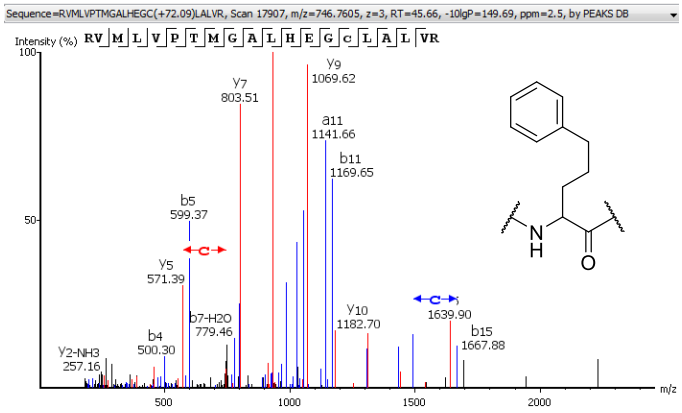
Installation of hhPhe on PanC



PanC-Dha47 **Full Conversion** PanC-hhPhe47

Calculated mass: 34104 Da. Observed mass: 34104 Da (+ Met Ox 34137 Da)

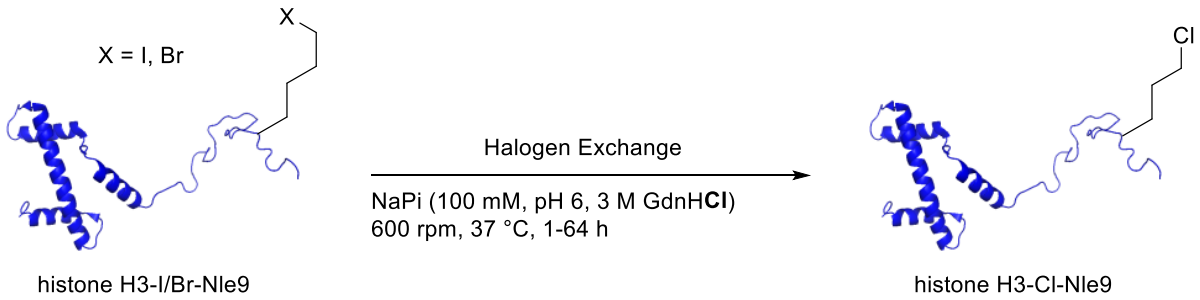




LC-MS/MS of PanC-hhPhe47: 97% sequence coverage with tryptic digest (1% FDR). Q Exactive (HCD).

Reactions with electrophile containing sidechains (alkyl halides)

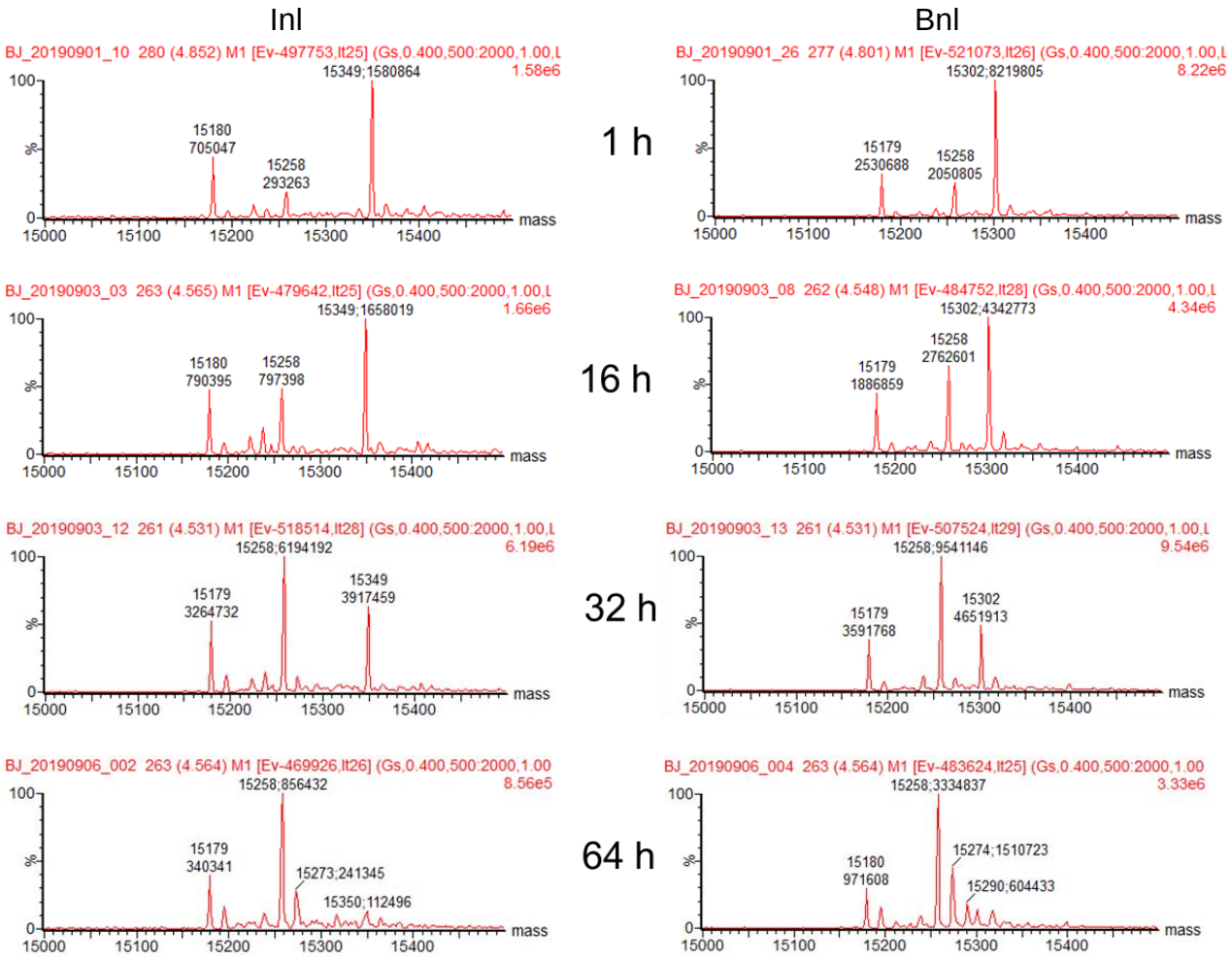
Formation of chloronorleucine (Cl-Nle) via halogen exchange



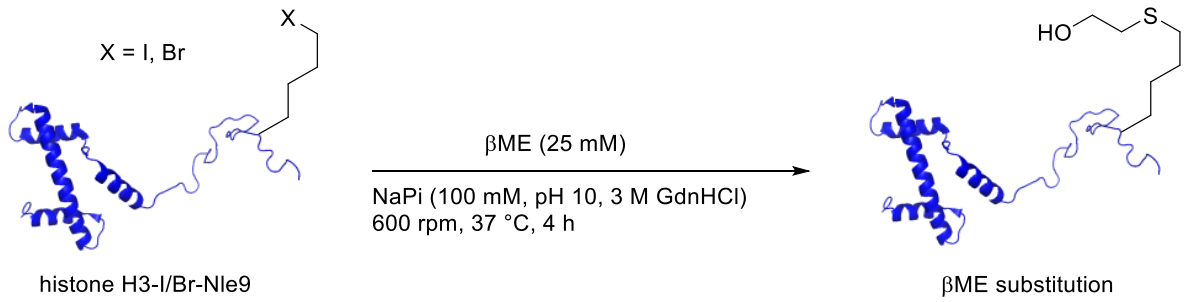
Calculated mass of histone H3-Cl-Nle9: 15258 Da. Observed mass: 15258 Da

Calculated Dha mass (residual): 15180 Da. Calculated histone H3-I-Nle9: 15350 Da.

Calculated histone H3-Br-Nle9: 15302 Da. Met oxidation: +16 Da



Reaction of βME with histone H3-I-Nle9 and histone H3-Br-Nle9



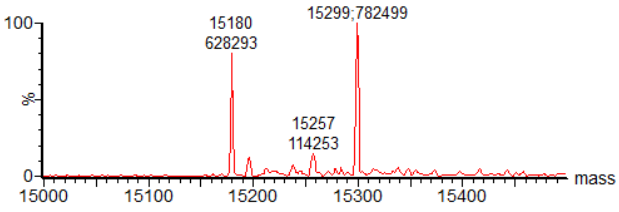
Calculated mass of βME substitution: 15300 Da

Observed mass: 15299 Da (H3-I-Nle9), 15300 Da (H3-Br-Nle9)

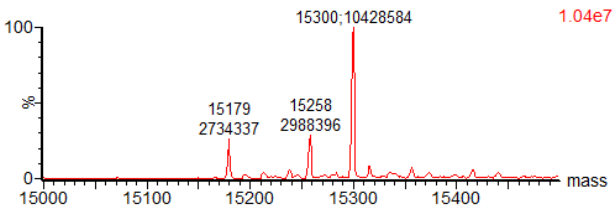
Calculated Dha mass (residual): 15180 Da

Calculated Histone H3-Cl-Nle9: 15258 Da

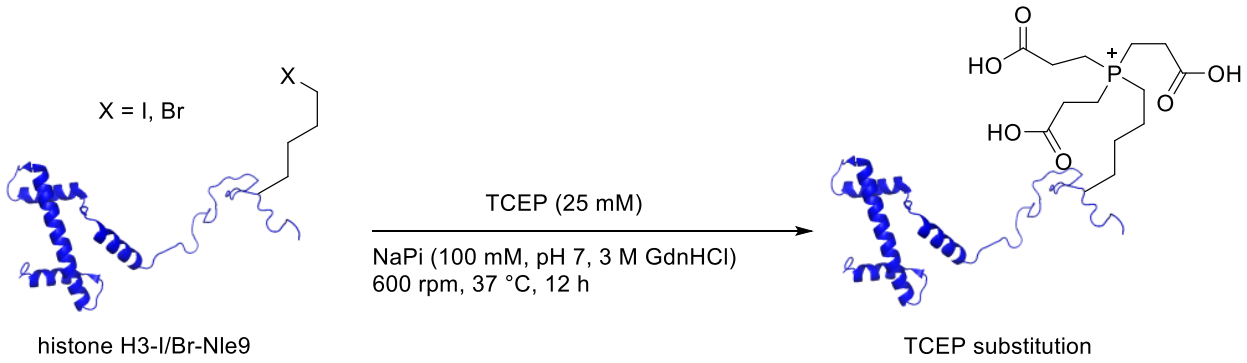
H3-I-Nle9



H3-Br-Nle9



Reaction of TCEP with histone H3-I-Nle9 and histone H3-Br-Nle9



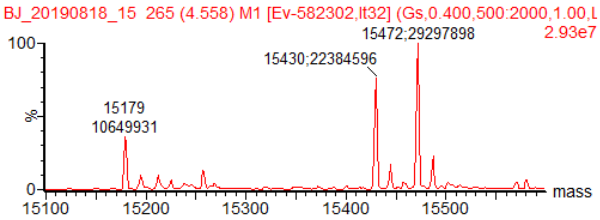
Calculated mass of TCEP substitution: 15473 Da. Observed mass: 15472 Da

Calculated Dha mass (residual): 15180 Da. Calculated histone H3-Cl-Nle9: 15258 Da

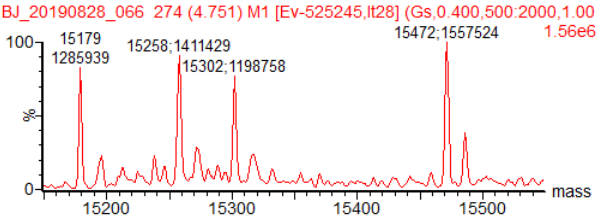
Calculated mass of TCEP addition to histone H3-Dha9 (background): 15430 Da

Calculated mass of H3-Br-Nle9 (unreacted): 15302 Da

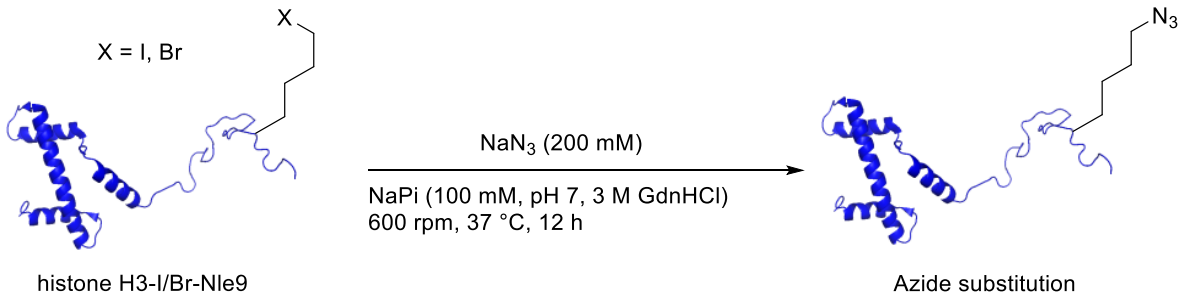
H3-I-Nle9



H3-Br-Nle9



Reaction of sodium azide with histone H3-I-Nle9 and histone H3-Br-Nle9



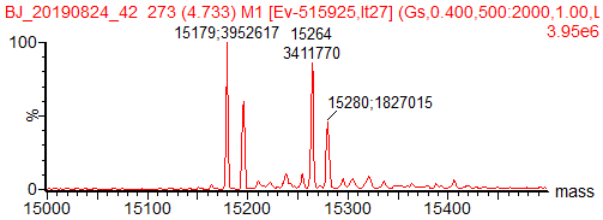
Calculated mass of azide substitution: 15265 Da. Observed mass: 15264 Da

Calculated Dha mass (residual): 15180 Da. Calculated histone H3-Cl-Nle9: 15258 Da

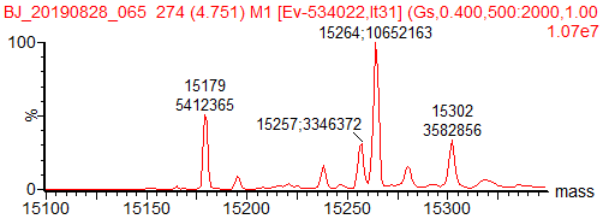
Calculated mass of histone H3-Br-Nle9 (unreacted): 15302 Da

Calculated Met oxidation: +16 Da

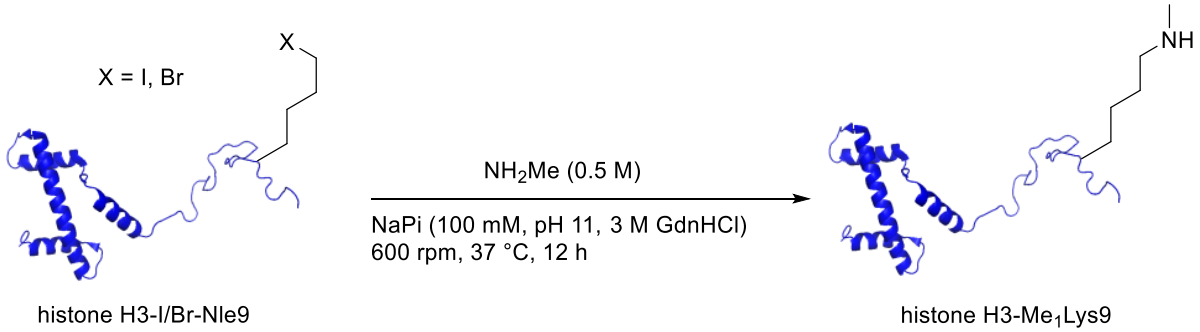
H3-I-Nle9



H3-Br-Nle9



Reaction of methylamine with histone H3-I-Nle9 and histone H3-Br-Nle9



Calculated mass of methylamine substitution: 15253 Da. Observed mass: 15253 Da

Calculated Dha mass (residual): 15180 Da.

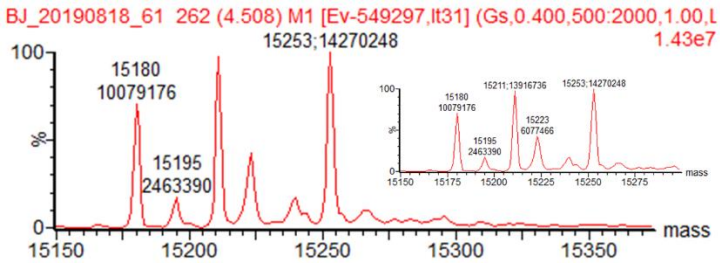
Calculated mass of histone H3-Cl-Nle9: 15258 Da. Met oxidation: +16 Da

Calculated mass of aza-Michael addition of methylamine to H3-Dha9: 15211 Da

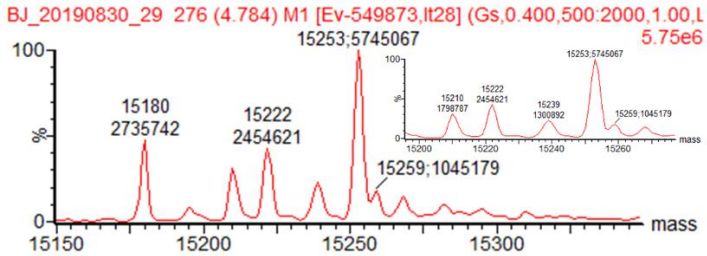
Calculated mass of elimination product: 15222 Da

Calculated mass of hydroxyl substitution produce: 15240 Da

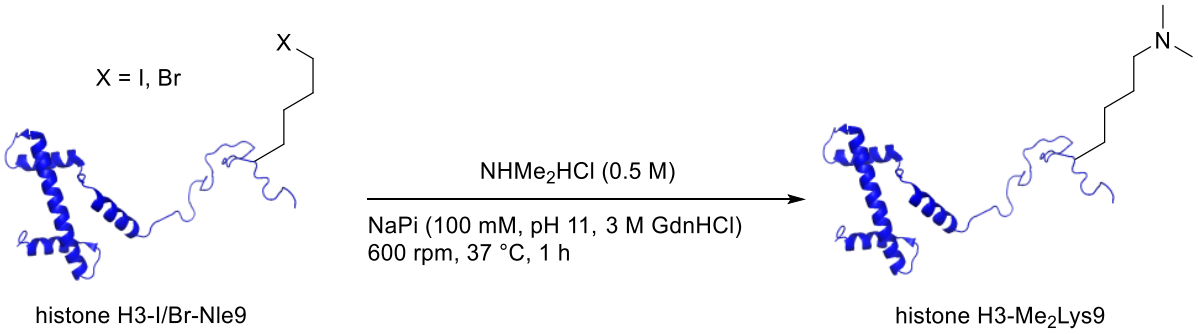
H3-I-Nle9 (inset zoomed in)



H3-Br-Nle9 (inset zoomed in)



Reaction of dimethylamine with histone H3-I-Nle9 and histone H3-Br-Nle9



Calculated mass of dimethylamine substitution: 15267 Da. Observed mass: 15267 Da

Calculated Dha mass (residual): 15180 Da.

Calculated mass of histone H3-Cl-Nle9: 15258 Da

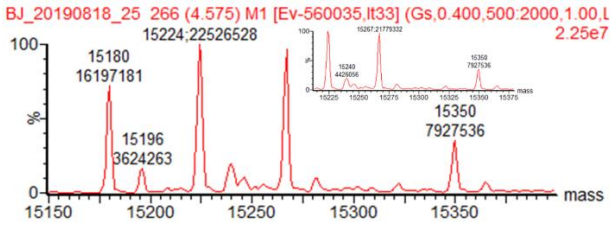
Calculated mass of aza-Michael addition of dimethylamine to H3-Dha9: 15225 Da

Calculated mass of hydroxyl substitution produce: 15240 Da

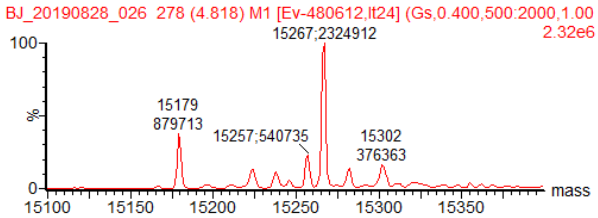
Calculated mass of histone H3-I-Nle9 (unreacted): 15350 Da

Calculated mass of H3-Br-Nle9 (unreacted): 15302 Da

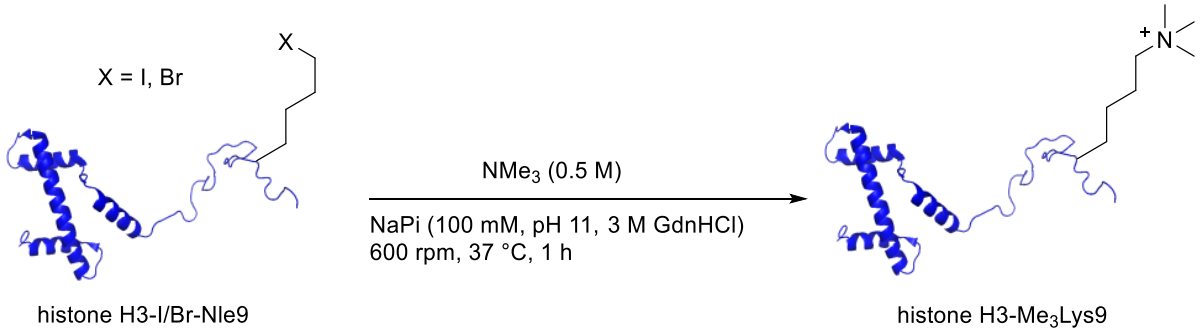
H3-I-Nle9 (inset zoomed in)



H3-Br-Nle9



Reaction of trimethylamine with histone H3-I-Nle9 and histone H3-Br-Nle9



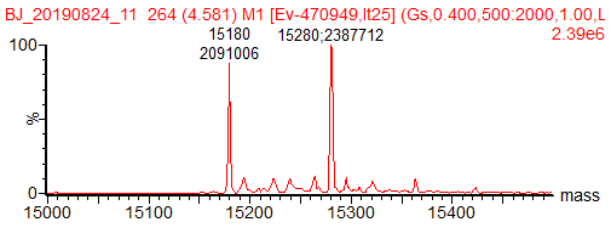
Calculated mass of trimethylamine substitution: 15280 Da

Observed mass: 15280 Da (H3-I-Nle9), 15281 Da (H3-Br-Nle9)

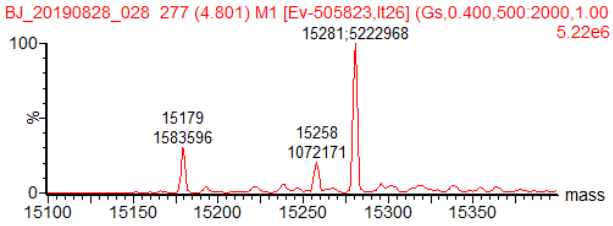
Calculated Dha mass (residual): 15180 Da

Calculated mass of histone H3-Cl-Nle9: 15258 Da (seen only for H3-Br-Nle9)

H3-I-Nle9



H3-Br-Nle9



Chapter 2: Investigating Novel Applications with Chemically Modified Histones

"Every protein is its own little monster"

- Professor Ryan Mehl (Oregon State University)

Relevant publications (* indicates equal contribution):

Josephson, B.*, Fehl, C.*, Isenegger, P.G.* *et al*. Light-driven post-translational installation of reactive protein side chains. *Nature* **585**, 530–537 (2020).

Mollner, T.A., Isenegger, P.G.*, **Josephson, B.*** *et al*. Post-translational insertion of boron in proteins to probe and modulate function. *Nat Chem Biol* **17**, 1245–1261 (2021).

Faulkner, S., Couturier, A.M., **Josephson, B.** *et al*. Intact nucleosomal context enables chromodomain reader MRG15 to distinguish H3K36me3 from -me2. *bioRxiv*, 2020.2004.2030.070136

Introduction

Current applications of chemically modified histones

Due to their biological ubiquity and importance in part due to their high PTM variety and content, histones have been a popular target for protein modification chemistry[15,51,54,94,95]. While a multitude of chemical strategies have allowed for site-specific installation of a variety of PTMs of interest, such as methylation, acetylation, phosphorylation, and ADP-ribosylation (**Figure 29, left**)[94,95,231], the utilized chemical strategies can often go beyond the structural and chemical limitations imposed by biology.

This allows for epigenetic systems to be investigated further through the incorporation of native residue and PTM variants (**Figure 29, center**). Phosphoserine mimics with a γCH_2 or γCF_2 in place of the native oxygen are resistant to enzymatic dephosphorylation[135], allowing permanent pSer installation. Various Lys and Arg residues and their PTMs (such as acetylation and methylation) with alternative atoms in their sidechains (such as S, Se, O, or NH) are close analogues of the native residues but are also often resistant to PTM removal[95,232]. Furthermore, altering sidechain length[233,234], varying acyl PTM substituents[235,236], or installing different alkyl PTM variants[237,238] can test the specificity and selectivity of their respective enzymes. These approaches, amongst many others, allow for the probing of histone PTM related function at a greater depth than allowed by the canonical modifications found in epigenetic systems.

Furthermore, beyond installing natural residues and PTMs or their respective unnatural variants, a third option for chemically modifying histones is to install reactive tags, which facilitate a wide range of bio-orthogonal chemistry (**Figure 29, right**). While there are too many to satisfactorily review here, common examples are the installation of site-specific

crosslinking facilitators, click chemistry functional groups for bioconjugation, fluorophores, or tags such as biotin used for later enrichment[94,95,222]. A few examples are highlighted in **Figure 29 (right)**. Installation of 'photolysine', followed by UV light treatment, promotes crosslinking to residue specific interaction partners[199,239]. Similarly, acyllysine variants containing alkyl halides and incorporated through GCE allow for proximity-based protein-protein crosslinking, a useful strategy that has yet to be used in histones in this context[223,240]. Additionally, common click chemistry pairs can be installed through many methods, such as by hijacking the native lysine methylating machinery to install alkyne groups on lysine[241], allowing for bioconjugation with 'off-protein' sources of the respective click chemistry partner. The tools available for producing chemically modified histones are capable of a wide variety of applications. But, as will be discussed in the next section, the method needs to be compatible with the epigenetic substrate of interest.

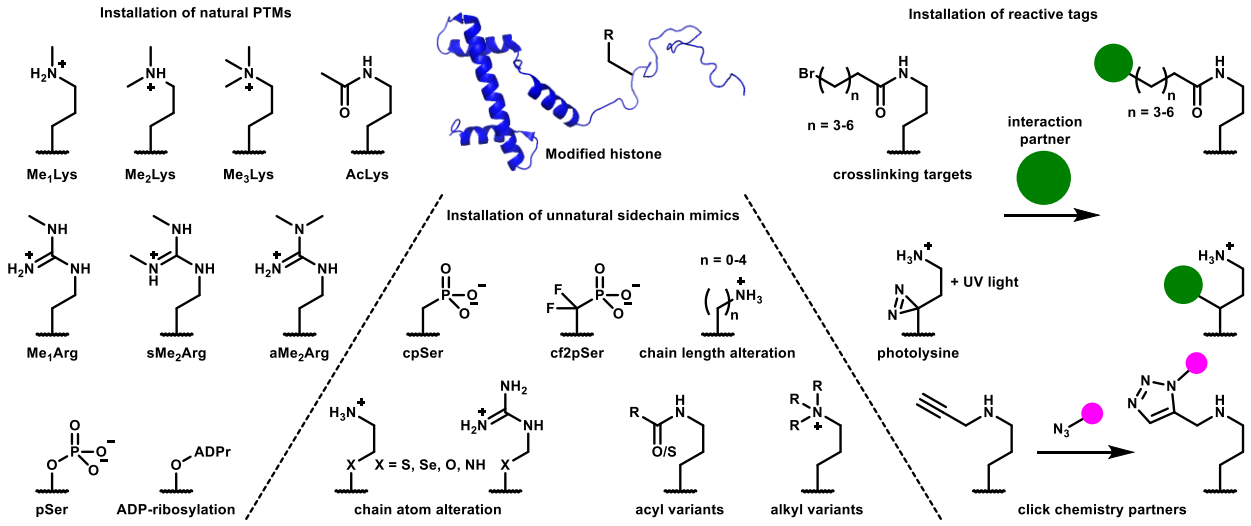


Figure 29: Different types of applications with modified histones.

Properties of chemical tools needed to study histone PTMs

Choosing a method to study histone modifications very much relies on the complexity of the epigenetic system of interest. Naturally, the more complicated the epigenetic system, the more barriers there are for protein modification chemistry, and the less tools available (**Figure 30**). In small peptide systems, it is easy to incorporate a wide variety of sidechains via peptide synthesis[95,109], a process that can be expanded to longer peptides or even full length proteins with more difficulty[95,110]. If a modified histone can be produced, there are time consuming but robust procedures for preparing histone octamers, nucleosomes[242], and nucleosome multimers[243,244] at various scales, although direct modification of these complexes should be the focus on more research. It is often vital to study these higher order complexes, as many crucial epigenetic processes require crosstalk between PTMs on different histones or necessitate vital interactions with DNA[245,246]. Moreover, the study of chromatin is most accurate in its native context, either through studies on chromatin extracts, or preferably, in unperturbed living cells or organisms.

Therefore, when designing or choosing a histone modification method, consideration must be given to its compatibility with the epigenetic substrate of interest[94]. A basic requirement is the lack of reactivity with the functional groups found in the canonical amino acids, but if full histone proteins or higher order complexes are desired, then protein compatibility, lack of protein complex disturbance and DNA compatibility are required as well. Furthermore, the desired modification must be stable in the conditions required for complex assembly or study. Installing desired histone modifications *in vivo* is the most difficult, as conditions and reagents must not only be biocompatible, but also must not perturb the highly sensitive chromatin states found in cells. Histone modification methods

utilizing GCE can accomplish this but require substantial engineering of the cell's expression system[96,98,100]. Recent *in vivo* work utilizing ligand-guided histone PTM installation[247,248] or intein mediated histone PTM installation[114-116] is also promising, but requires considerable expertise and experimental design.

More methods for installing desired histone PTMs on increasingly complex epigenetic substrates are sorely needed. Moreover, the methods should be used to study various epigenetic processes of interest, preferably on the most complex epigenetic substrate compatible with the method. Researchers developing methods to install new unnatural or reactive sidechains should also strive to show how they allow for the study of epigenetic processes in unique or advantageous ways.

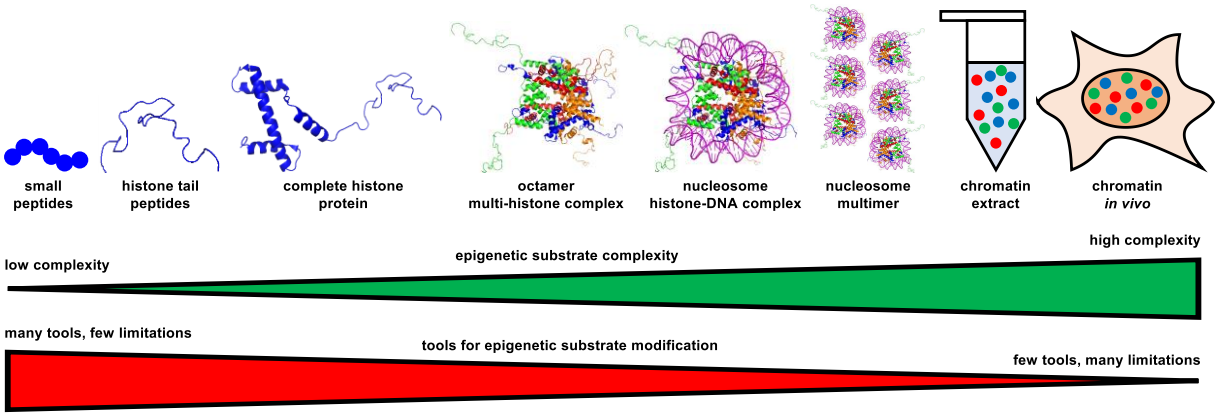


Figure 30: Spectrum of epigenetic complexity and its association with compatible methods to introduce desired histone modifications.

Goals of this Chapter

Chapter 2 seeks to study a variety of epigenetic processes by exploiting site-specific protein modification methods recently developed in our group (**Figure 31**)[135,208,221]. While the focus will primarily be on the epigenetic system of interest, efforts will be made to showcase how the specific protein modification methods are unique and useful tools for studying the system of interest in novel ways. Additionally, **Chapter 2** aims to study epigenetic processes with all three types of sidechain installation strategies discussed earlier (**Figure 29**): Installing native histone PTMs to study their function, installing unnatural sidechain and PTM mimics to allow for their study in novel ways, and installing reactive sidechains allowing for site-specific 'on-protein' chemistry.

Native residues and PTMs of interest will be installed onto histones using the light-driven radical Dha reaction developed in **Chapter 1**[208], as well as our group's older radical Dha modification method (**Figure 31, A and B**)[135]. These PTM containing histones will be first used to test the functions of PTM processing enzymes, such as the Sirt2 deacylase, as well as be assembled into nucleosomes to test the effects of DNA interactions on different methylated states of histone H3-Lys36 on the accessibility of that residue[249].

Unnatural sidechain and PTM mimics will be installed onto histones using the light-driven radical Dha modification reactions either developed in **Chapter 1** (**Figure 31B**) or by colleague Dr. Patrick Isenegger (**Figure 31C**)[208]. The unnatural acyllysine PTM, BzLys[76], will be installed onto histone H3-Lys18, allowing for its susceptibility to Sirt2 deacylase removal to be screened. In addition, γCF_2 containing AcLys PTM mimics will be installed into histones and their PTM processing monitored by ^{19}F-NMR. Other γCF_2 containing sidechains, such as the oxidized Met PTMs, will be characterized by ^{19}F-NMR as well.

Additionally, reactive chemical tags, such as halogen containing sidechains installed by the light-driven radical Dha modification reaction (**Figure 31B**)[208], or the Dha borylation reaction developed by colleague Dr. Tim Mollner (**Figure 31D**)[221], will be installed into histones. The electrophilic alkyl halide containing sidechains, building off work in **Chapter 1**, will be assessed as targets for site-specific proximity-driven protein-protein crosslinking with the histone demethylase KDM4A. The boronoalanine (Bal) sidechains will be used as structural NMR probes in reconstituted nucleosomes treated with Aurora B Kinase by tracking Bal interactions with the amide backbone throughout the disordered histone tail in response to phosphorylation[221].

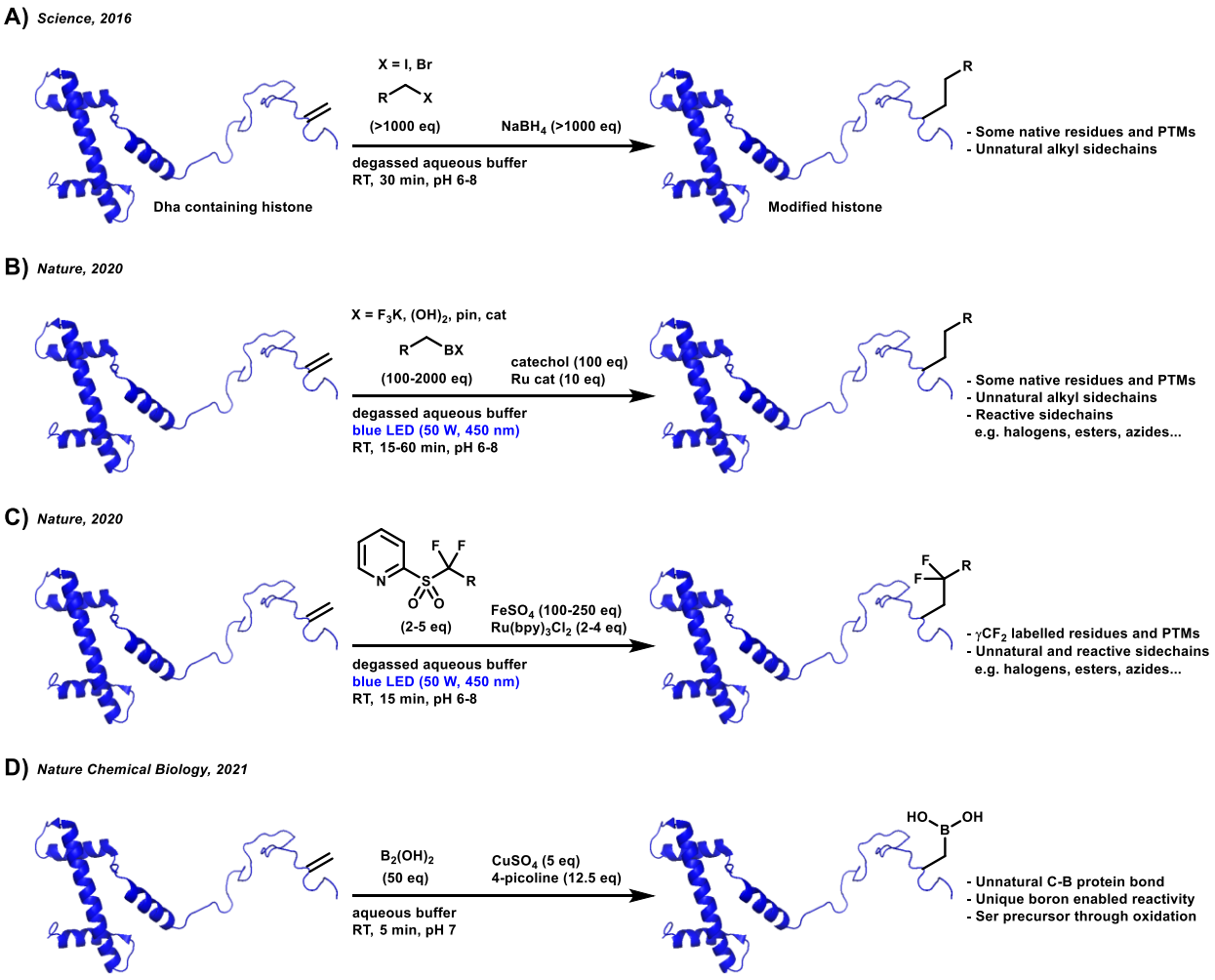


Figure 31: Overview of the chemical methods used in this Chapter to modify histones and study various epigenetic phenomena. **A)** The first method allowing βC-γC radical bond formation reaction on Dha[135]. **B)** The reaction developed in **Chapter 1** allowing for reactive sidechain installation on Dha in addition to native residues and PTMs[208]. **C)** A complementary reaction to **(B)** which installs γCF_2 labelled sidechains[208]. **D)** A Dha borylation reaction recently developed in our group[221].

Protein-protein crosslinking with “on-protein” electrophiles

Overview of current protein-protein crosslinking methods

The ability to site-specifically crosslink two proteins together, with applications in either protein complex engineering or interaction partner capture, has been a major focus of protein modification methods[240]. Numerous strategies utilizing noncanonical amino acid crosslinking targets have been developed, with most strategies relying on modified versions of Tyr (**Figure 32A**) or extensions of acylated Lys (**Figure 32B**) residues incorporated via GCE techniques[240].

The incorporation of electrophilic alkyl halides into these Tyr or Lys sidechain analogues has been of substantial focus, with numerous strategies allowing for targeting the nucleophilic Cys residues of protein targets, as well as other nucleophilic residues such as Lys, Tyr, and His[223-225,250-254]. Other moieties, such as Michael acceptors[255], isothiocyanates[256], aryl carbamates[257], ketones[254], and a cyclopentadiene (capable of driving dimerization through a Diels-Alder reaction)[258], have also been incorporated via GCE (**Figure 32A and B**). Notably, a Tyr analogue containing both alkyl halide and alkyne dual functionality allowed for crosslinking to interaction partners followed by enrichment via utilization of the alkyne for click chemistry enrichment[225].

In a related strategy, various amino acids, or their PTMs, containing a diazirine functional group in their sidechain, can drive crosslinking to surrounding residues after activation via UV light irradiation (**Figure 32C**), often targeting aliphatic amino acids such as Gly[199,259,260]. These crosslinking residues are perhaps the best mimics of native sidechains and PTMs, but due to their reliance on peptide synthesis or global

incorporation into proteins via native machinery, site-specific installation in full length proteins remains difficult[260]. GCE can be used to install photo-lysine sidechains site-specifically, but it requires elaborate synthesis and deprotection schemes[260].

While there are many options for incorporating sidechains capable of crosslinking, most require drastic extensions off Lys or Tyr residues, due to their dependence on amber codon suppression methodology[98], and therefore do not properly mimic native sidechains or PTMs, occluding potential sidechain or PTM specific binding partners. We reasoned that, through our chemical methodology developed in **Chapter 1**, we could install alkyl halide containing sidechains directly onto the protein backbone that could more properly mimic the structure and connectivity of native or modified Lys residues, providing a more direct and general strategy for capturing site- and PTM-specific interaction partners through proximity-driven protein-protein crosslinking (**Figure 32D**).

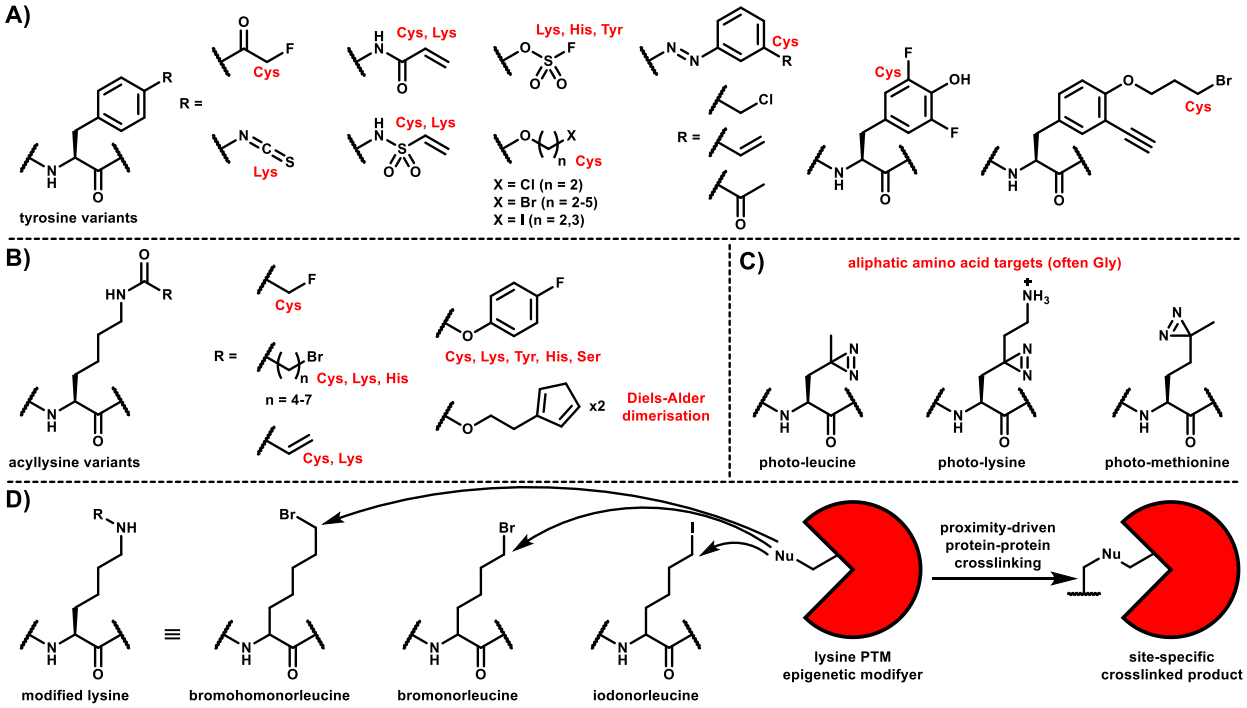


Figure 32: Unnatural amino acids used for crosslinking based on tyrosine **(A)** and acyllysine **(B)** analogues installed by GCE. **C)** Examples of diazirine containing unnatural amino acids that drive crosslinking after UV light illumination. **D)** Modified lysine mimicking halogen containing sidechains accessible through the light-driven radical Dha reaction developed in **Chapter 1** may allow for capture of specific lysine PTM modifying enzymes.

Crosslinking KDM4A to histone H3 via electrophilic sidechains

With our ability to install electrophilic alkyl halide containing sidechains in hand (from **Chapter 1**)[208], we decided to test their use as crosslinking targets from nucleophilic sidechains (such as Cys) in histone interaction partners. The histone demethylase KDM4A was chosen as a model interaction partner (kindly provided by Professor Akane Kawamura), due to the proximity of three Zn binding Cys residues to the methylated lysine binding pocket (**Figure 33A**)[261,262]. As KDM4A is specific for methylated Lys residues (histone H3-Lys9/36)[261], we chose to install bromohomonorleucine (Bhn), as its length and connectivity best matched that of a methylated Lys residue (**Figure 32D**). We further decided to install Bhn at three sites, the KDM4A target site Lys9, as well as non-target sites Lys4 and lys27, to test the specificity or promiscuity of KDM4A crosslinking.

KDM4A (2 µM) was incubated with either histone eH3.1-Bhn4, -Bhn9, -Bhn27, or -WT (4 µM) in HEPES buffer (50 mM, pH 7.4) for 1 h at 37 °C, with the goal of covalently capturing KDM4A via a proximity-driven reaction between one of the Zn bound Cys residues[262] and the Bhn sidechain (**Figure 33B**). The reaction was quenched with sample buffer, boiled (10 min), and analyzed by SDS-PAGE (**Figure 33C**). Pleasingly, for every sample containing both KDM4A and a Bhn-containing histone eH3.1, a distinct band showed up at the mass corresponding to the combined mass of KDM4A and the histone (**Figure 33C, ***). This band was not present in the incubation of histone eH3.1-WT and KDM4A, or in any of the incubations with KDM4A or the histones alone, indicating that crosslinking successfully and selectively occurred as desired. Interestingly, there was no observable preference of KDM4A for any of the three Bhn sites (Lys4/9/27), perhaps hinting that it samples multiple Lys residues along the histone tail, but is only active on specific sites.

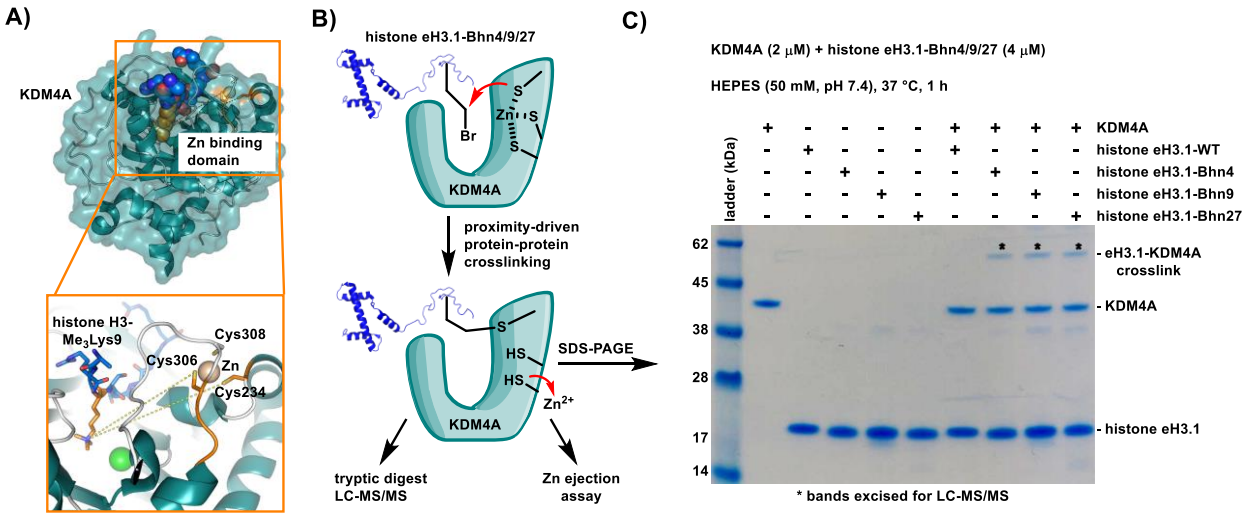


Figure 33: A) Structure of KDM4A bound to its substrate, Me_3Lys9 (PDB: 2OQ6) in a histone tail peptide, showing the proximity of three Zn binding Cys residues to the active site pocket. **B)** Workflow of crosslinking KDM4A with histones containing the electrophilic Bhn residue installed at various Lys sites, followed by analysis by SDS-PAGE **(C)**, LC-MS/MS (**Figure 35**) and Zn ejection assays (**Figure 36**).

Next, we tested the crosslinking dependence on incubation time and temperature at each of the three Bhn sites (Lys4/9/27). As expected, longer incubation times (2 h vs 15 min) led to a more pronounced crosslinking band, and lower temperatures (4 °C) produced no apparent crosslinking products, in contrast to the strong bands observed at higher temperatures (37 °C) (**Figure 34A**). Again, all three Bhn sites (Lys4/9/27) produced indistinguishable results.

We then tested the crosslinking dependence on both the necessity of a strong nucleophilic residue (Cys), and identity of the crosslinked protein as an interaction partner. All three histone eH3.1-Bhn proteins (Bhn4/9/27) and a WT control were either incubated with known histone H3 interaction partner histone $H4^{242}$, which contains no Cys

residues, or the protein BSA, which contains a free Cys residue[263] but is not a histone H3 interaction partner (**Figure 34B**). No crosslinked products were seen for either protein, at any Bhn site, even after a 2 h incubation at 37 °C, confirming that crosslinking requires both a nucleophile and the proximity provided by an interaction partner. Lastly, the suspected crosslinked bands (**Figure 33C, ***) were confirmed by identifying both KDM4A (via its 6His tag) and histone eH3.1 in the bands by western blot (**Figure 34C**).

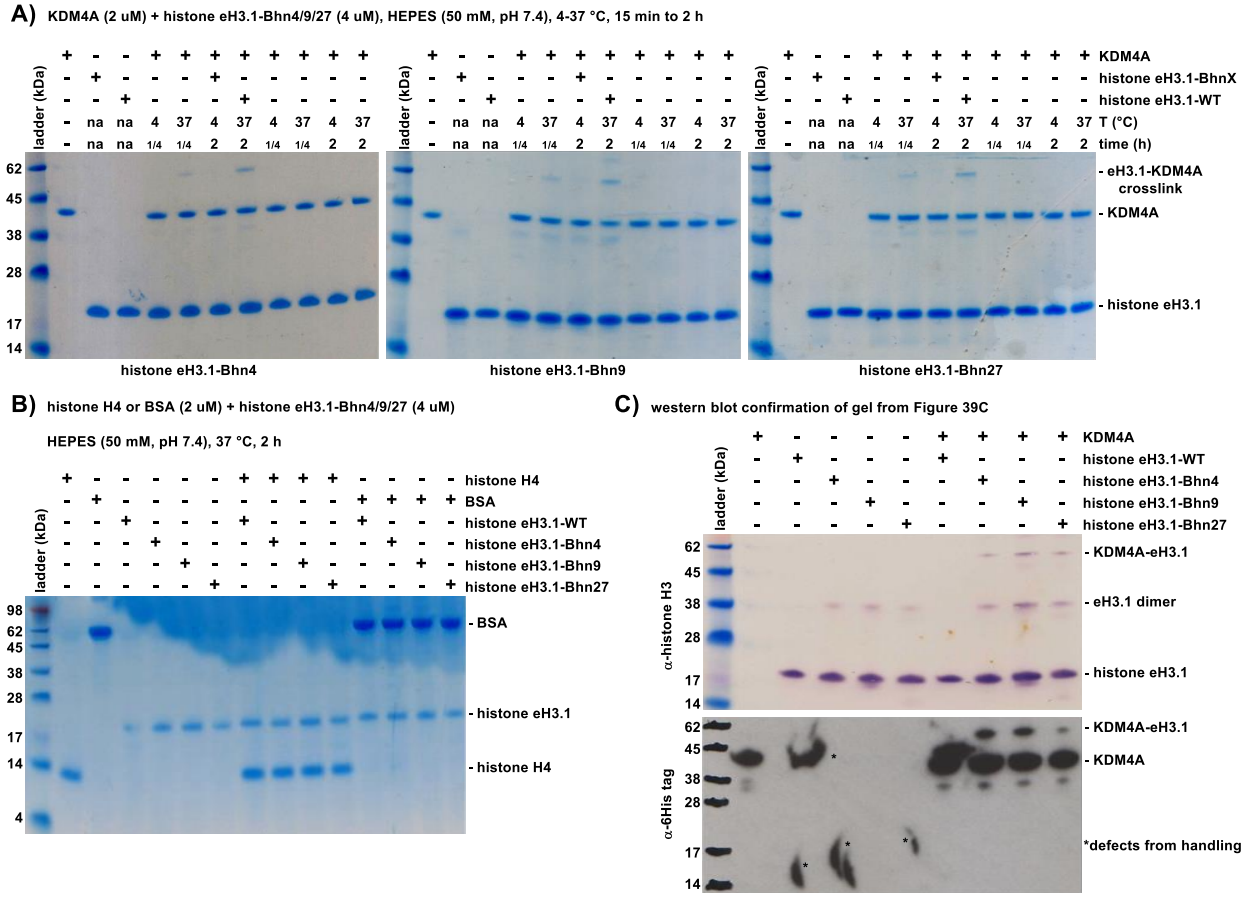


Figure 34: A) Incubation time and temperature screenings for KDM4A crosslinking to three histone eH3.1-Bhn sites (4, 9, 27). **B)** Incubation of the interaction partner histone H4 (no Cys residues), and unrelated Cys containing protein BSA with histone eH3.1-Bhn histones show no signs of crosslinking. **C)** Western blots against histone eH3.1 (top) and KDM4A (bottom, via 6His tag) confirm crosslinked bands contain both proteins.

Validation of crosslinking via LC-MS/MS and Zn ejection assay

To confirm our suspicion that the Zn binding Cys residues proximal to the KDM4A active site were the crosslinking sites, we excised the KDM4A-histone eH3.1-Bhn4/9/27 crosslinked bands from the gel (**Figure 33C, ***) and performed LC-MS/MS (carried out and analyzed by Dr. Simon Nadal and Professor Shabaz Mohammed). Due to the suspected substitution crosslinking reaction on the Bhn alkyl halide sidechain, the predictable linker structure allowed for the crosslinking sites to be identified.

As expected, two of the three Zn binding Cys residues (Cys-234 and Cys-306) were identified as the crosslinking sites on KDM4A by searching LC-MS/MS spectra for the mass of KDM4A and histone eH3.1 peptides plus the known linker mass (**Figure 35A-C**). Interestingly, while both histone eH3.1-Bhn4 and -Bhn9 gave crosslinked products for both Cys residues, -Bhn27 only gave results for one Cys residue, (Cys 234).

Surprisingly, we observed a self-crosslinking product between histone eH3.1-Bhn4 and -Ser10/Thr11 (indistinguishable), indicating formation of a Williamson-type ether (**Figure 35D**). This is unexpected due to its low reaction rate[264,265], indicating strong effective molarity enhanced reactivity due to the apparent proximity of the two histone sidechains. This may be due to the KDM4A-histone eH3.1 interaction existing as a dimer[266,267], holding the histone tails together and driving ether formation (**Figure 35E**).

Taken altogether, the Bhn sidechain shows exceptional crosslinking capability when incubated with interaction partners containing nearby nucleophiles, even allowing effective-molarity-driven ether formation. This contrasts the lack of reactivity seen in **Chapter 1** with ‘off-protein’ nucleophiles often needing high mM or M concentrations at high pH, supporting Bhn as a highly selective proximity-driven crosslinking sidechain.

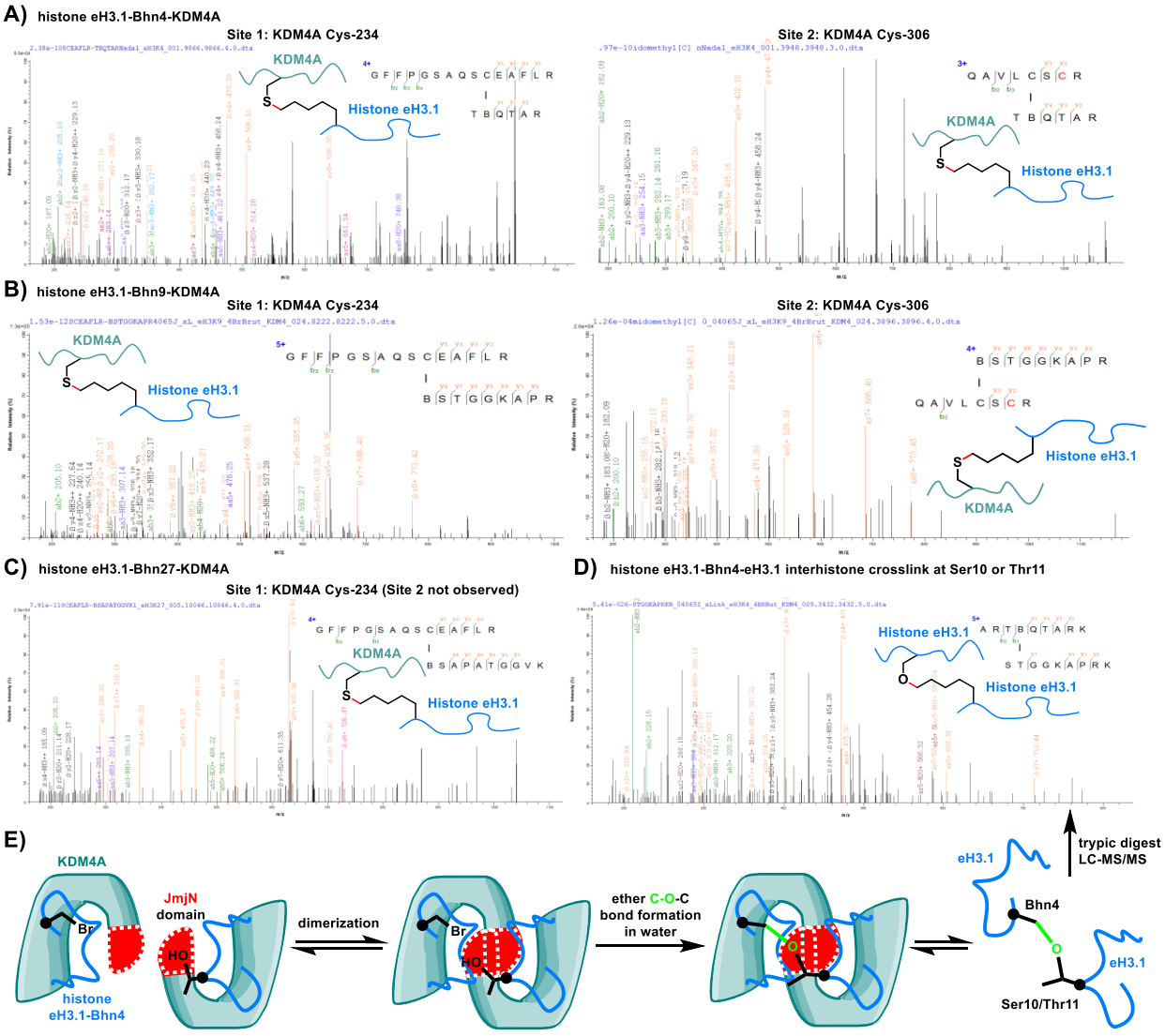


Figure 35: Representative LC-MS/MS spectra for KDM4A-histone eH3.1 crosslinked peptides formed with histone eH3.1-Bhn4 **(A)**, -Bhn9 **(B)**, and -Bhn27 **(C)**. Note that crosslinked products arose from Cys-234 and Cys-306, except for **(C)** in which only Cys-234 products were seen. **D)** An unexpected self-crosslink product was observed between histone eH3.1-Bhn4 and eH3.1-Ser10/Thr11, perhaps arising from known dimerization events driving ether formation through heightened effective molarity **(E)**.

To further characterize the crosslinking of KDM4A with Bhn-containing histone eH3.1, we performed a Zn ejection assay (assisted by Grace Roper), which measures the presence of Zn(II) released into solution as the Zn bound Cys residues get alkylated, a method commonly used to assessed KDM4A inhibitors which alkylate these Cys residues[262]. KDM4A (2 μM) was treated with 20 μM of either histone eH3.1-Bhn9, -WT, 4-bromobutylboronic acid (the small molecule precursor to the Bhn sidechain), or no compound in the presence of TSQ, a fluorescent probe for free Zn(II), all in HEPES buffer (50 mM, pH 7.5, 1.1% v/v DMSO) at 37 °C (**Figure 36A**).

The release of free Zn(II) was strongly detected only in the sample containing histone eH3.1-Bhn9, indicating Cys alkylation and Zn ejection. Histone eH3.1-WT and no compound controls produced no such response. Interestingly, the 4-bromobutylboronic acid compound also produced no response, even though it contained an identical alkyl halide functional group. This indicates that the desired Cys alkylation can only occur when the alkyl halide is installed on the histone sidechain, provided proper context for proximity-driven reactivity. Notably, this is an analogues but more selective process to that of the KDM4A Cys alkylator and inhibitor Ebselen (**Figure 36B**)[262], indicating the potential use of site-specifically modified proteins containing context-dependent alkylating sidechains such as Bhn as enzyme inhibitors.

To quantify the rate of Cys alkylation by histone eH3.1-Bhn9 via measured Zn ejection, we carried out a Zn(II) calibration (**Figure 36C, D**) then repeated incubations of KDM4A and histone eH3.1-Bhn9 in triplicate while measuring the accumulation of released Zn(II). After normalizing to no enzyme control measurements, the time dependent release of Zn(II) from KDM4A in the present of histone eH3.1-Bhn9 was calculated to be 9.270 ±

0.025 nM/min, contrasting the much more rapid Zn(II) release observed with Ebselen, which we measured at >1663 nM /min. This indicates that histone eH3.1-Bhn9 is a slow but selective and irreversible crosslinker, and inhibitor, of KDM4A.

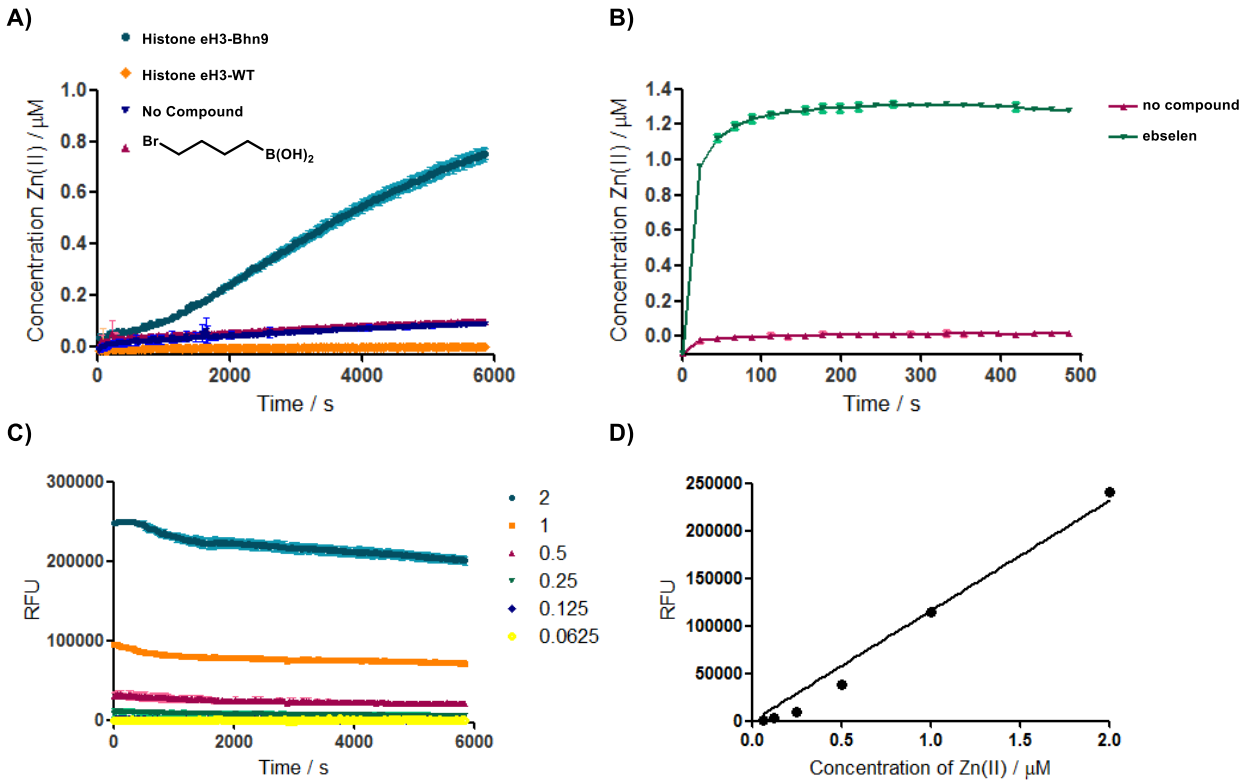


Figure 36: Zn(II) ejection assay. **A)** Ejection of Zn(II) is detected only after treatment of KDM4A with histone eH3.1-Bhn9. **B)** Ejection of Zn(II) was observed when KDM4A was treated with Ebselen (25 µM), an alkylator of the Zn binding Cys residues. **C)** A Zn(II) calibration (numbers indicate concentration in µM) was carried out to allow quantification of ejection Zn(II) from **(A)**. **D)** The calibration curve from **(C)**, plotted at time = 483 s.

Potential for site-specific discovery of unknown interaction partners

Inspired by our results from the KDM4A-histone eH3.1-Bhn4/9/27 crosslinking, we next tested whether histone eH3.1-Bhn9 could crosslink to interaction partners in a more complex epigenetic setting. Histone eH3.1-Bhn9, -WT, or no histone control (20 µg), pre-immobilized on α-HA beads (via the C-terminal HA tag on histone eH3.1) was incubated with HeLa cell nuclear lysate (250 µL, 0.5 mg/mL, kindly provided by Dr. Simon Nadal) for 4 h at 37 °C. After bead purification, histone eH3.1-Bhn crosslinked interaction partners were eluted and detected by western blot with an α-FLAG antibody (histone eH3.1 contains a C-terminal FLAG tag), with crosslinked products revealed by bands appearing at higher molecular weights than the lone histone (**Figure 37**). Multiple higher molecular weight bands were indeed seen in both the eluted fraction and aliquots at the end of incubation with histone eH3.1-Bhn9 (**Figure 37, indicated by ***), but were not present for either the histone eH3.1-WT or no histone control.

This suggests that histone eH3.1-Bhn9 was a successful crosslinking target for site-specific, currently unknown interaction partners in nuclear extracts. Future work could leverage this method to enrich interaction partners for various histone sites, as well as other proteins. Furthermore, pairing this method with LC-MS/MS as described above could not only reveal the site-specific interaction partners, but also reveal the crosslinking sites and residues on the respective interaction partners, providing detailed insights into epigenetic interactions with specific histone PTM sites.

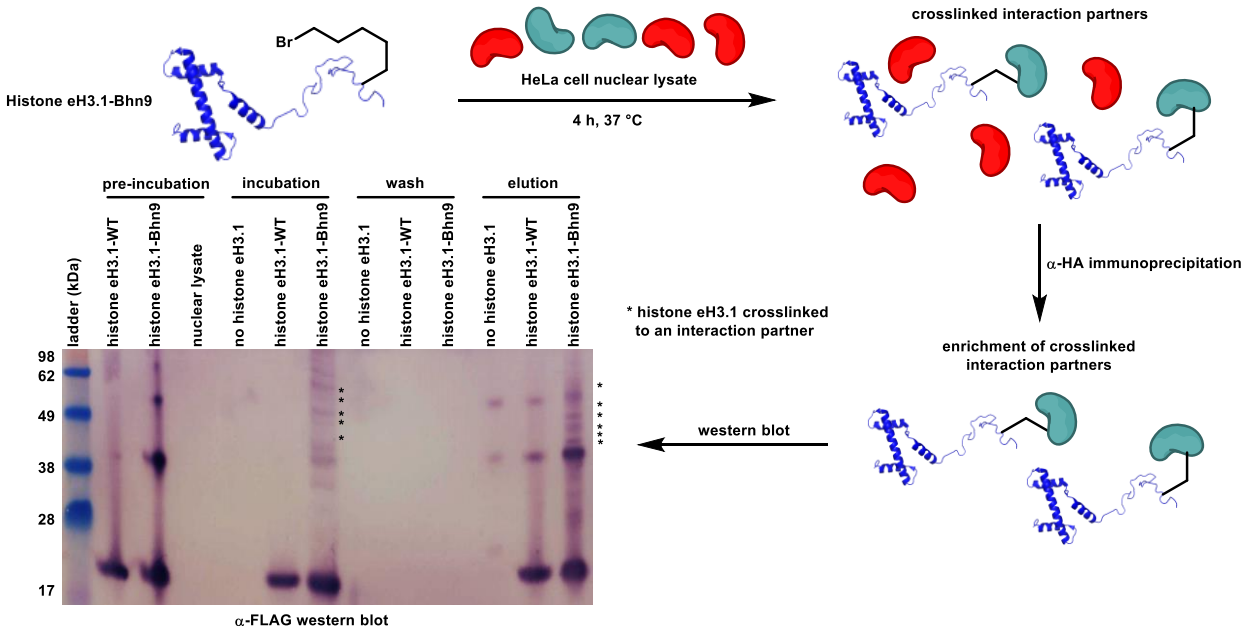


Figure 37: Histone eH3.1-Bhn9 was incubated in HeLa cell nuclear lysate, immunoprecipitated with crosslinked partners an α-HA antibody (histone eH3.1 is C-terminal FLAG and HA tagged), then western blotted (α-FLAG) to reveal crosslinked interaction partners with bands at higher molecular weights than that of the lone histone.

Investigating histone H3 tail structure in response to phosphorylation using Boronoalanine (Bal) as an NMR Probe

Previous work within the group

Recent work in our group undertaken by Dr. Tim Mollner developed a reaction that selectively borylates Dha residues on proteins, resulting in a boronic acid functionalized alanine residue with an unnatural Cβ-Bγ bond, named Boronoalanine (Bal) (**Figure 38A**)[221]. Due to the unique chemical properties[268] and near biological absence of the element B[269], Bal functionalized proteins could be leveraged for a variety of unique applications, such as binding to 'off-protein' diols, mediating proteolytic stability, and producing Ser upon oxidation in the presence of ROS such as hydrogen peroxide[221].

Of particular interest for this thesis work, the electron deficient (Lewis-acidic) B atom in the boronic acid (sp^2, $6e^-$) of Bal could dynamically engage with Lewis-basic residues both in the buffer and in the protein via Lewis-acid base pair (PLABP). NMR studies on small molecule Bal (**Figure 38B**) revealed exchange of the boronic acid to both boronate (fast exchange) from water and oxaborolane (slow exchange) from engagement with the amide backbone carbonyl, with pK_a values of 8.94 and 8.50, respectively, matching observations in previous literature[270,271]. This hinted at the ability of Bal to sample nearby amide backbone carbonyls within the protein, supported by the observation of -18 Da and -36 Da peaks in intact protein LC-MS spectra of Bal proteins, indicating single or double engagement of Bal releasing 1 or 2 water molecules, respectively[221].

To test this further, a $^{13}C^{15}N$-labelled histone H3-Bal9 protein was prepared, along with oxidized Bal analogue $^{13}C^{15}N$-labelled histone H3-Ser9. Overlaying $^{13}CO^{15}N$ NMR

spectra of these two proteins revealed the signal loss of several peaks in the Bal9 sample (**Figure 38D, gray circles**), confirming Bal engagement with the local peptidic backbone through PLABP leading to the signal intensity loss at nearby dynamically engaged residues, believed to arise from intermediate exchange due to the high apparent concentrations of carbonyls in the protein backbone.

It was reasoned that this Bal engagement with nearby backbone carbonyls through PLABP mediated dative-contact induced chemical exchange (DICE) could be utilized as an NMR probe for detecting transient structural content in intrinsically disordered proteins. Similar methods involving paramagnetic relaxation enhancement relying on incorporated nitroxide radicals can provide structural information on intrinsically disordered regions over long ranges, but struggle to provide reliable short range information[272]. We decided to assess Bal as a DICE NMR probe within the histone H3 tail, an intrinsically disordered region prone to structural adaptation upon PTM formation[273-275].

Previous work in our group conducted by Dr. Lukas Lercher identified two regions of the histone H3 tail that contained residual alpha helical structure in the context of a full nucleosome (**Figure 38E**). Furthermore, phosphorylation of histone H3 with Aurora B Kinase, which acts upon histone H3 at sites Ser10 and Ser28 and is an important epigenetic cell cycle regulator[276-278], conferred structural changes to the histone tail as observed by ^{1}H-^{15}N-HSQC NMR (**Figure 38F**). We chose this as the model system to investigate Bal enabled DICE NMR, as it would not only test the compatibility of Bal in a full nucleosome and its usefulness as a probe for intrinsically disordered protein structure, but could perhaps provide key insights into the structural consequences for the histone H3 tail after the epigenetically crucial Aurora B Kinase phosphorylation event[277].

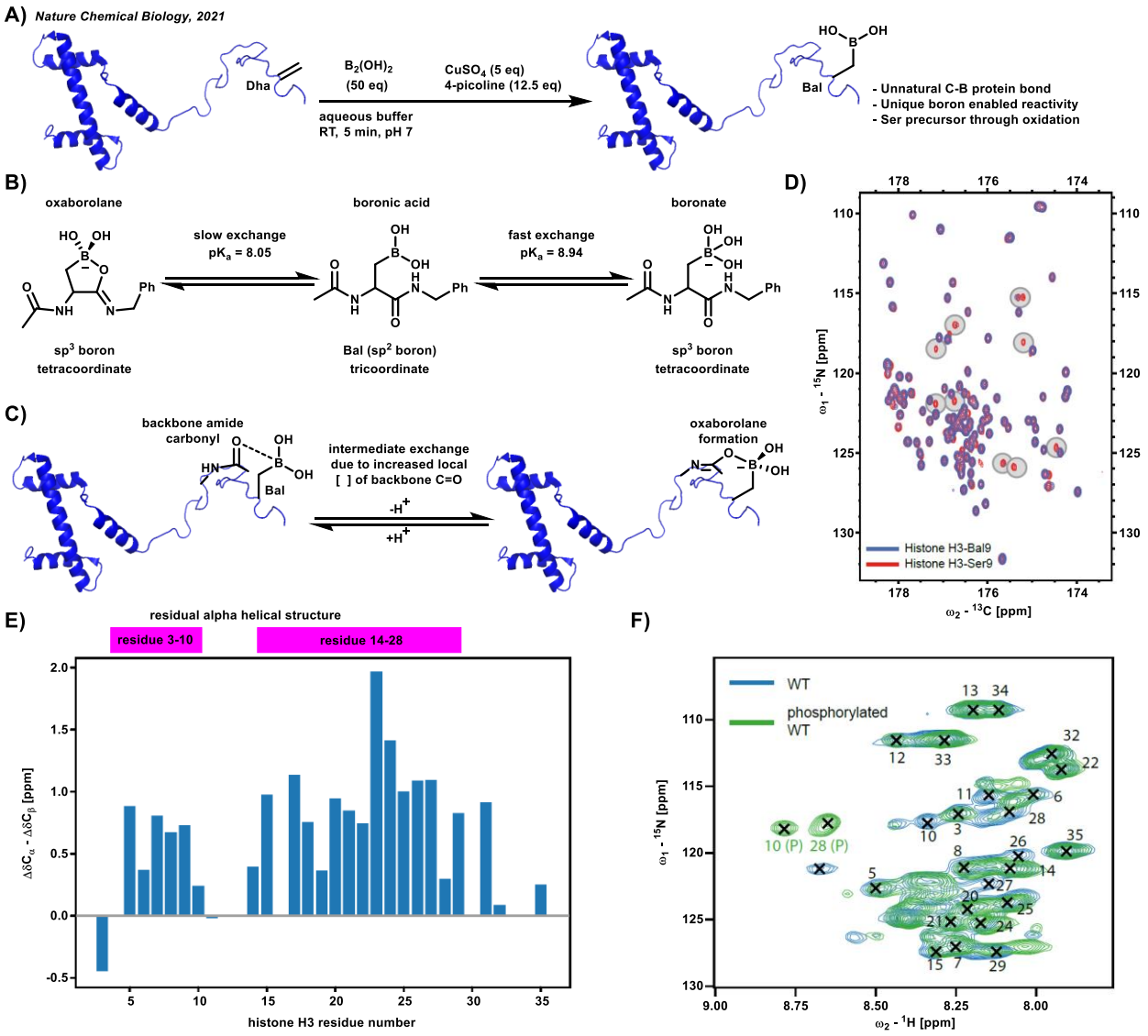


Figure 38: A) Dha borylation reaction. **B)** Bal exchanges to boronate and oxaborolane. **C)** Oxaborolane formation occurs on backbone carbonyls. **D)** $^{13}CO^{15}N$ NMR spectra comparing ($^{13}C^{15}N$)-histone H3-Bal9 and -Ser9, PLABP mediated intermediate exchange leads to signal loss with -Bal9 (circled). **E)** Measured $\Delta\delta C_\alpha$ - $\Delta\delta C_\beta$ chemical shifts reveal the histone H3 tail exhibits regional alpha helical structure in a nucleosome. **F)** ^{1}H-^{15}N-HSQC spectra comparing WT and Aurora B Kinase phosphorylated histone H3 in nucleosomes. Figure was adapted with permission from data provided by Dr. Tim Mollner.

Creation of Bal containing nucleosomes

We chose to install Bal onto the histone H3-Lys9 site, due not only to its proximity to the residue of interest, Ser10, the target of Aurora B Kinase phosphorylation, but also because of the known histone PTM crosstalk between Me_3Lys9 and pSer10[73,279]. We reasoned that -Bal9 mediated DICE NMR investigations of nucleosomes treated with Aurora B Kinase may provide key structural insights into this interaction, while also testing if the proximity of Bal9 interfered with proper Aurora B Kinase function at Ser10.

Dr. Tim Mollner kindly provided ^{15}N-histone H3-Bal9 and -WT. Octamer reconstitution and nucleosome assemblies were carried as per standard protocols[242], producing ~1 mg of each desired nucleosome (**Figure 39A**). Heat shift incubations (4/37/45 °C, 1 h) on aliquots of each nucleosome all gave single bands at the expected size on 6% TBE gels stained for both DNA (SYBR SAFE) and protein (Coomassie), indicating Bal did not interfere with nucleosome assembly or introduce structural aberrations. While some free DNA was present after nucleosome assembly, it was not expected to interfere with future analysis, and no further purification was carried out (see **Experimental Details** for further octamer reconstitution and nucleosome assembly characterization details).

^{15}N-histone H3-Bal9 and WT nucleosomes were prepared for ^{1}H-^{15}N-HSQC NMR by dialysis into the required NMR buffer (1 mM EDTA, 10 mM NaPi pH 6.5) with the last dialysis containing 5% v/v D_2O. After NMR measurement, both nucleosomes were phosphorylated with Aurora B Kinase (performed by Dr. Tim Mollner), reanalyzed on 6% TBE gels as before, and dialyzed again into NMR buffer for measurement.

To confirm that Aurora B Kinase could phosphorylate histone H3-Bal9 on the adjacent residue Ser10, free histone H3-Bal9 and WT were subjected to Aurora B Kinase

phosphorylation. A western blot using an α-histone H3-pSer10 antibody (**Figure 39B**) showed that both histone H3-WT and Bal9 could be phosphorylated. This is notable because Bal did not inhibit Aurora B Kinase activity on an adjacent residue, or interfere with antibody recognition of pSer10, hinting at its benign nature as a structural probe.

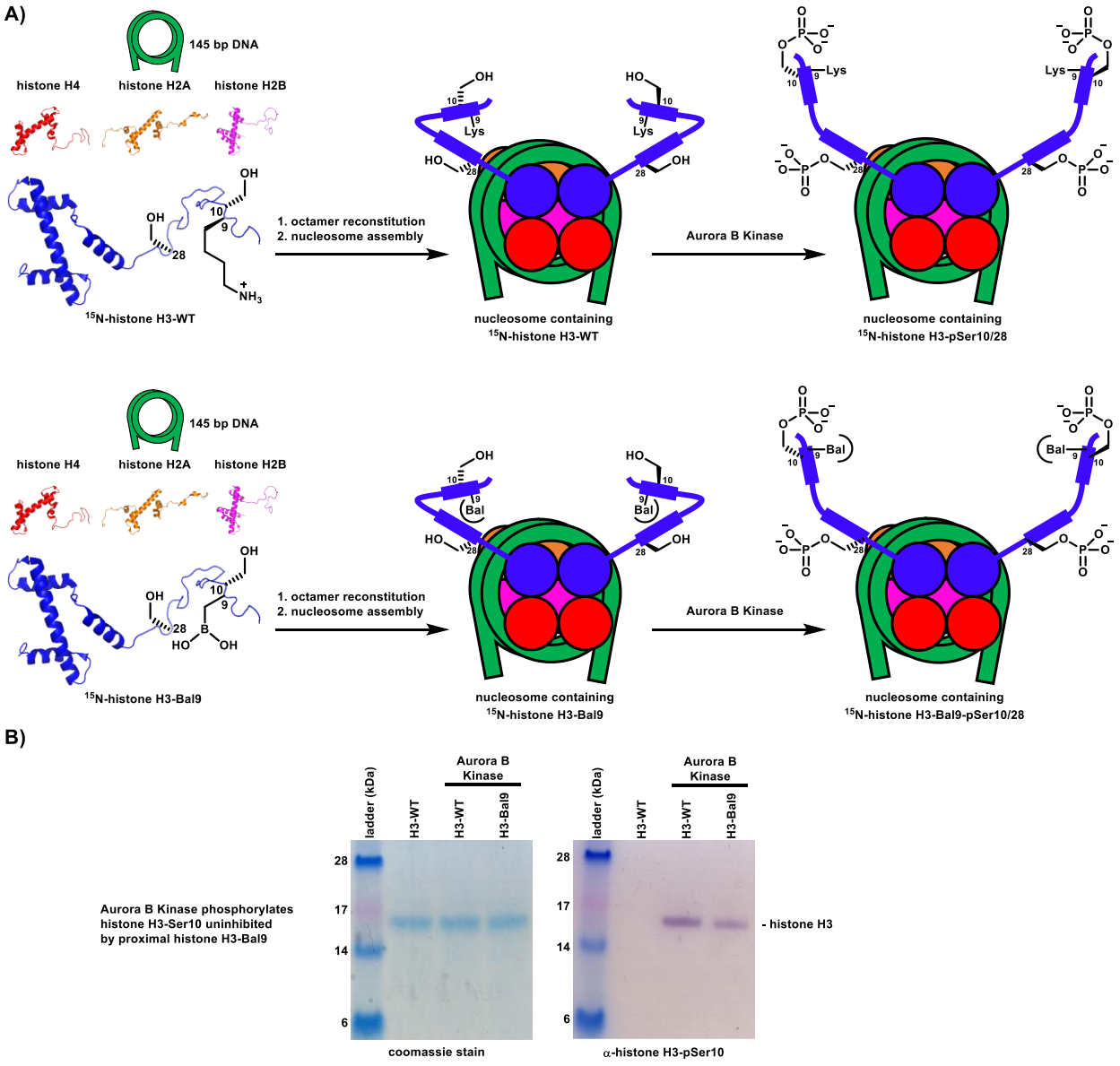


Figure 39: A) Construction of ^{15}N-histone H3 containing nucleosomes for structural comparisons after Aurora B Kinase phosphorylation utilizing Bal as an NMR probe via DICE. **B)** An α-histone H3-pSer10 western blot (100 ng histone per lane).

Structure insights into the histone tail in response to phosphorylation revealed by Bal mediated DICE NMR

We now had ^{1}H-^{15}N-HSQC NMR results for ^{15}N-histone H3-Bal9 and WT nucleosomes, before and after Aurora B Kinase phosphorylation (Charlie Buchanan from the Professor Andy Baldwin group performed and analyzed NMR results). First, we compared the spectra for unphosphorylated ^{15}N-histone H3-Bal9/WT by plotting the intensity ratios between them for each residue (**Figure 40A**). This revealed a 'double-dip' pattern of signal intensity loss along the histone tail, corresponding to the two transiently alpha helical regions doubling back on each other connected by a beta-turn, allowing Bal mediated PLABP engagement of the residues further up the histone tail and resulting in their repressed signal intensity. This matched the earlier data procured by Dr. Lukas Lercher (discussed above). Furthermore, Bal improved upon previous methods utilizing paramagnetic relaxation enhancement as it provided detailed short-range structural information. Notably, the beta turn at residues 10-12 (adjacent to Bal9) conveyed a characteristic 'low-high-low' pattern of signal intensity loss, indicative of backbone engagement allowed by a beta-turn configuration (see Extended Data Figure 9f in the corresponding publication for further analysis outside the scope of this Chapter)[221].

We next compared the spectra for Aurora B Kinase phosphorylated ^{15}N-histone H3-Bal9/WT by plotting the intensity ratios between them for each residue (**Figure 40B**). Surprisingly, the 'double-dip' pattern observed in the unphosphorylated sample comparison was greatly diminished, with only nearby residues to Bal9 experiencing a dip in signal intensity, an effect which decreased linearly up the histone tail. This indicates that the transient doubled-back alpha helical structure regions opened up, no longer

exposing Bal9 to residues further up the histone tail, an effect which to our knowledge has yet to be described in studies on Aurora B Kinase function on histones.

In conclusion, it was remarkable that the installation of Bal9 in place of Lys9 on histone H3 not only did not impede enzymatic phosphorylation of the proximal Ser10, but also still permitted detection of the pSer10 product with a specific monoclonal antibody. This exhibits the exceptional compatibility of Bal to be imbedded as a non-intrusive structural 'observer' within a highly intricate and complex environment such as different states of modified chromatin, as shown here with isotopically, chemically, and enzymatically modified nucleosomes investigated with Bal mediated DICE NMR.

This allowed here the detection and characterization of a previously uncharacterized structural consequence of Aurora B Kinase phosphorylation on the histone H3 tail. In particular, the crosstalk between histone H3-Me_3Lys9 (site of Bal installation) and pSer10 (produced by Aurora B Kinase) has been of high interest in epigenetics. Phosphorylation of Ser10 leads to inhibited recruitment of heterochromatin protein 1 (HP1), blocking chromatin condensation and regulating the cell cycle[56,73,279]. The traditional assumption was that pSer10 sterically or electronically obstructed HP1 binding to Me_3Lys9, but studies on peptide substrates have produced conflicting results[279,280], sometimes even showing greater binding with histone H3-Me_3Lys9-pSer10 than on Me_3Lys9 alone[281]. Our results show that Aurora B Kinase phosphorylation may have an overlooked effect on the transient structure of the histone H3 tail, highlighting a more complex structural influence that may lead to HP1 ejection rather than simply the occlusion of HP1 from methylated Me_3Lys9 by the adjacent pSer10.

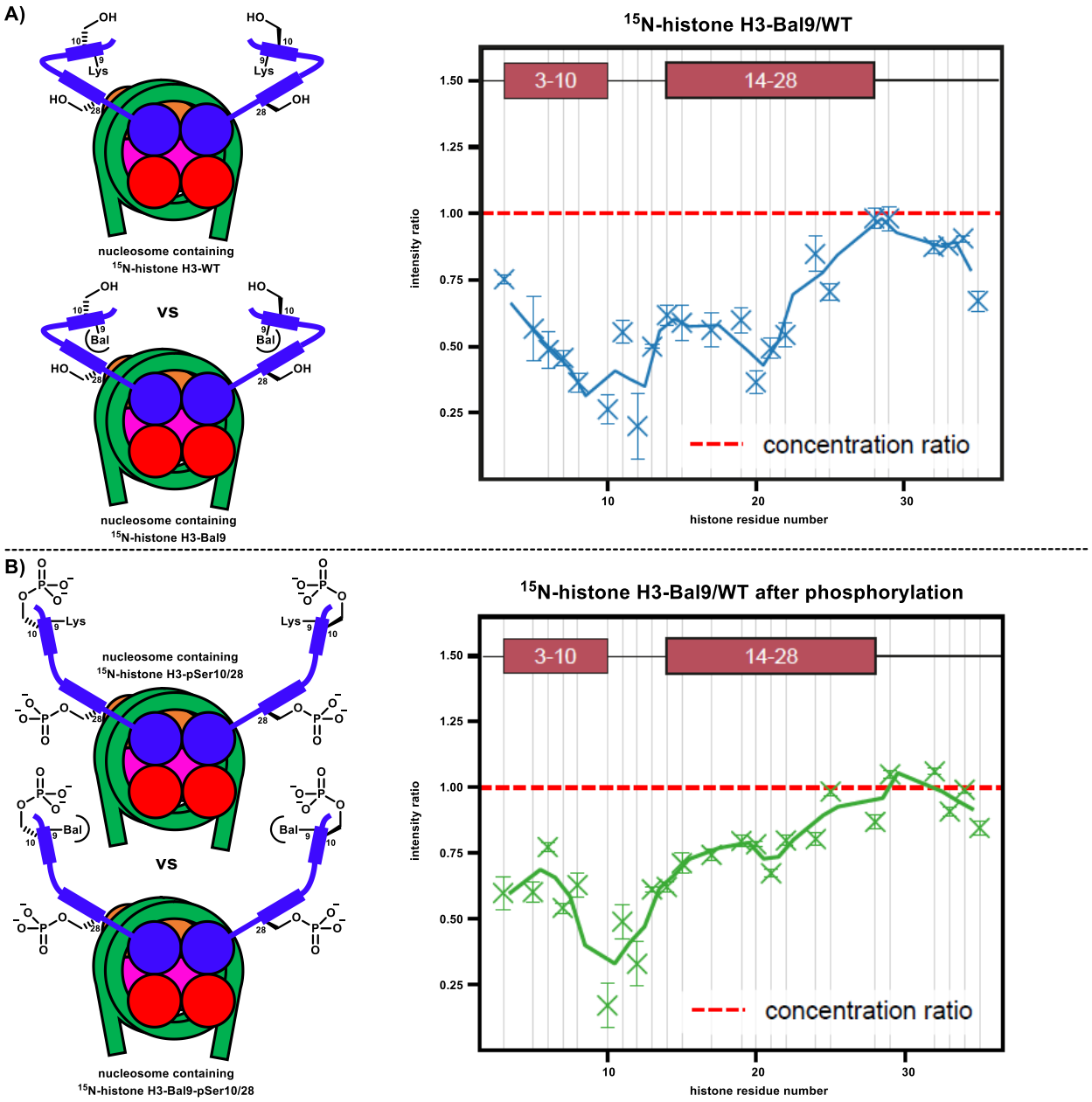


Figure 40: Comparison of histone H3 residue NMR intensities between nucleosomes containing ^{15}N-histone H3-WT/Bal9 **(A)** and nucleosomes containing ^{15}N-histone H3-WT/Bal9 after Aurora B Kinase phosphorylation **(B)**. Bal mediated backbone binding provides both short- and long-range structural information through loss of signal intensity (DICE). Blue boxes on histone tails indicate residual alpha helical structure. Figure was adapted with permission from data provided by Dr. Tim Mollner.

Investigating histone modifying enzyme function with installed native and unnatural histone PTMs

Investigating histone deacylation with Sirt2

Histone lysine acylation (especially non-acetyl acylation) is of growing interest in the field of epigenetics[28,282], with many recent studies focusing on the occurrence and function of non-enzymatic lysine acylation[283]. For example, a recent study identified the unnatural histone PTM, lysine benzoylation (BzLys), accumulating in chromatin after exposure to the common food preservative sodium benzoate, which non-enzymatically added to histone lysine residues after transforming to benzoyl-CoA[76]. Peptide studies and *in vivo* experiments revealed that BzLys caused differential expression changes compared to the analogous AcLys PTM and could be removed by the histone deacylase Sirt2[76,284]. We reasoned that our light-driven radical Dha reaction (developed in **Chapter 1**) was an ideal tool for creating full length histones with site-specifically introduced BzLys residues, allowing the study of the BzLys PTM in more depth.

To accomplish this, we chose to target the histone eH3.1-Lys18 site, a known Sirt2 deacylation site[284]. Installation of histone eH3.1-BzLys18, as well as an AcLys18 control, was carried out successfully using standard protocols developed in **Chapter 1**, with both yielding the desired modified histones with ~90% conversion (**Figure 41, top**). Both modified histones were then dialyzed into HEPES buffer (20 mM, pH 7.4), and Sirt2 was added along its required cofactor NAD^+ (0.5 µg Sirt2, 150 µM NAD^+, 20 µM histone, 50 µL volume, 37 °C, 600 rpm). Aliquots were taken at three time points (5 min, 30 min, 12 h) and immediately analyzed via intact protein LC-MS (**Figure 41, bottom**).

The results matched the previous findings that Sirt2 was able to deacylate BzLys, this time confirmed on a full-length site-specifically modified histone. However, while the AcLys removal had gone to completion within the first 5 min, BzLys removal took much longer, only reaching completion at the 12 h timepoint (**Figure 41**). Interestingly, both deacylation reactions halted at ~50% conversion to Lys18 (**Figure 41, green**). This is likely due to the unavoidable mixture of $_{L}$- and $_{D}$-sidechains installed via the light-driven radical Dha chemistry[208]. The remaining 50% of AcLys18 and BzLys18 is believed to be the $_{D}$-sidechain, indicating that Sirt2 is very specific for the $_{L}$-sidechains for both PTMs.

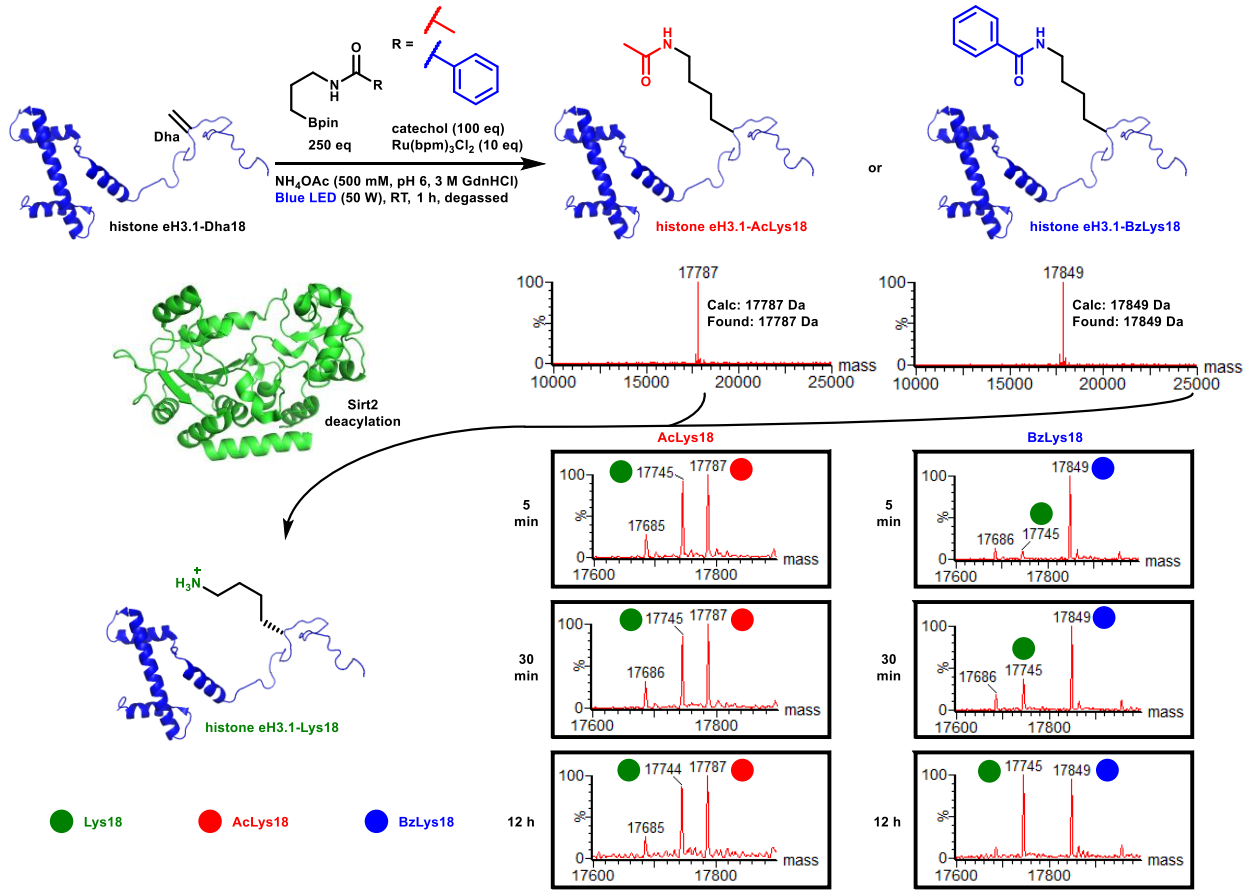


Figure 41: Creation of histone eH3.1-AcLys18 (red) and BzLys18 (blue), followed by deacylation with Sirt2. Intact protein LC-MS deconvoluted spectra are given for starting material and three deacylation time points, showing conversion of ~50% to Lys18 (green) for both PTMs, likely full conversion of the L-diastereomer. Histone eH3.1-BzLys18 deacylation was observed to be much slower than that of AcLys18. The unlabeled mass in the LC-MS spectra corresponds to residual Dha (17686 Da).

Production of histone H3-Me_2/Me_3Lys36 containing nucleosomes

Histone lysine methylation is a dynamic PTM that can serve different epigenetic functions depending on both its installation site and extent of methylation[15,54,61]. Case in point, histone H3-Lys36 has a specific methyltransferase, SET domain containing 2 (SETD2), that installs Me_3Lys36 on a Me_2Lys36 precursor[285]. Furthermore, mutation of SETD2 is implicated in a range of cancers[286], and deletion is embryonically lethal[287], suggesting a critical role of histone H3-Me_3Lys36 distinguishable from that of Me_2Lys36. A chromodomain protein linked to genome stability, MORF Related Gene on chromosome 15 (MRG15)[288,289], is a likely binder of histone H3-Me_3Lys36. MRG15 function is affected by upregulating SETD2 or the corresponding histone H3-Me_3Lys36 demethylase KDM4A and localizes to the same genomic sites as both SETD2 and histone H3-Me_3Lys36[288,290,291]. However, histone H3 tail peptides did not show discrimination of MRG15 binding between histone H3-$Me_{2/3}$Lys36. To further explore this phenomenon, we decided to synthesize nucleosomes containing histone H3-$Me_{2/3}$Lys36, hypothesizing that full nucleosome context may lead to MRG15 binding discrimination between histone H3-$Me_{2/3}$Lys36 due to structural consequences that were not revealed by peptides.

Histone H3-Cys36 was kindly provided by Dr. Sarah Faulkner (the leader of this project), and Dha formation was carried out as previously described[135]. As this project began before the light-driven radical Dha reaction (**Chapter 1**) was fully developed, we used our older method to install the desired methylated lysine PTMs[135]. Alkyl iodides corresponding to $Me_{2/3}$Lys sidechain radical precursors were initiated with sodium borohydride, successfully producing histone H3-$Me_{2/3}$Lys36 upon reacting with the corresponding Dha starting material (**Figure 42A**). The reaction was repeated to obtain ~2 mg of each

modified histone, followed by octamer reconstitution and nucleosome assembly (assisted by Dr. Sarah Faulkner). Three nucleosomes were created (**Figure 42B**), containing either histone H3-WT, Me_2Lys36 or Me_3Lys36, and wrapped by biotinylated 186 bp DNA. Characterization of the three nucleosomes by denaturing and native gel (**Figure 42C**) revealed expected bands similar between all three samples, suggesting that Lys36 methylation did not significantly affect octamer or nucleosome assembly or introduce structural aberrations. The nucleosomes were immobilized to beads (via the biotinylated DNA) and incubated with 6His-tagged MRG15 (hMRG15), followed by IP, elution, and western blotting against either hMRG15 (α-6His), or histone H3 as a control (α-histone H3) (**Figure 42D**, performed by Dr. Sarah Faulkner).

The results showed preferential enrichment of MRG15 with nucleosomes containing histone H3-Me_3Lys36, but not for histone H3-Me_2Lys36 or WT, a discrimination not observed with peptides or free histones[249]. This indicates that full nucleosome context allows MRG15 binding discrimination between histone H3-$Me_{2/3}$Lys36. Due to the proximity of the Lys36 site to the DNA backbone, we reasoned that differential hydrogen bonding interactions between the methylated Lys residues and the phosphate groups in the DNA backbone may lead to differences in Lys sidechain accessibility, an effect that would not be observable in peptide or lone histone studies. Molecular dynamics simulations (performed by Tom Watts) supported this hypothesis, with histone H3-Me_2Lys36 capable of forming H-bonds to the DNA backbone and becoming buried, while histone H3-Me_3Lys36 was incapable of forming H-bonds, remaining more solvent exposed (**Figure 42B**). This matched the binding experiment findings and suggested a new mechanism for differential Lys methylation function based on DNA H-bonding.

While this investigation is still ongoing and will benefit from structural data of the MRG15-histone H3-$Me_{2/3}$Lys36 nucleosome interaction, it clearly exhibits the need for histone PTM studies to utilize appropriately complex substrates of choice. The discrimination between histone H3-$Me_{2/3}$Lys36 H-bonding interactions with the DNA backbone was not revealed in peptide or free histone binding experiments, while a full nucleosome properly provided the context needed to distinguish the subtle structural and interactive differences leading to MRG15 recruitment preferentially to histone H3-Me_3Lys36.

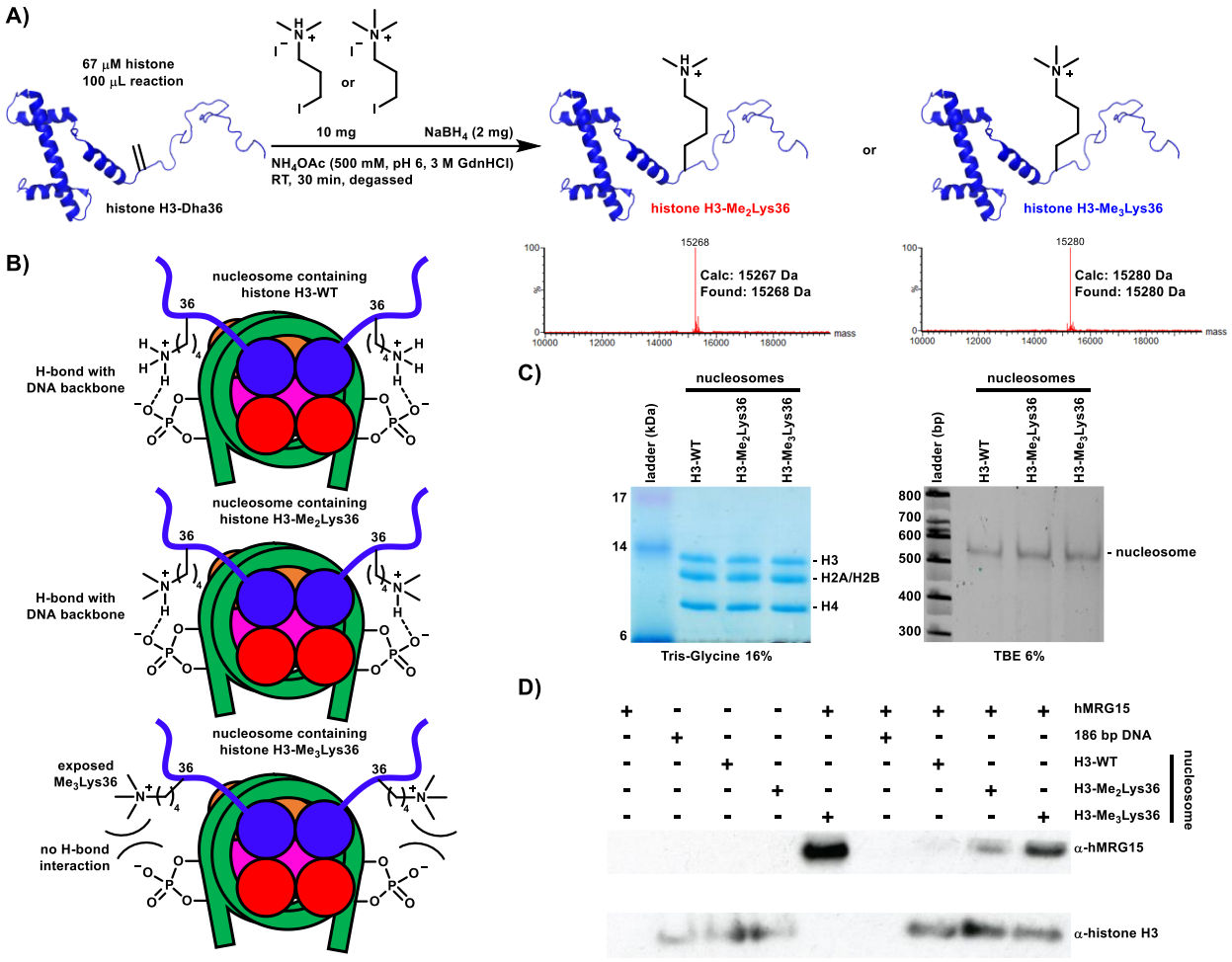


Figure 42: A) Creation of histone H3-Me$_2$Lys36 and Me$_3$Lys36, with intact protein LC-MS deconvoluted spectra of the products. **B)** Representation of the three nucleosomes created for this study, indicating the H-bonding interactions with the DNA backbone, or lack thereof for histone H3-Me$_3$Lys36, that led to shielding or exposure to MRG15 interaction. **C)** Characterization of the nucleosomes by denatured (left) and native (right) gels. **D)** Enrichment of hMRG15 after incubation with bead-bound nucleosomes (via their biotinylated 186 bp DNA) was preferential with histone H3-Me$_3$Lys36 over Me$_2$Lys36 and WT, revealed by western blot. Figure was adapted with permission from data provided by Dr. Sarah Faulkner.

Tracking PTM status, enzyme processing, and complex formation with γCF_2-containing PTM sidechains

Previous work within the group

During the entire course of the development and optimization of the light-driven radical Dha modification reaction described in **Chapter 1**, colleague Dr. Patrick Isenegger was developing another light-driven strategy for Dha modification in tandem. This complementary method, published in the same manuscript as the aforementioned reaction[208], resulted in the difluoroalkylation of Dha residues in proteins (**Figure 43A**). It utilized mono- or difluoroalkylpyridylsulfone derivatives and bromo-mono- and bromo-difluorocarbonyls as radical precursors, which could be reductively activated upon blue light illumination by the same photocatalyst used in method described in **Chapter 1**, $Ru(bpy)_3Cl_2$. The additive iron sulfate both promoted the initial reduction of the Ru photocatalyst to the required Ru^I state and facilitated reduction of the resultant α-carbon radical after difluoroalkyl radical addition to Dha, producing protein sidechains with a γCXF-R (X = H or F) group (**Figure 43B**).

There were multiple advantages to this Dha fluoroalkylation reaction. It was efficient, often requiring only 2 equivalents of the radical precursor, and a similar amount of $Ru(bpy)_3Cl_2$ photocatalyst. Additionally, as with the previously described reaction, it was tolerant towards a wide range of chemical functionality. Many native, PTM containing, and unnatural reactive sidechains were easily installed into proteins in this manner, accessing fluorinated derivatives and PTMs of Lys, Arg, Glu, Gln, and Met amino acids, as well as allowing the incorporation of functional groups such as halogens, azides, other radical precursors, and many oxidation states of sulfur (sulfide, sulfoxide, sulfone, and sulfate)

(**Figure 43**). While the development of this reaction was not part of the work in this thesis, the fluorinated protein products were used extensively in applications in this section. The installed γCF_2 group provided a unique and minimally sized tag that not only allowed ^{19}F-NMR tracking of specific sidechains and their PTM status (discussed below).

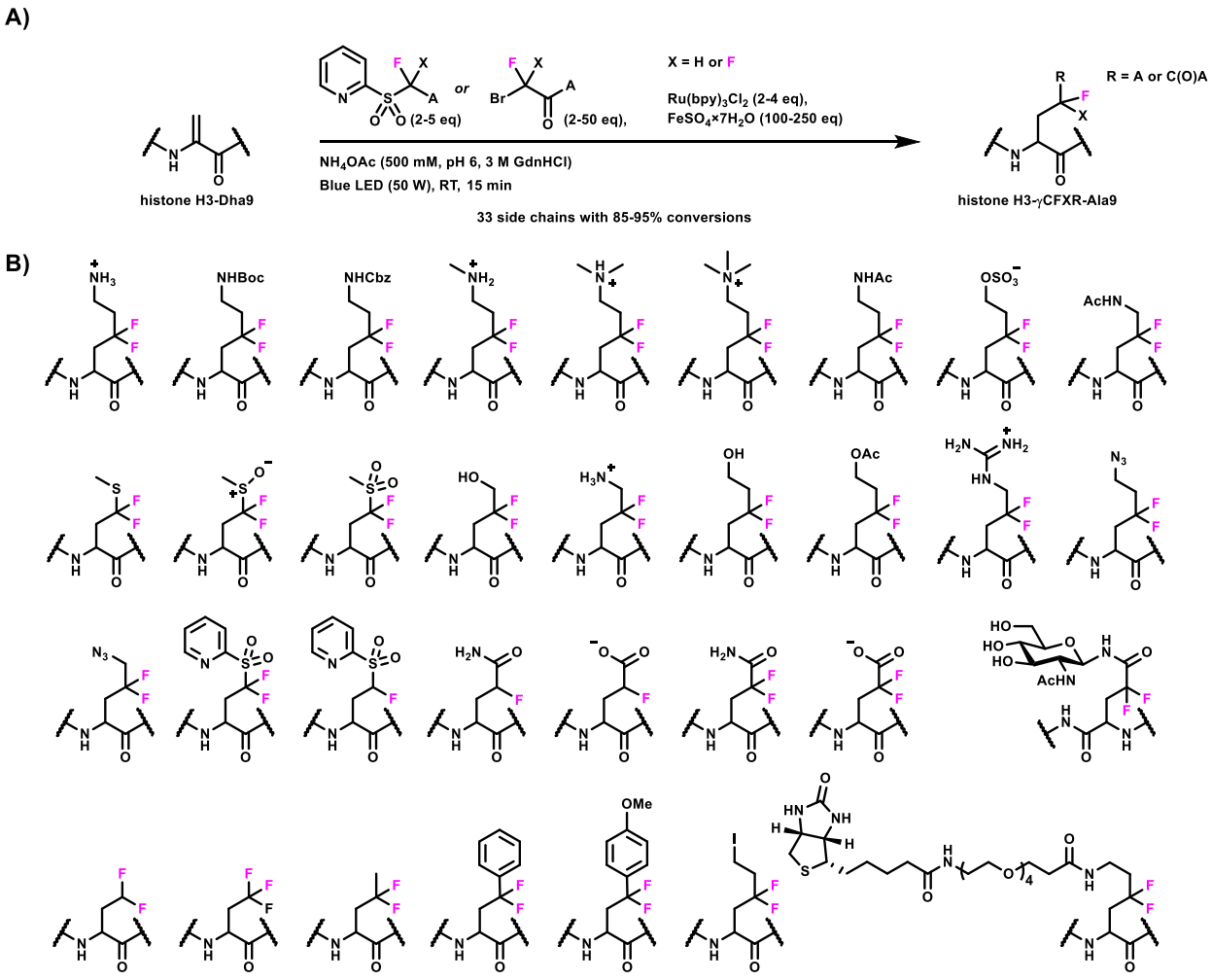


Figure 43: A) General reaction conditions for mono- and difluoroalkylation of Dha residues in proteins. **B)** Sidechain scope for the reaction described in **(A)**.

Rationale for incorporating fluorine into proteins for ^{19}F-NMR

While there are many methods for site-specifically installing PTMs on to histones[94,95], a major limitation still exists for studying the PTM status of specific residues in native contexts. In nature, PTMs are dynamic[51], being installed and removed as needed to generate the transcriptional landscape required for the cell. Tools for unambiguously detecting these minute molecular changes (in the case of methionine oxidation, as small as a single oxygen atom) in solution and in real time are extremely lacking. For example, two of the most common methods for detecting the PTM status of specific sidechain on histones (or other proteins), LC-MS/MS[292] and antibody binding[293], are either destructive to the sample (in the case of LC-MS/MS), or occlude the relevant PTM site throughout the binding event (in the case of antibodies), eliminating its ability to bind to its interaction partners or be processed by relevant epigenetic enzymes. Furthermore, when more complexity is introduced, such as when proximal residues Lys9 and Ser10 on histone H3 are trimethylated and phosphorylated, respectively, the modification of one residue can restrict antibody detection of the other[89].

An ideal chemical tool for detecting the PTM status of a specific sidechain would fulfil the following requirements: 1) easy to access and introduce into full length histone proteins, 2) amenable to a wide variety of native and modified protein sidechains, 3) highly sensitive to changing chemical environments (such as binding events or PTM processing), 4) small enough to avoid interfering with native PTM function and structure, 5) non-destructive detection in real time with high signal to noise ratio in the presence of other biomolecules. We rationalized that our light-driven radical Dha reaction allowing installation of γCF_2-containing sidechains (**Figure 43**)[208] would fulfill these requirements when paired with the

nondestructive technique of ^{19}F-NMR. The γCF$_2$-sidechain installation method on Dha containing histones is fast, operationally simple, and utilizes minimal amounts of readily accessible difluoropyridylsulfone sidechain precursors that encompass a plethora of relevant histone PTMs[208]. Furthermore, ^{19}F is the third most receptive NMR nucleus (behind only ^{1}H and ^{3}H), highly sensitive to its environment with a large chemical shift dispersion, and with a comparable size to the hydrogen atoms it replaces at the gamma carbon[294]. Additionally, the near absence of fluorine in biology ensures an exemplary signal to noise ratio, even in complex biological environments.

This section will utilize γCF$_2$-containing sidechains in various applications measured by ^{19}F-NMR: tracking enzymatic histone PTM processing (deacetylation of AcLys by Sirt2), distinguishing different PTMs on the same sidechain (different oxidized Met PTMs), and following protein complex formation (histone octamer reconstitution) (**Figure 44**). This section was performed in partnership with Dr. Patrick Isenegger, who developed the Dha difluoroalkylation reaction and advised the various applications.

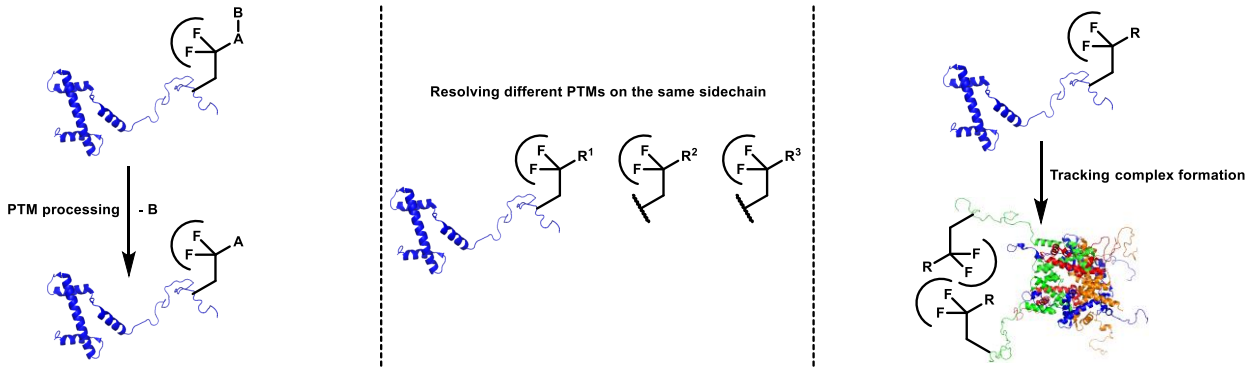


Figure 44: Three types of histone related applications measured by ^{19}F-NMR that are tested in this section.

Tracking histone PTM processing via ^{19}F-NMR

We chose to test ^{19}F-NMR tracking of sidechain PTM removal with the same model system used in the previous section, deacetylation of histone eH3.1-AcLys18 with Sirt2[284]. We first synthesized histone eH3.1-γCF_2-AcLys18 (**Figure 45A, blue**) and γCF_2-Lys18 (**Figure 45A, green**) from Dha18 (1.5 mg scale), utilizing the relevant difluoroalkylpyridylsulfone reagents (5 eq), achieving nearly full conversion as observed by intact protein LC-MS deconvoluted spectra (**Figure 45A**). ^{19}F-NMR measurements were then collected, providing the starting material and product reference spectra for Sirt2 deacetylation (**Figure 45B, right-top and -middle**). Then, the histone eH3.1-γCF_2-AcLys18 sample was subjected to a scaled up Sirt2 deacetylation (histone 625 µg, Sirt2 100 µg, NAD^+ 150 µM, 1 mL, HEPES 50 mM, pH 7.5, 30 °C, 1 h) followed by ^{19}F-NMR measurement (**Figure 45B, right-bottom, purple**).

Pleasingly, ^{19}F-NMR spectra for the starting material and product references (**Figure 45B, right-top and -middle**) were clearly distinguishable, giving credence to the sensitivity of the ^{19}F nucleus, which could detect the presence or absence of the acetyl PTM five bonds away. Furthermore, both $_L$- and $_D$- diastereomers, characteristic of Dha modification chemistry[127], were clearly revealed in the spectra after simulation to determine the relevant resonance frequencies and coupling constants (performed by Professor Andrew Baldwin). The ^{19}F-NMR spectra collected from the Sirt2 reaction (**Figure 45B, right-bottom**) gave a mixture of peaks corresponding to those measured for both the histone eH3.1-γCF_2-AcLys18 and γCF_2-Lys18, evident when the spectra were overlayed (**Figure 45B, left**). This spectrum was again simulated, taking the starting material and product species with scaled intensities. This revealed that only the $_L$-γCF_2-AcLys18 reacted with

Sirt2, selectively producing L-γCF2-Lys18 (**Figure 45B**). These results indicate that γCF2-sidechains are ideal probes to tracking histone PTM processing. In this example, they not only did not interfere with native Sirt2 function, but could easily distinguish removal of Lys acetylation, as well as resolve the selectivity of Sirt2 for L- over D-AcLys[208].

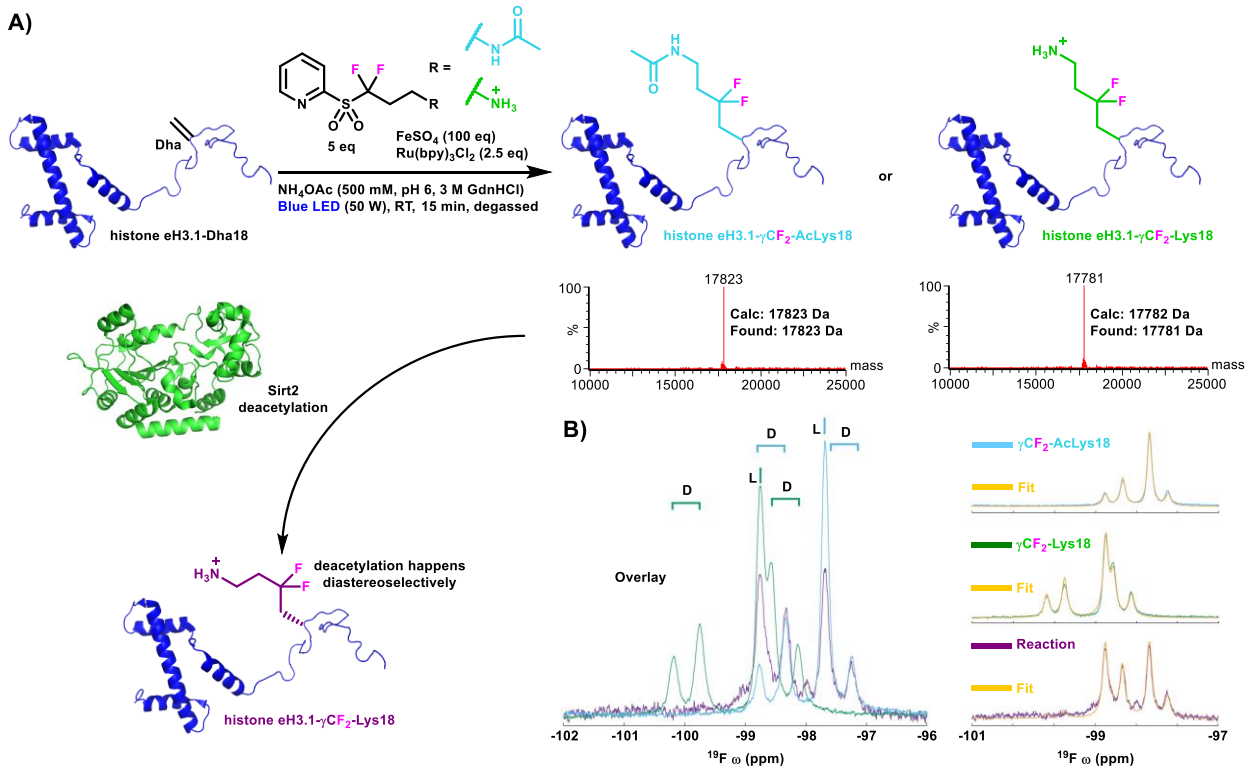


Figure 45: A) Reaction conditions for creation of histone eH3.1-γCF2-AcLys18 (blue) and control histone eH3.1-γCF2-Lys18 (green). Sirt2 deacetylates histone eH3.1-γCF2-AcLys18 to give purely L-γCF2-Lys18 (purple). **B)** ^{19}F-NMR of histone eH3.1-γCF2-AcLys18 and γCF2-Lys18 gave clearly distinguishable spectra. ^{19}F-NMR of the Sirt2 reaction shows shifting of half the γCF2-AcLys18 peaks to the positions of half of the measured γCF2-Lys18 peaks, indicating selective Sirt2 deacetylation of L-γCF2-AcLys18, creating L-γCF2-Lys18. Fitting was performed by Professor Andrew Baldwin.

Significance of methionine as oxidation PTM site and relevant oncohistone mutation

Oncohistones, mutated histones with amino acid changes that drive the onset of cancer through changing the physical properties or enzymatic interactions of their parent histone, have been of growing focus in the epigenetics community[295,296]. In one of many examples, it is known that a simple Lys to Met mutation in histone H3 at site 27 causes enhanced binding to its methyltransferase, Polycomb Repressive Complex 2 (PRC2), sequestering it and inhibiting its native function, leading to pediatric glioblastoma (**Figure 46A**)[74,297]. Post-translational modifications on these mutant residues, such as Met oxidation and others, while likely relevant, have been lacking in study.

As described in the introduction, it is well known that the redox environment in cells can drive PTM formation on a variety of protein residues. While the presence of reactive oxygen species (ROS) can certainly cause wanton damage to proteins by oxidizing various amino acids, it is also known that many residues can mediate the redox environment through promoting their own oxidation[22]. Furthermore, specific enzymes exist that can add or remove these oxidative PTMs. This is perhaps best exemplified by Met, which can be oxidized to the corresponding sulfoxide, and more rarely, to the sulfone, in the presence of ROS[35,36] (**Figure 46B**). Methionine sulfoxide reductase enzymes not only are specific to each diastereomer of the methionine sulfoxide but can also catalyze the reverse reaction in the right conditions[35,39,298], hinting at the level of precision in which methionine oxidation is regulated in living systems. However, this reversibility, the small atomic distinction between oxidation states (a single oxygen atom) and methionine sulfoxide diastereomers, and the susceptibility of methionine to be

oxidized during sample processing makes this residue and its oxidative PTMs extremely difficult to study.

We reasoned that, with our Dha difluoroalkylation method in hand, we could site-specifically install γCF_2-Met variants of each oxidation state. The γCF_2 group would allow tracking of the PTM status via ^{19}F-NMR in real time, perhaps with sensitivity that would allow distinction between oxidation states as well as sulfoxide diastereomers.

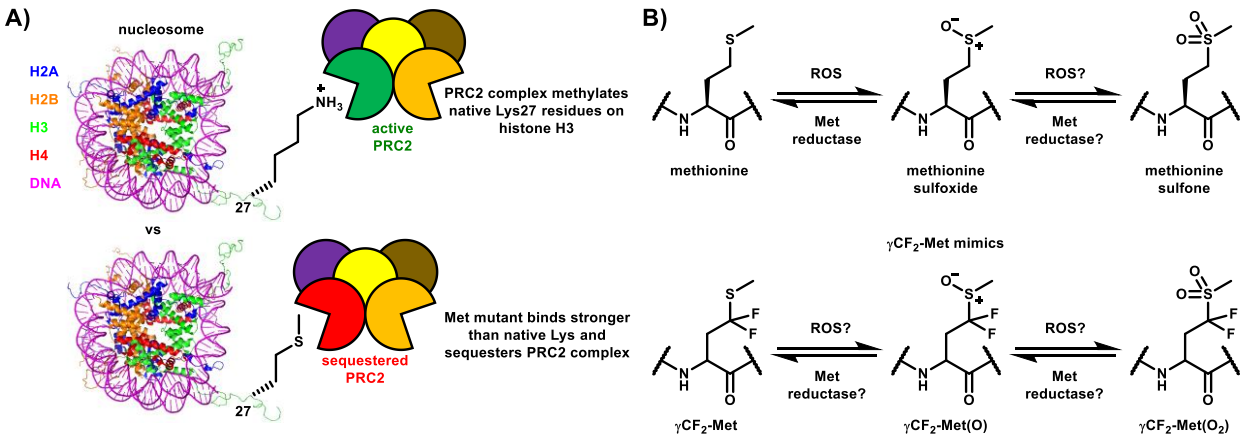


Figure 46: A) The histone H3-Lys27 to Met27 oncohistone mutation binds and sequesters the PRC2 methyltransferase complex. **B)** Schematic illustrating the reversible oxidation states of Met. Installation γCF_2-Met oxidation state mimics may allow tracking of Met oxidation PTMs via ^{19}F-NMR.

Distinguishing Met oxidation PTMs via ^{19}F-NMR

We reasoned that γCF_2-Met, along with its oxidized variants γCF_2-Met(O) (Met sulfoxide) and γCF_2-Met(O_2) (Met sulfone) would be useful ^{19}F-NMR probes for distinguishing the different oxidation states of Met on the same sidechain. Early attempts installed the γCF_2-Met derivatives on the oncohistone site histone H3-Lys27[74,297], but interactions with the nearby folded histone core made interpreting the ^{19}F-NMR spectra difficult (data not shown). While these results were intriguing on their own and provide an avenue for future histone H3-Met27 oncohistone studies, we decided instead to first assess the γCF_2-Met variants in a cleaner setting.

We chose to install γCF_2-Met, γCF_2-Met(O), and γCF_2-Met(O_2) onto the lesser studied histone H3-Lys4 site[299], figuring that the terminus of the disordered histone tail, along with conducting ^{19}F-NMR measurements in denaturing conditions, would produce the cleanest ^{19}F-NMR spectra without structural interference. Large scale installation on histone eH3.1-Dha4, using standard Dha difluoroalkylation conditions (**Figure 43A**), went to nearly full conversion for each γCF_2-Met variant as assessed by intact protein LC-MS (**Figure 47A**), yielding 4 mg of histone after buffer exchanging to the required ^{19}F-NMR buffer (NH_4OAc 250 mM, pH 7, 3 M GdnHCl, prepared in D_2O). SDS-PAGE analysis also revealed the expected bands for each modified histone, at the same size as Cys4 and Dha4 precursors (**Figure 47A**). The modified histones then had their ^{19}F-NMR spectra recorded (600 MHz, 2800 scans, 12 h) (**Figure 47B**).

The ^{19}F-NMR measurements produced easily distinguishable spectra for each γCF_2-Met variant, with both diastereotopic fluorine atoms, both diastereomers ($_L$- and $_D$-sidechains), and both sulfoxide epimers (for γCF_2-Met(O)) clearly accounted for (**Figure 47B**). Histone

eH3.1-γCF_2-Met4 produced one singlet and one quartet, arising from two environments for the γCF_2 group, one strongly coupled (quartet) and the other with chemical shifts of the two fluorine atoms too close to be resolved (singlet). Histone eH3.1-γCF_2-Met(O)4 gave the most complex spectra containing four γCF_2 species, as expected due to the addition of the two sulfoxide epimers, shifted upfield relative to γCF_2-Met4 due to the additional oxygen. Interestingly, two sets of peaks gave a relative intensity ratio of ~1:0.4 compared to the other two, indicating preferential $_D$- vs $_L$- sidechain installation, although we did not resolve between the two. This effect was also seen with histone eH3.1-γCF_2-Met(O_2)4, which gave two strongly coupled doublets at a ~1:0.4 relative intensity ratio. In conclusion, all three oxidized γCF_2-labelled Met sidechains were easily installed and measured by ^{19}F-NMR, producing easily distinguishable spectra by their chemical shifts and splitting.

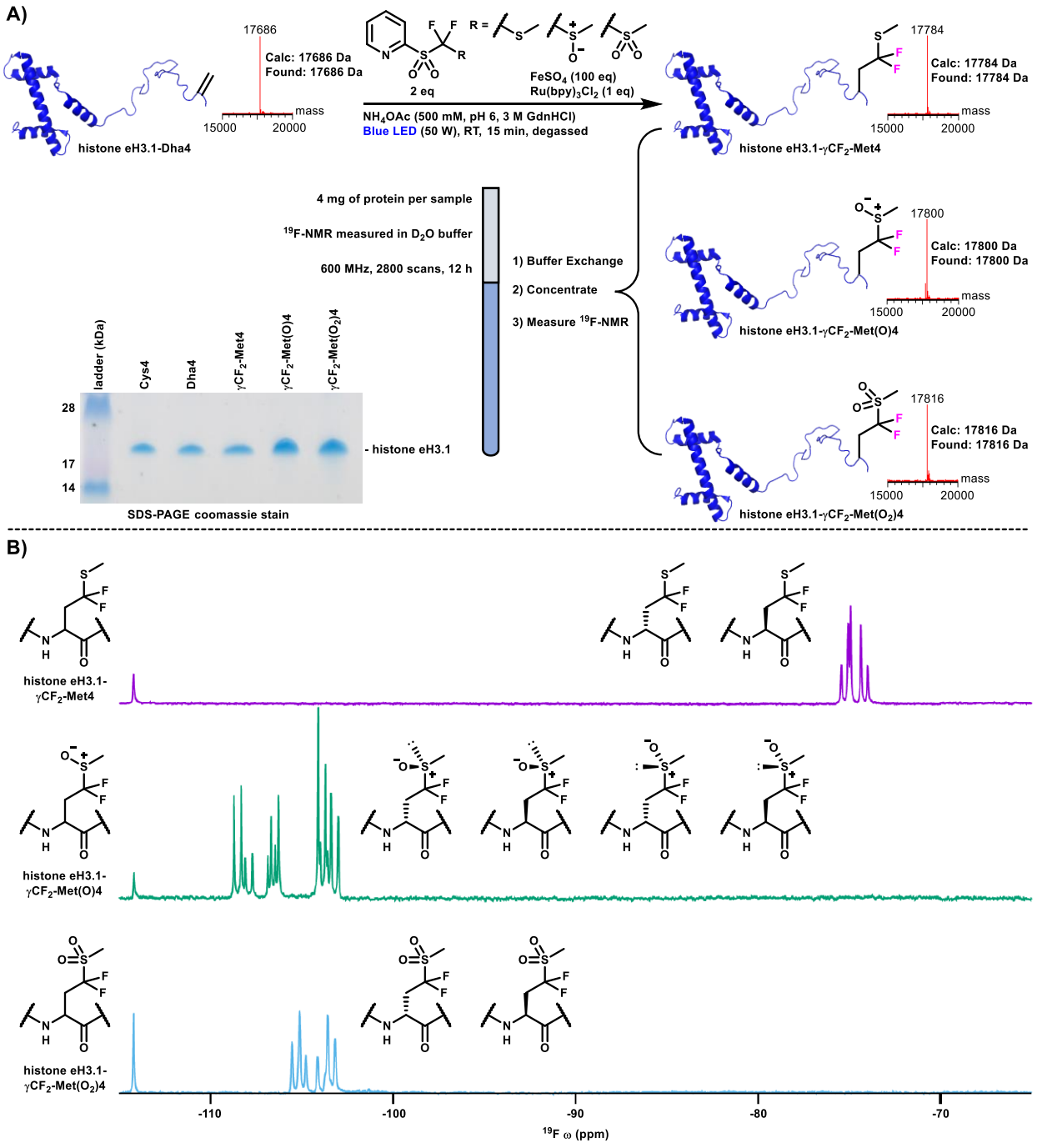


Figure 47: A) Details for the synthesis and ^{19}F-NMR measurement of histone eH3.1-γCF$_2$-Met4, γCF$_2$-Met(O)4 and γCF$_2$-Met(O$_2$)4, accompanied by intact protein LC-MS deconvoluted spectra and SDS-PAGE characterization. **B)** ^{19}F-NMR spectra of the modified histones from **(A)**, with structures of expected species inset.

Following protein complex (histone octamer) formation via ^{19}F-NMR

Lastly, we reasoned that the γCF_2 containing sidechains would allow for the modified proteins to be tracked by ^{19}F-NMR throughout their incorporation into multiprotein complexes. To test this, we investigated histone H3 as a denatured and folded free histone, followed by tetramer formation with histone H4, then octamer formation with histones H2A and H2B (**Figure 48A**).

The smallest Dha difluoroalkylation substrate, γCF_2-hAla, was installed onto histone H3-Dha9 (**Figure 48A**), and ^{19}F-NMR spectra were recorded for the histone in both folded and unfolded states (see **Experimental Details** for free histone, tetramer, and octamer formation and ^{19}F-NMR measurement details). Then histone tetramer and octamer reconstitutions were performed via literature known protocols[242,300], with ^{19}F-NMR measurements after each step. Both the histone tetramer and histone octamer were first purified by size exclusion chromatography (Superdex S75, 16/60) and analyzed via SDS-PAGE gel before ^{19}F-NMR measurements were taken. An intact protein LC-MS deconvoluted spectra of the final octamer confirmed the presence of all four histones, and retention of the γCF_2-hAla modification on histone H3 (**Figure 48A**).

The ^{19}F-NMR results for all four measured species (**Figure 48B**) were quite similar, which may be expected due to the γCF_2-hAla position on the end of the disorder histone H3 tail. Subtle differences in the chemical shifts between species, such as slight downfield shift of the folded histone H3 in contrast to the other three species, may be due to different tumbling rates of the protein complexes. In hindsight, future endeavors using this approach to track protein complex formation should install the γCF_2-sidechain closer to folded regions or interaction areas during the protein complex formation, which would

likely generate greater differences in the ^{19}F-NMR spectra due to the sensitivity of the fluorine nucleus to its environment[294]. Nevertheless, we were pleased in the sensitivity and noise free ^{19}F-NMR results, especially since the final octamer consisted of less than 1 mg of material, indicating that γCF_2-sidechains are extremely sensitive probes requiring far less material for protein NMR experiments[294].

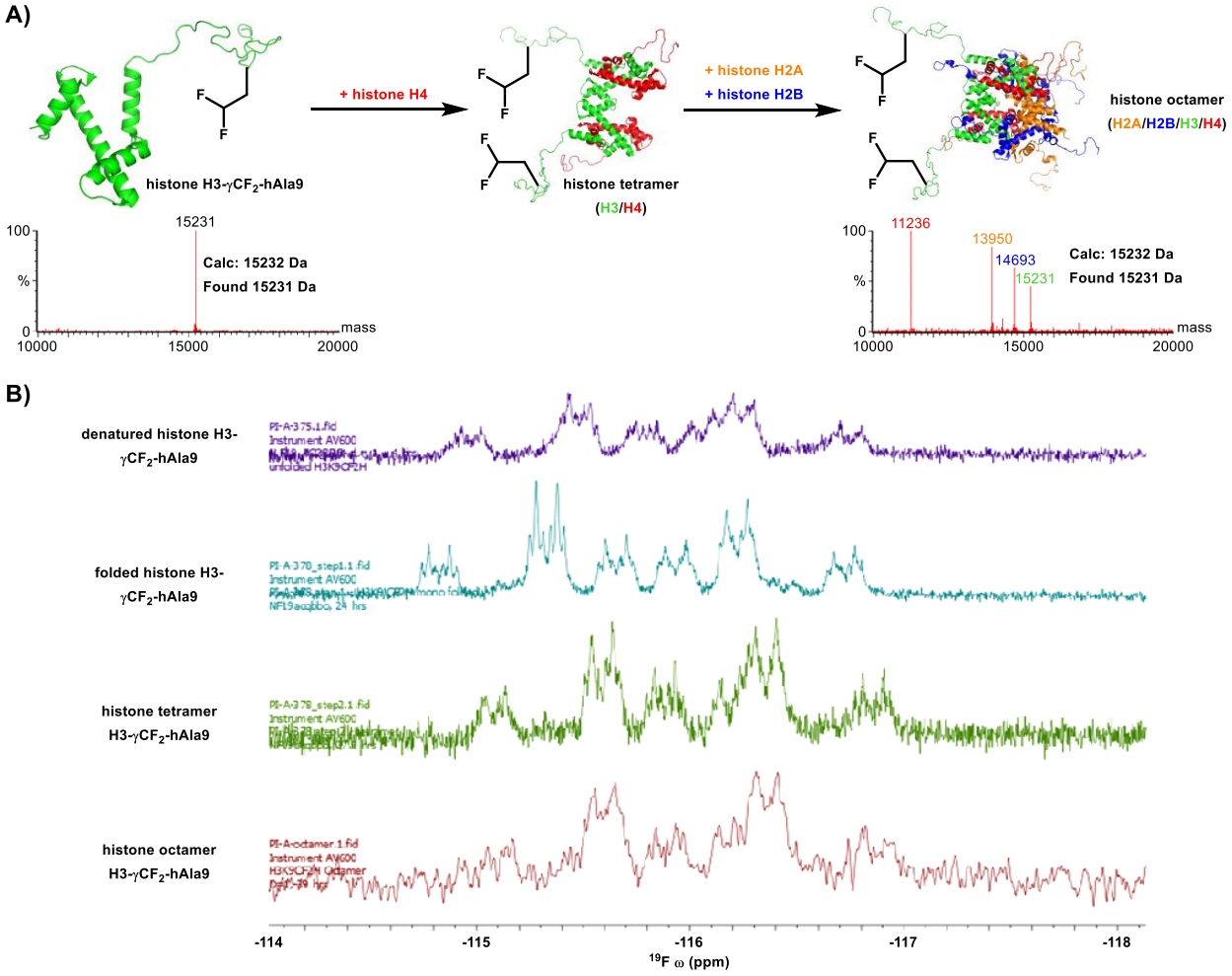


Figure 48: A) Representation of histone H3-γCF_2-hAla9 formed into a histone tetramer and octamer for ^{19}F-NMR measurement, with intact protein LC-MS deconvoluted spectra for the histone and octamer. **B)** ^{19}F-NMR spectra of the species in **(A)**, with histone H3-γCF_2-hAla9 measured both in a denatured and folded state.

Conclusions

Summary and advantages brought about by this chapter

This chapter consisted of an assortment of histone and epigenetics focused applications utilizing recent protein modification methods developed in our group, such as the light-driven radical Dha modifying reaction developed in **Chapter** 1[208], the corresponding Dha difluoroalkylation reaction[208], an older radical Dha modifying reaction from our group[135], and the recently published Dha borylation reaction[221]. The motive for this chapter was two-fold: testing the compatibility of our protein modification strategies in novel ways and on increasingly complex epigenetics substrates (such as nucleosomes or nuclear extracts) and characterizing important epigenetic processes and mechanisms through the applications afforded by our protein modification techniques. This chapter consisted of four independent sections that each ventured to fulfil these motives.

The first section was an extension of the section in **Chapter 1** that investigated the reactivity of electrophilic sidechains installed with our light-driven radical Dha modification reaction[208]. In this section, the use of electrophilic sidechain bromohomonorleucine (Bhn), mimicking the connectivity and structure of a methylated Lys residue, was installed onto various histone sites and used to crosslink the histone demethylase KDM4A[261] by alkylating the Zn-binding Cys residues near its active site. This crosslinking was confirmed by SDS-PAGE, western blot, Zn ejection assay, and LC-MS/MS, and was seen to be proximity-driven, as the sidechain radical precursor to Bhn did not promote the KDM4A Cys alkylation, lacking the effective molarity enhancement given by the histone recognition. Furthermore, incubation of histone eH3.1-Bhn9 in HeLa cell nuclear extract produced many crosslinked products, as revealed by western blot, hinting that this

method is compatible with complex chromatin substrates and could potentially identify unknown histone PTM interactors in future applications.

The second section utilized the unique chemical properties of the Boronoalanine (Bal) sidechain installed via the borylation of Dha residues in proteins[221]. Bal was capable of oxaborolane formation with proximal amide backbone carbonyls through PLABP, repressing the ^{1}H-^{15}N-HSQC NMR signal intensity for sites near the installed Bal residue. This was used to investigate the histone H3 tail structure in response to phosphorylation by Aurora B Kinase on a full nucleosome, by installing Bal as a DICE NMR probe on an adjacent residue (Lys9) to phosphorylation (pSer10). Results showed that two regions of residual alpha helical structure doubled back on each other in the unphosphorylated histone H3 tail, but opened in response to phosphorylation, revealed by the changes in NMR signal intensity losses for each histone tail residue due to Bal mediated PLABP. This has implications for the mechanism of inhibited binding of HP1 to histone H3-Me_3Lys9 in the presence of pSer10[73,279,280], which may be due in part to these subtle histone tail structural changes in response to phosphorylation.

Section three utilized two radical Dha modification methods[135,208] to install natural and unnatural histone PTMs. First, the unnatural histone PTM, benzoyllysine[76], was compared to natural PTM acetyllysine in its ability to be deacylated by Sirt2 on histone H3-Lys18. This revealed slower rates of Sirt2 deacylation for BzLys compared to AcLys for the first time on a full-length site-specifically modified histone protein. In addition, the mechanism of MRG15 binding discrimination between histone H3-$Me_{2/3}$Lys36 was investigated in a full nucleosome context utilizing our group's previous radical Dha modification chemistry[135] to install the methylated lysine residues. This revealed that

MRG15 preferentially bound Me_3Lys36 over Me_2Lys36 in a nucleosome, but not in peptide or free histone experiments. We reasoned, with further insight from molecular dynamics simulations, that this was due to the differential hydrogen bonding ability between $Me_{2/3}$Lys36 to the DNA backbone in the nucleosome, sequestering Me_2Lys36 while exposing Me_3Lys36[249], revealing the importance of picking the appropriately complex epigenetic substrate when studying epigenetic mechanisms.

The fourth section explored the use of γCF_2-sidechains, installed by the light-driven Dha difluoroalkylation reaction[208], as site-specific 'on-protein' ^{19}F-NMR probes for a variety of applications. First, the deacetylation of histone H3-γCF_2-AcLys18 by Sirt2, selectively producing $_L$-γCF_2-Lys18, was resolved by ^{19}F-NMR. Next, the different oxidation states of Met (sulfide, sulfoxide, and sulfone) were installed on the oncohistone site histone eH3.1-Lys4[299], with ^{19}F-NMR distinguishing not only the oxidation state of Met, but even the two epimers of the Met sulfoxide. Lastly, histone H3-γCF_2-hAla9 was reconstituted into histone tetramers and octamers, with ^{19}F-NMR successfully identifying it in the multiprotein complexes, even at sub-milligram scales. This section revealed that γCF_2-sidechain enabled ^{19}F-NMR allows sensitive and noise free tracking of specific sidechains in proteins, allowing their dynamic modification states to be determined.

Prospects for future work

The first lesson learned from the work in this chapter is that increasing epigenetic complexity reveals unexpected epigenetic mechanisms that would not be possible to observe in simpler systems. In both studies with intact nucleosomes (MRG15 recognition of histone H3-Me$_{2/3}$Lys36 and Bal enabled DICE NMR of histone tail residual structure after Aurora B Kinase phosphorylation), subtle structural effects in intact nucleosomes led to the observed functions. There are many chemical methods for creating desired modified histones[94,95], but every effort should be made to study the modified histones in the most complex and realistic setting possible allowed by the method. New protein modification methods should spend as much time assessing their method in complex biological settings as developing the chemistry for the method.

Second, the proximity-driven protein-protein crosslinking enabled by installed electrophilic sidechains such as alkyl halides has yet to reach its full potential[208]. While many protein-protein crosslinking strategies exist[240], there are few straightforward methods that can mimic the PTMs of different sites to drive recognition and capture of interaction partners. HeLa cell nuclear lysate incubations with histone H3-Bhn9 revealed many unknown interaction partners by western blot, and it would not be difficult to follow up these results with proteomics to identify not only the interaction partner identities but also their crosslinking sites, providing insight into their binding mechanisms.

Lastly, the ^{19}F-NMR tracking of γCF$_2$-sidechains installed by the light-driven radical Dha difluoroalkylation reaction is extremely promising. Histone PTMs (and PTMs of other proteins) are dynamic, and this ^{19}F-NMR probe is the perfect reporter for the PTM status at a specific site of interest. The rarity of fluorine in biology, its similarity to hydrogen in

size, and its attributes as an NMR active nucleus gave it exquisite intensity, sensitivity, and signal-to-noise ratios in our ^{19}F-NMR measurements. Future studies should leverage this γCF_2-sidechain enabled method, either through the protein modification chemistry used here[208], or even global proteome incorporation via feeding in precursor γCF_2-amino acids into cells, to non-destructively and sensitively identify and quantify the PTMs of specific sidechains in more complex biological settings.

Statement on collaboration

The work outlined in this thesis chapter was done with the assistance of many others, who both performed some of the work before I started the project and assisted with various aspects throughout the course of the work. Their contributions are explicitly stated in the text when appropriate and outlined in full here.

Crosslinking studies with alkyl halide containing histone sidechains was made possible due to the LC-MS/MS analysis performed by Dr. Simon Nadal from the Professor Shabaz Mohammed group, who also supplied HeLa cell nuclear extracts. Dr. Grace Roper from the Professor Akane Kawamura group supplied the KDM4A enzyme and assisted with Zn ejection assays.

The histone tail structure studies utilized the Boronoalanine installation chemistry developed by Dr. Tim Mollner, who assisted with preparation of Bal containing histones used to assemble nucleosomes, and the Aurora B kinase phosphorylation reactions. Charlie Buchanan of the Professor Andrew Baldwin group performed the NMR analysis on those Bal containing nucleosomes.

Dr. Sarah Faulkner provided histone H3-Cys36 mutants for the installation of di- and trimethyllysine and carried out nucleosome preparation and characterization with various members of the group of Professor Fumiko Esashi as well as Tom Watts. For histone acylation studies, Dr. Benjamin Bower provided Sirt2, and Dr. Patrick Isenegger assisted with synthesis of the acylated Lys chemical precursors.

Dr. Patrick Isenegger developed the Dha difluoroalkylation reaction used to install γCF_2-containing sidechains used for ^{19}F-NMR and was a partner throughout all the ^{19}F-NMR histone sidechain tracking studies.

Experimental Details

General Methods

Details for general experimental procedures, intact protein LC-MS, production (expression, purification, and Dha formation) of histone proteins are given in the **Experimental Details** section of **Chapter 1**.

Gel electrophoresis

Prior to gel loading, protein samples were first boiled in sample buffer (NuPAGE LDS Sample Buffer (4X), with added reducing agent, Thermo Fisher, 10 min at 95 °C). Crosslinking and modified protein characterization gels were ran on NuPAGE 12% Bis-Tris 1.0 or 1.5 mm gels with a variety of well numbers (Thermo Fisher, 200 V, 40 min, 4 °C). Nucleosome samples were run on Novex 6% TBE gels (ThermoFisher, 150 V, 90 min, 4 °C) pre-ran in 0.5X TBE running buffer (Thermo Fisher, 150 V, 90 min, 4 °C) and stained with SYBR Safe (Thermo Fisher, for DNA staining) and InstantBlue Coomassie Protein Stain (Abcam, for protein). Denatured nucleosome gels were run on Novex 16% Tris-Glycine gels with Tris Glycine Running Buffer (Thermo Fisher, 200 V, 2 h, 4 °C). Ladders used were the SeeBlue Plus2 Pre-Stained Protein Standard (Thermo Fisher, for proteins) and 100 bp DNA Ladder (New England Biolabs). Gels were ran in Novex MiniCell tanks, with a Bio-Rad PowerPac controller.

Western blotting

Western blotting was performed using an iBlot2 gel transfer device and iBlot2 transfer stacks, nitrocellulose, mini and regular size (Thermo Fisher). Membranes were blocked in TBS-T (Tris buffered saline + 0.1% Tween-20, 1 h, RT), incubated with primary antibodies (overnight, 4 °C), washed three times with TBS-T (5 min), incubated with a secondary antibody (1 h, 4 °C), and washed three more times with TBS-T then once with milliQ water (5 min) before imaging. Imaging was done one with a ChemiDoc XRS+ System (Bio-rad). Antibodies were used as per the manufacturer's recommendations: α-Histone H3 (96C10) Mouse mAb for histone detection, Mouse monoclonal α-polyHistidine-Alkaline Phosphatase, Clone HIS-1 (Sigma, A5588) for KDM4A detection (6His tag), Rabbit Anti-Mouse IgG (H+L) HRP conjugate (Promega, W4021) and Goat Anti-Mouse IgG H&L Alkaline Phosphatase (Abcam, ab97020) as secondary antibodies.

Zn Ejection Assay

The Zn ejection assay was performed using the N-(6-nethoxy-8-quinolyl)-p-toluenesulfonamide (TSQ) (Enzo) Zn(II) fluorophore as described with minor modifications[261,262]. In brief, assays were performed in 384 well black μCLEAR® non-binding plates (Grenier) using a reaction volume of 100 μL at 37 °C on a BMG CLARIOstar (360ex/490em). The plate was shaken (5 s, 700 rpm) before each reading, taken every 22 s for 270 cycles. Reactions consisted of 10 μM TSQ, 25 μM Ebselen or 20 μM histone eH3.1-Bhn9/-WT/4-bromobutylboronic acid, and those with enzyme contained 2 μM KDM4A all with 1.1 % (v/v) DMSO in HEPES buffer (50 mM, pH 7.5). Compounds and TSQ were added to the plate before initiating the assay with addition of KDM4A using the CLARIOstar injector. An internal calibration curve of $ZnCl_2$ (0-2 μM) in HEPES buffer (50 mM, pH 7.5) was included in each experiment to quantitate the concentrations of Zn(II) ejected. Data were normalized by subtracting a no enzyme control for that compound at each time point. Mean ± standard deviation (n=3 technical replicates) was plotted for each time point using GraphPad Prism 5.0, representative data from three biological replicates is shown. The rate of Zn(II) ejection was calculated from the slope of linear regression plotted over the linear region of Zn(II) ejection (from 946-3982 s) plotted in GraphPad Prism 5.0. Time-dependent release of Zn(II) from KDM4A was observed (9.270 ± 0.025 nM/min) when incubated with histone eH3.1-Bhn9 but not with eH3.1-WT or 4-bromobutylboronic acid. This contrasts with the rapid Zn(II) ejection rate with Ebselen, a Zn(II) chelating small molecule inhibitor of KDM4A activity[262], at > 1663 nM /min.

KDM4A-histone crosslinking LC-MS/MS

Crosslinked samples were searched with Peaks 8.5 to confirm the presence of both proteins and to confirm sample purity. Crosslinked samples were processed with the pLink 2.3.5 software package. Using the pConfig module, the amino acid ‘B’ of mass 205.01023 and composition H(12)C(7)N(1)O(1)Br(1) was defined. Note that the isotopic pattern doesn’t matter for crosslinking analysis as HBr gets eliminated. The linker 4BrBut was defined as linking the alpha amino acid B to the beta amino acids STYCHRKWDENQ with the linker composition H(-1)Br(-1). First search mass tolerance was set to 20 ppm and fragment mass tolerances at 20 ppm. A mass filter of 10 ppm was applied. E-values were computed and a global FDR of 1% set. Trypsin (or LysC for histone eH3.1-Bhn4) was selected as a protease with a maximum number of 3 missed cleavages. Carbamidomethylation (Cysteine), Oxidation (Methionine), Deamidation (Asparagine, Glutamine), Carbamylation (lysine, peptide N-term), Amidation (C-terminus) were set as variable modifications. RAW files were searched against a database containing the sequence of the modified histone eH3.1 construct and the expressed KDM4A. Resulting spectra were manually analyzed and validated using pLabel 2.3.5. As an empirical rule, an e-value higher than e-03 indicates a potential identification, higher than e-06 a reliable identification and higher than e-10 a very good identification.

Histone eH3.1-Bhn9 Nuclear Lysate Immunoprecipitation

Histone samples (20 µg or either Human Histone eH3.1-WT, eH3.1-Bhn9, or no Histone control) were immobilized on Anti-HA Magbeads (Pierce 88836, 50 µL/sample pre-equilibrated in buffer used for immobilization) via their HA epitope tag for 30 min at RT in HEPES buffer (50 mM, pH 7.5). The beads were then incubated with HeLa nuclear lysate (250 µL, 0.5 mg/mL, 4 hr, 37 °C, 600 rpm) to promote the crosslinking. The HeLa nuclear lysate was prepared as previously described[300]. After incubation, the beads were washed (5x with 500 µL HEPES buffer + 0.1% Tween20, 1x with milliQ water). The histones + interaction partners were eluted off the beads with Glycine (0.1 M, pH 2.0, 100 µL, 10 min, 37 °C) and quenched with Tris buffer (1 M, pH 8.5, 15 µL). The elution and quenching were repeated once more with the beads. To analyze the crosslinking and immunoprecipitation, controls of all histones and lysate, as well as samples from all conditions for incubation, last wash, and elution were checked via SDS-PAGE with either a Coomassie Blue staining or by western blot with an α-FLAG antibody (Histone eH3.1 samples are FLAG-HA epitope tagged) to detect higher MW bands corresponding to the mass of the Histone eH3.1 covalently crosslinking to an unknown interaction partner.

Widom 601 145 bp large-scale DNA preparation for nucleosomes

The plasmid containing the 8 × 145 bp Widom 601 sequence in a pUC vector was transformed into XL-10-Gold Ultracompetent cells (Agilent) following the manufacturer's instructions and spread on a Luria Bertani (LB) -agar plate containing ampicillin (100 µg/mL) and incubated overnight at 37 °C. The next day, colonies were transferred to four starter culture tubes containing Terrific Broth (Sigma Aldrich) and ampicillin and shaken overnight at 250 rpm, 37 °C. The day following that, 4 x 1 L flasks containing the same media were inoculated with the starter cultures and shaken for 24 h at 250 rpm, 37 °C. The cells were pelleted (5,000 g, 8 min, 4 °C), and the pellet resuspended in Alkaline Lysis Solution 1 (50 mM glucose, 10 mM EDTA, 25 mM Tris·HCl pH 8.0, 60 mL per liter of media producing the cell pellet). Twice the volume of Alkaline Lysis Solution II (200 mM NaOH, 1% v/v SDS) was added and the mixture shaken vigorously to remove any clumps, before incubation on ice for 20 min with occasional additional shaking. Alkaline Lysis Solution III (4 M NaOAc, 2 M AcOH, 210 mL per liter of media producing the cell pellet) was added and the solution inverted ten times before a further 20 min incubation on ice. The mixture was centrifuged (10,000 g, 30 min, 4 °C), and the supernatant filtered through Miracloth (Merck Millipore), followed by the addition of 0.52 volumes of isopropanol. After a 15 min incubation at room temperature, the precipitated DNA was pelleted by centrifugation (11,000 g, 30 min, 20 °C) and allowed to dry open to the air for 1 h. The pellet was fully dissolved in TE 10/50 (10 mM Tris, 50 mM EDTA pH 8.0, 40 mL) and split between two 50-mL falcon tubes. RNase A (NEB, 10 mg/mL, 120 µL) was added and the solutions incubated overnight at 37 °C. The next day, any precipitant was removed by centrifugation (10,000 g, 15 min, 20 °C) and the DNA in the

supernatant was precipitated by adding 1/5 the original volume of 4 M NaCl and 2/5 the original volume of 40% PEG6000. After mixing at 37 °C for 5 min and incubating on ice for 30 min, the DNA was pelleted by centrifugation (3000 g, 20 min, 4 °C). The pellet was dissolved in TE 10/0.1 (10 mM TRIS, 0.1 mM EDTA pH 8.0, 15 mL) by rocking overnight at 37 °C. To digest the 145 bp inserts out of the plasmid backbone, the following reagents were added: Tris·HCl (to 6 mM from a 1 M stock, pH 8.0), $MgCl_2$ (6 mM from a 4 M stock), NaCl (150 mM from a 4 M stock) and DTT (1 mM from a 1 M stock). Then, the restriction enzyme EcoRV-HF (10,000 units, NEB) was added and the solution incubated at 37 °C for 20 h. The digest was checked for completion at this stage by analyzing an aliquot via agarose gel electrophoresis. Following digestion, 0.192 volumes of NaCl (4 M) and 0.346 volumes of PEG6000 were added, and the solution incubated on ice for 1 h to precipitate the larger plasmid backbone. The precipitant was pelleted by centrifugation (27,000 g, 20 min, 4 °C) and the 145 bp insert containing supernatant was poured into ice cold ethanol (125 mL) to induce precipitation with a 20 min incubation on ice. The supernatant/ethanol mixture was divided among four 50 mL centrifuge tubes and centrifuged (27,000 g, 20 min, 4 °C), and the pellet air dried for 20 min and dissolved in TE 10/0.1 (5 mL). An aliquot of the 145 bp insert containing pellet was analyzed at this stage by agarose gel electrophoresis to validate efficient separation from the larger vector backbone DNA. Next, 2/5 volumes of phenol and 1/5 volumes of chloroform were added to the dissolved insert solution and vortexed to mix. The suspension was then centrifuged (5,000 g, 10 min, RT) to separate phases, and the upper aqueous later extracted with a pipette and mixed 1:1 v/v with chloroform. The mixture was vortexed and centrifuged as before and the aqueous layer subjected to one more round of chloroform (1:1 v/v)

purification. The final aqueous layer was extracted by pipette, and a 1/10 volume of NaOAc (3 M, pH 5.2) and 3 volumes of ethanol were added. The solution was vortexed and allowed to precipitate at −20 °C for several days. The precipitant was pelleted by centrifugation (10,000 g, 30 min, 4 °C), air dried for 30 min and resuspended in water (1 mL, 4.5 mg of 145 bp insert final yield).

Bal containing histone octamer formation

Lyophilized histones were dissolved in Unfolding Buffer (6 M GdnHCl, 5 mM dithiothreitol (DTT), 20 mM Tris-HCl pH 7.5) in equimolar ratios (20 µM, 72 nmol each) and incubated at room temperature for 1 h. The solution was transferred to a dialysis cassette (ThermoFisher, Slide-A-Lyzer Dialysis Cassette, MWCO 7 kDa) and dialyzed three times against Refolding Buffer (2 M NaCl, 1 mM EDTA, 1 mM DTT, 10 mM Tris-HCl pH 7.5) for 2 h each time at 4 °C, with the last dialysis taking place overnight. The resulting solution was concentrated (VivaSpin 6, MWCO 5 kDa) to approximately 1 mL and purified via size exclusion chromatography (Cytiva, HiLoad 16/600 Superdex 200 pg), eluting with Refolding Buffer. Fractions containing the octamer and absent of other residual histone species (tetramer, dimer and monomers) were revealed by SDS–PAGE analysis. The purified octamer was concentrated to at least 2 mg/mL (measured by A_{280} spectrophotometry with the combined histone molecular weight and extinction coefficients) and stored at 4 °C until further use.

BaI containing nucleosome assembly

Histone octamers (1.5 mg, 2 mg/mL in Refolding Buffer) were mixed with an equimolar amount of 145 bp DNA (2 mg/mL in 2 M KCl) and dialyzed (4 °C, 2 h per dialysis step) against Reconstitution Buffer (1 mM EDTA, 1 mM DTT, 10 mM Tris-HCl pH 7.5) containing 2, 0.85, 0.65 and last 0.25 mM KCl. The resulting solution was then centrifuged (5 min, 10,000 rpm, 4 °C) to remove any precipitant, if any, and nucleosome aliquots (10 µL) were heat-shifted (1 h at 4, 37 or 55 °C) to ensure a homogenously populated nucleosome with respect to the position of the DNA on the histone octamer. The incubated aliquots were then analyzed by Native–PAGE to reveal the presence of both DNA (Invitrogen SYBR Safe Staining) and octamer (Coomassie Blue staining) in the nucleosome band. The heat-shifted samples revealed that every reconstituted nucleosome required no further heat shifting beyond 4 °C, as evident by the lack of several close nucleosome bands resolving to one band at higher temperature incubations. Coomassie Blue staining revealed that most protein was contained in the nucleosome band, while SYBR Safe staining revealed that in every case, there was some residual free 145 bp DNA remaining in the sample. Any lost octamer was presumed to have crashed out during dialysis, evident by the presence of some precipitant in most cases post-dialysis.

Bal nucleosome NMR sample preparation

Nucleosomes were prepared for NMR by dialyzing three times (2 h each, last one overnight, 4 °C) into NMR buffer (1 mM EDTA, 10 mM NaPi pH 6.5), the last dialysis was performed with a slightly more concentrated buffer stock that, when 5% v/v D_2O was added, gave the desired NMR buffer concentration. After dialysis, the nucleosomes were concentrated (VivaSpin 500, MWCO 5 kDa) to 190 µL, and then 10 µL of D_2O was added. The samples were transferred to 3 mm Bruker Match NMR tubes. After NMR measurement, some nucleosomes were further modified via phosphorylation with Aurora B kinase. These nucleosomes were reevaluated by Native–PAGE (without further heat shifting) to confirm integrity and dialyzed again as described above into the NMR buffer for further NMR measurement.

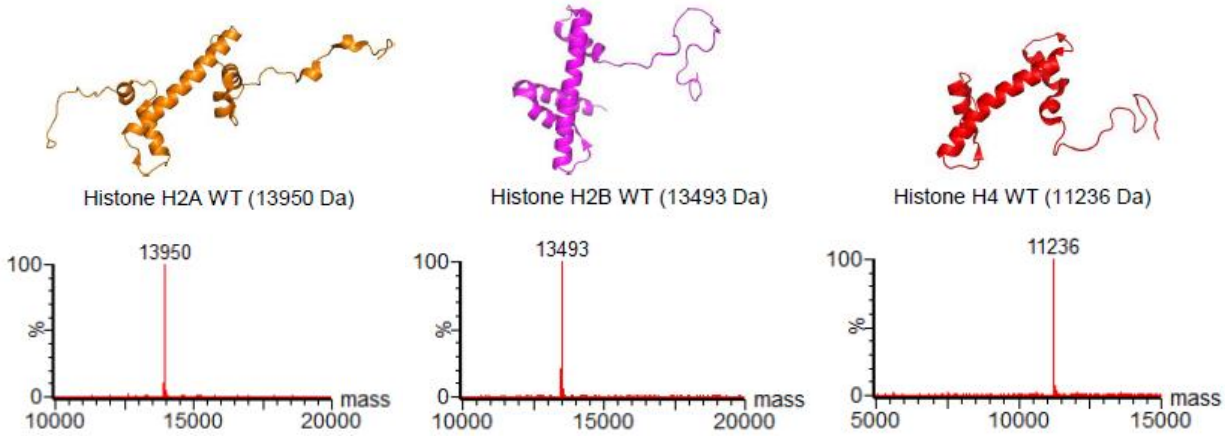
Phosphorylation of histones and nucleosomes with Aurora B Kinase

For free histone phosphorylation, a solution of Histone H3-Ser10 (either native or from oxidized Histone H3-Bal10) was prepared in Kinase Buffer (10 mM Tris, 30 mM NaCl, 1 mM DTT, 3 mM $MgCl_2$, 2 mM ATP, 10 mM β-glycerophosphate, pH 7.5). To 500 µL of this solution in a microcentrifuge tube were added 0.5 µg of Aurora B kinase (Abcam, ab51435), and the mixture was vortexed. The tube was incubated at 30 °C, 400 rpm for 2 days before the samples were analyzed by SDS–PAGE and western blot.

For nucleosome phosphorylation, after NMR measurement, samples were retrieved from the NMR tubes and diluted with 1,800 µL of Kinase Buffer. In the case of Bal-containing nucleosomes, dilution was conducted in a nitrogen filled glove box with kinase buffer that was degassed overnight (4 °C). This precaution was taken to prevent oxidation of Bal to Ser, which was observed on Histone H3-Bal9 on prolonged exposure to kinase buffer at elevated temperature (30 °C). Recombinant human Aurora B kinase (3 µg, Abcam, ab51435) was added and the mixture was shaken (400 rpm) at 30 °C for 48 h, after which the samples were again prepared for NMR measurement.

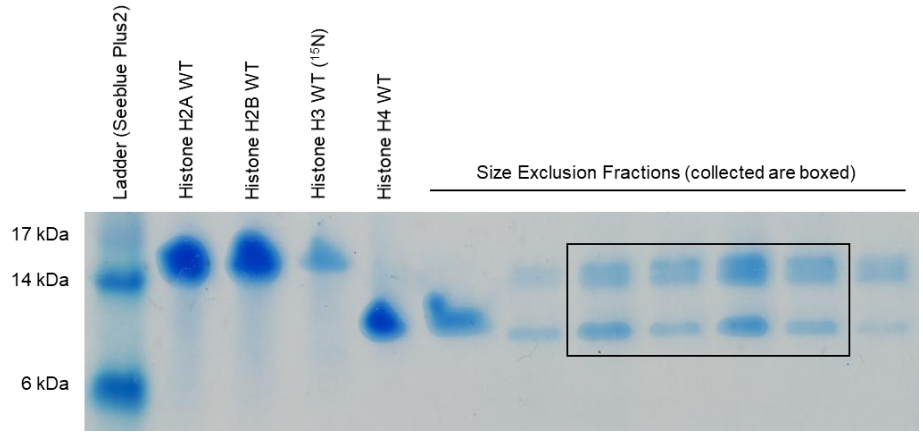
Histone production for Bal NMR studies

WT histones for histone octamer reconstitution and nucleosome assembly were prepared as previously described[242,300]. Production of Bal containing histones was carried out by Dr. Tim Mollner[221]. Intact protein LC-MS deconvoluted spectra are provided below for WT histone H2A, H2B, and H4.

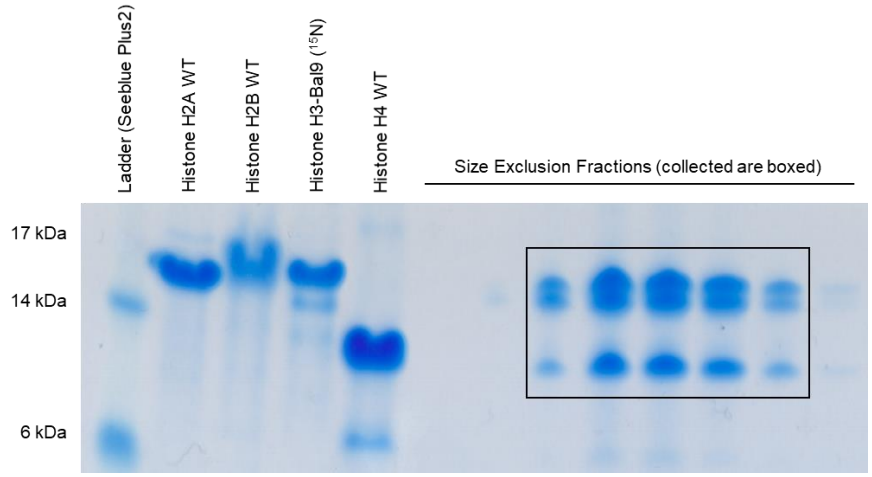


Histone octamer gel characterization for Bal NMR studies

Octamer containing ^{15}N Labelled Histone H3-WT

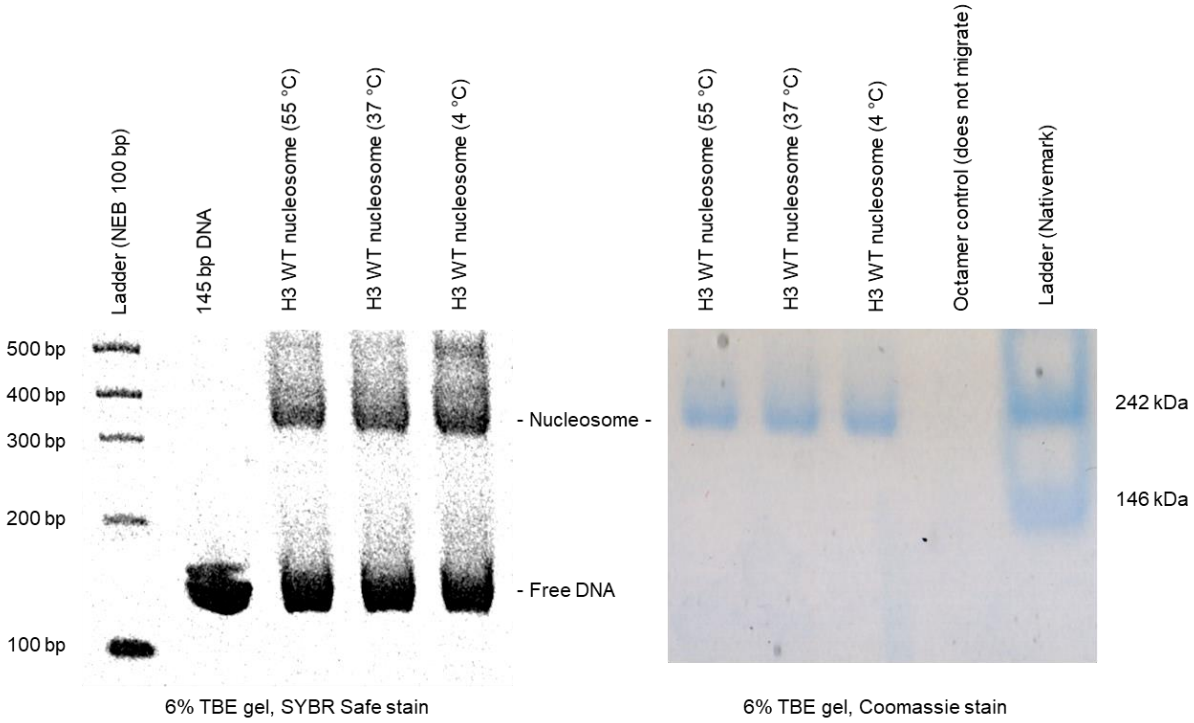


Octamer containing ^{15}N Labelled Histone H3-Bal9

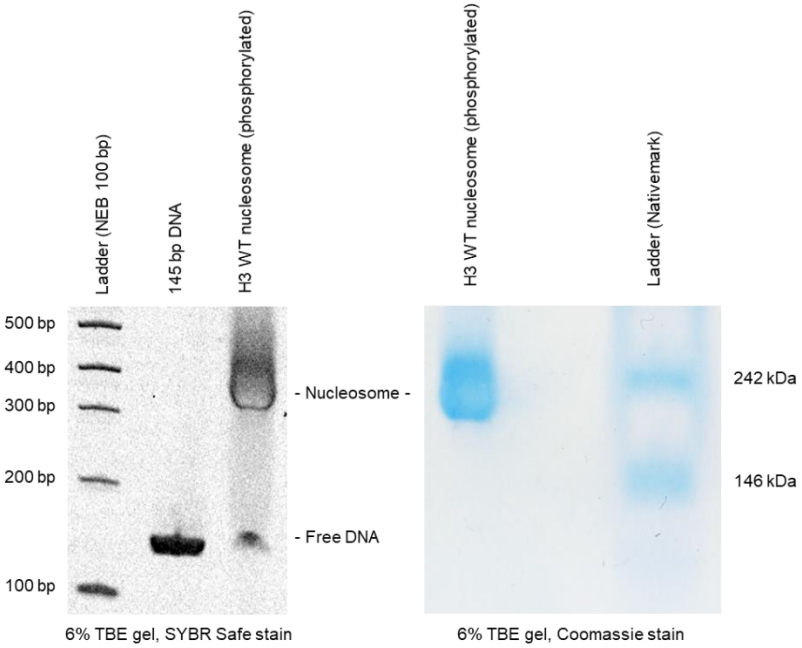


Nucleosome gel characterization for Bal NMR studies

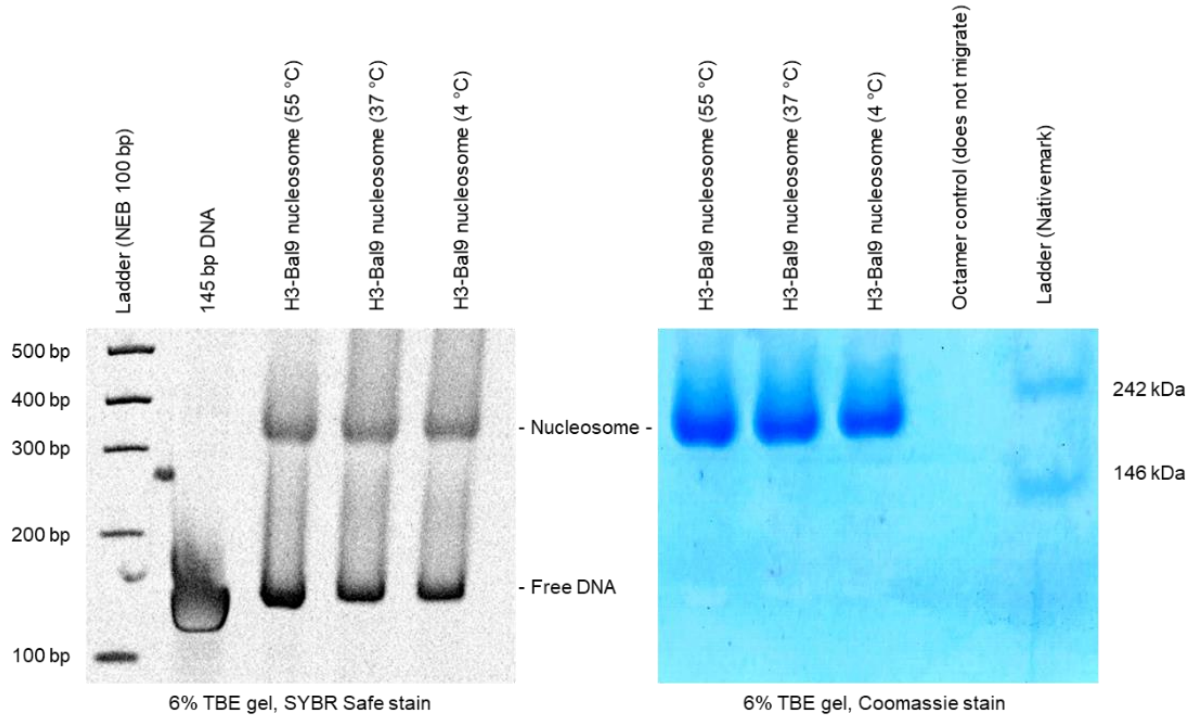
Nucleosome containing ^{15}N Labelled Histone H3-WT



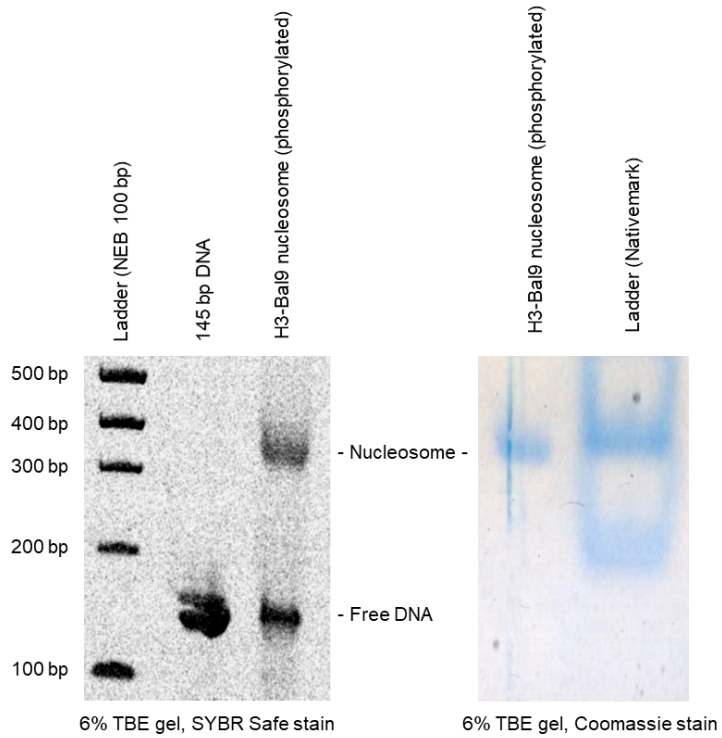
Nucleosome containing ^{15}N Labelled Histone H3-WT after phosphoryation



Nucleosome containing ^{15}N Labelled Histone H3-Bal9



Nucleosome containing ^{15}N Labelled Histone H3-Bal9 after phosphorylation



Histone octamer formation for ^{19}F-NMR measurement

The expression and purification of core histones and histone tetramer and octamer formation was carried out essentially as previously described[242,300]. In brief, after ^{19}F-NMR measurement of the unfolded histone H3-γCF_2-hAla9 (NH_4OAc 500 mM, pH 6, GdnHCl 3 M, prepared in D_2O), 2 mg of the histone was buffer exchanged into 1 mL of Unfolding Buffer (7 M GdnHCl, 10 mM Tris, 1 mM EDTA, 10 mM DTT, 1 mM Benzamidine, pH 7.5) then buffer exchanged into Tris Buffer (150 mM NaCl, 10 mM Tris, 1 mM EDTA, 2 mM βME, pH 7.5) with a PD10 G25 Minitrap. An equivalent volume of Deuterated Tris Buffer (as above but made with 100% D_2O) to make a final buffer with 50% D_2O, which was concentrated to a volume of 0.75 mL (Vivaspin 6, 5 kDa MWCO). The mixture was centrifuged (15000 rpm, 10 min, 4 °C) to pellet any precipitant, and an internal standard of trifluoroethanol (1 eq) was added. The concentration was measured (Nanodrop, 2.0 mg/mL), folding checked via Circular Dichroism (CD), and filtered into an NMR tube. For the Histone H3-DfeGly9-H4 tetramer reconstitution, the modified histone H3 and histone H4-WT (1:1 molar ratio, 2.5 mg of histone H3-γCF_2-hAla9) were mixed in Unfolding Buffer (6 mL), incubated for 30 min at RT, then dialyzed into Refolding Buffer (3x into 1 L, 2 h each with one overnight). The resultant solution was centrifuged (15000 rpm, 10 min, 4 °C) to pellet any precipitant, concentration was measured (0.5 mg/mL, 1 mL), and purified via Size Exclusion (Superdex S75, 16/60 pre-equilibrated in Refolding Buffer). Tetramer containing fractions (visualized via SDS-PAGE analysis) were combined, concentrated, and resuspended into 50% Deuterated Refolding Buffer (made with 1:1 H_2O/D_2O). Trifluoroethanol (0.1 µL in 1 mL) was added as an internal NMR standard. The final concentration was measured (Nanodrop, 2.5 mg/mL), and folding checked via Circular

Dichroism (CD) before filtering into an NMR tube. For the histone H3-γCF_2-hAla9-H4-H2A-H2B octamer reconstitution, all histones were dissolved in Unfolding Buffer (1:1:1.1:1.1 molar ratio, 25 nmol of modified histone H3) and incubated for 30 min at RT before dialyzing into Refolding Buffer (3x into 1 L, 2 h each with one overnight). The resultant solution was centrifuged (15000 rpm, 10 min, 4 °C) to pellet any precipitant before purifying via Size Exclusion as above. Fractions containing histone H3-γCF_2-hAla9-H4-H2A-H2B octamer were collected, concentration measured (Nanodrop, 0.8 mg total), an internal trifluoroethanol standard was added (0.1 µL) and the NMR sample was prepared and measured as above. After the NMR, the octamer was analyzed by SDS-PAGE and CD to check proper folding.

Sirt2 deacylation of modified histones

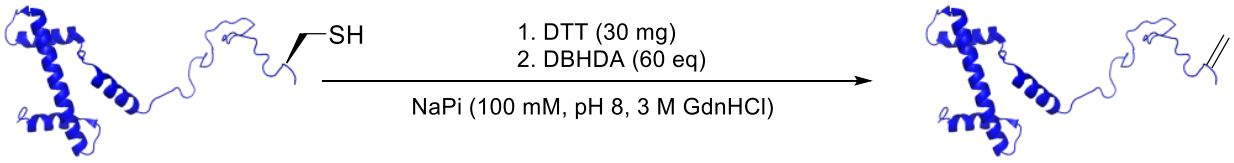
The reaction products of histone eH3.1-Ac/BzLys18 installation were dialyzed thrice against HEPES buffer (20 mM, pH 7.4), twice for 2 h, once overnight at 4 °C. The solutions were pre-warmed to 37 °C, and Sirt2 (0.5 µg from a stock in the same buffer) and NAD+ (150 µM final concentration from a 10X stock in buffer) were added to solutions containing the modified histones (20 µM histone, 50 µL final reaction volume). The reactions were shaken at 37 °C, 600 rpm, with aliquots of the crude reaction mixture taken out at 3 time points (5 min, 30 min, 12 h), diluted (1:50 in H_2O + 1% v/v formic acid) and immediately analyzed via LC-MS intact protein LC-MS. Both reactions stopped at ~50% deacylation, which we take as evidence of the 50:50 D:L diastereomer ratio after the installation of the modifications on Dha, as the D diastereomer is not recognized by Sirt2 as a substrate. For the Sirt2 deacetlyation of histone eH3.1-γCF_2AcLys18, the reaction was carried out essentially the same as above but at a much larger scale (histone 625 µg, Sirt2 100 µg, NAD+ 150 µM, 1 mL, HEPES 50 mM, pH 7.5, 30 °C, 1 h) to produce enough material for subsequent protein for ^{19}F-NMR.

General procedure for large-scale ^{19}F-NMR measurement of γCF_2- labelled histones

A glass vial (5 mL) was charged with $FeSO_4 \cdot 7H_2O$ (100 eq) and transferred into a glovebox. Then, an aliquot of Dha modified histone (1-5 mg, 0.5-1 mL, typical protein concentration of 1-5 mg/mL), difluoroalkylpyridylsulfone reagent (5 eq in 1 M DMSO stock) and $Ru(bpy)_3Cl_2$ (1-5 eq in 10 µL water) were added to the glass vial. Afterwards, the vial was sealed with a plastic cap, transferred out of the glovebox and irradiated with blue LED light (50W) for 15 minutes. The crude reaction mixture was treated with either βME (10 µL), vortexed for 30 s and purified by PD MiniTrap G-25 column (GE Healthcare) (equilibrated with buffer in D_2O following gravity protocol) by adding 500 µL of the reaction product to the column (multiple columns were sometimes necessary), adding 200 µL of NMR buffer and discarding the eluant, then adding 500 µL of NMR buffer and collecting the eluant, yielding the pure fluorinated histone. After concentration of the protein sample to a volume of 0.5 mL using a vivaspin column (mwco = 5000 Da) the sample is ready for recording a ^{19}F-NMR spectra.

Protein modification details

Formation of histone eH3.1-Dha4

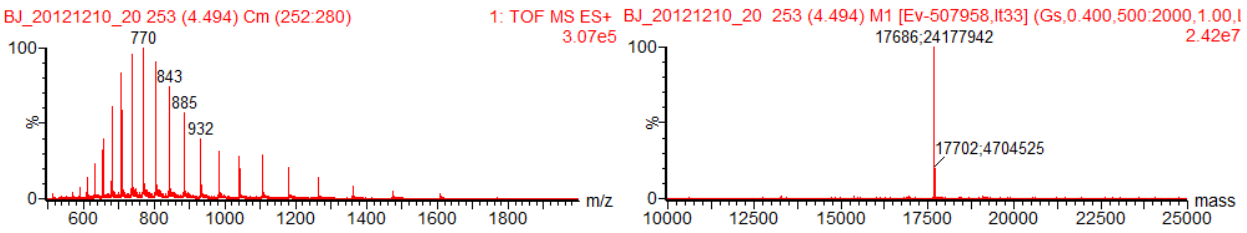


Human histone eH3.1-Cys4 Human histone eH3.1-Dha4

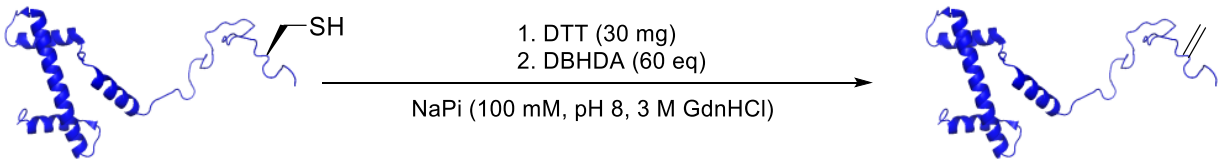
Lyophilized *Homo sapiens* histone eH3.1-Cys4 (10 mg) was added to denaturing phosphate buffer (100 mM NaPi, pH 8, 3 M GdnHCl, 500 µL) and mixed until fully dissolved. DTT was added (30 mg) and shaken for 30 min at RT, 500 rpm, to reduce disulfide bonds, before being removed via desalting into 1 mL of the same buffer (PD Minitrap G25, GE Healthcare). The resulting protein concentration was determined (Nanodrop) and immediately followed by the addition of DBHDA (60 eq from a freshly prepared 0.5 M DMSO stock) and subsequent shaking (500 rpm) at 25 °C for 45 min, then 37 °C for 2 h. The protein was desalted as before to remove the excess DBHDA and exchange into the desired buffer. Protein yield and concentration was determined by Nanodrop and conversion was determined by LC-MS analysis.

Calculated mass: 17686 Da

Observed mass: 17686 Da



Formation of histone eH3.1-Dha9

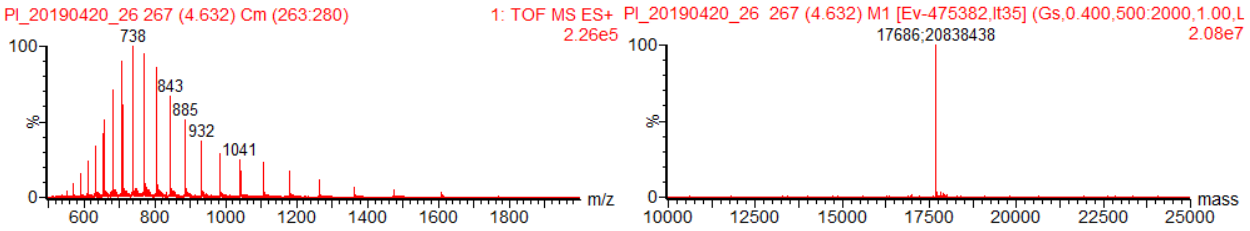


Human histone eH3.1-Cys9 Human histone eH3.1-Dha9

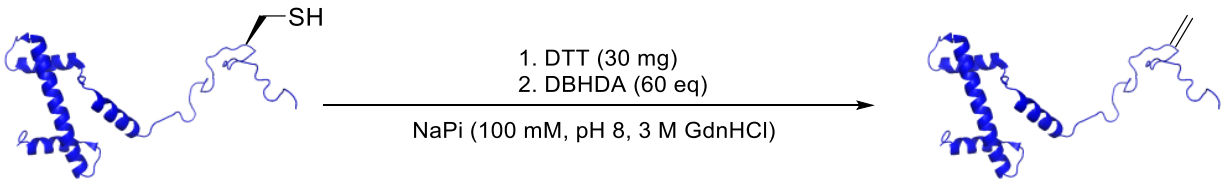
Lyophilized *Homo sapiens* histone eH3.1-Cys9 (10 mg) was added to denaturing phosphate buffer (100 mM NaPi, pH 8, 3 M GdnHCl, 500 µL) and mixed until fully dissolved. DTT was added (30 mg) and shaken for 30 min at RT, 500 rpm, to reduce disulfide bonds, before being removed via desalting into 1 mL of the same buffer (PD Minitrap G25, GE Healthcare). The resulting protein concentration was determined (Nanodrop) and immediately followed by the addition of DBHDA (60 eq from a freshly prepared 0.5 M DMSO stock) and subsequent shaking (500 rpm) at 25 °C for 45 min, then 37 °C for 2 h. The protein was desalted as before to remove the excess DBHDA and exchange into the desired buffer. Protein yield and concentration was determined by Nanodrop and conversion was determined by LC-MS analysis.

Calculated mass: 17686 Da

Observed mass: 17686 Da



Formation of histone eH3.1-Dha18

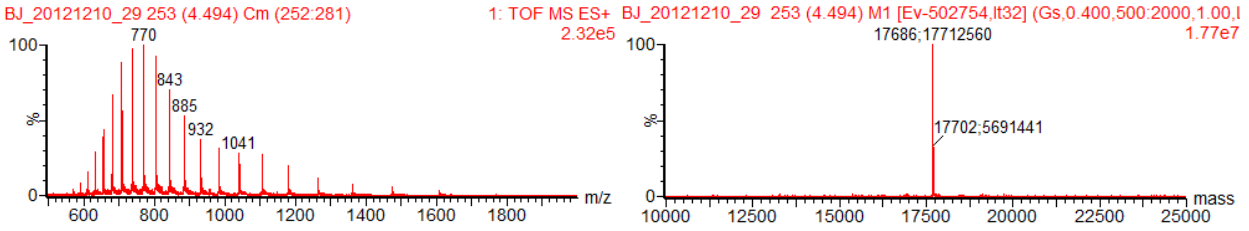


Human histone eH3.1-Cys18 Human histone eH3.1-Dha18

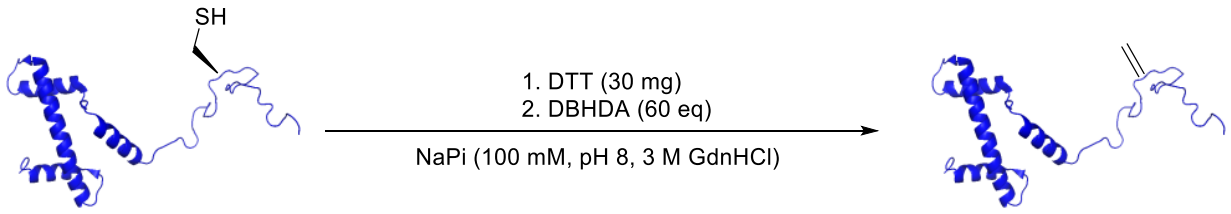
Lyophilized *Homo sapiens* histone eH3.1-Cys18 (10 mg) was added to denaturing phosphate buffer (100 mM NaPi, pH 8, 3 M GdnHCl, 500 µL) and mixed until fully dissolved. DTT was added (30 mg) and shaken for 30 min at RT, 500 rpm, to reduce disulfide bonds, before being removed via desalting into 1 mL of the same buffer (PD Minitrap G25, GE Healthcare). The resulting protein concentration was determined (Nanodrop) and immediately followed by the addition of DBHDA (60 eq from a freshly prepared 0.5 M DMSO stock) and subsequent shaking (500 rpm) at 25 °C for 45 min, then 37 °C for 2 h. The protein was desalted as before to remove the excess DBHDA and exchange into the desired buffer. Protein yield and concentration was determined by Nanodrop and conversion was determined by LC-MS analysis.

Calculated mass: 17686 Da

Observed mass: 17686 Da



Formation of histone eH3.1-Dha27

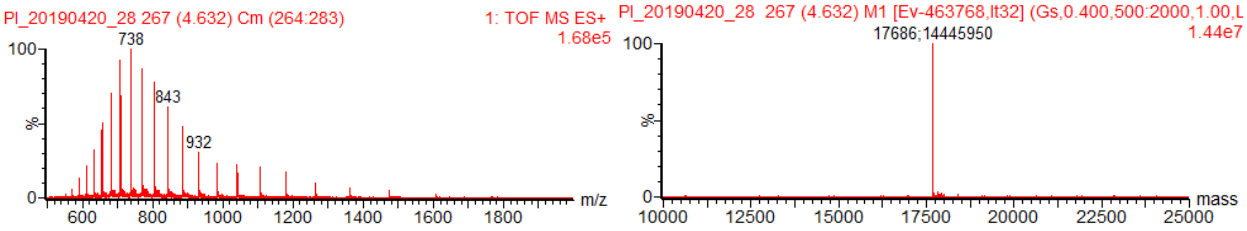


Human histone eH3.1-Cys27 Human histone eH3.1-Dha27

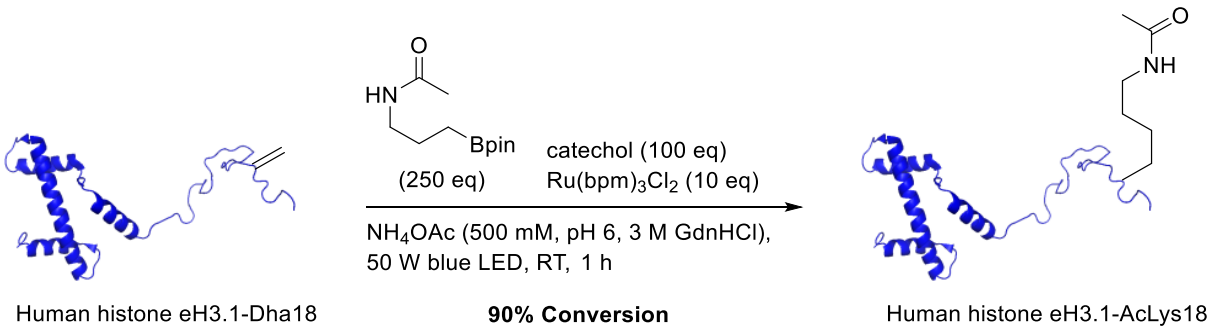
Lyophilized *Homo sapiens* histone eH3.1-Cys27 (10 mg) was added to denaturing phosphate buffer (100 mM NaPi, pH 8, 3 M GdnHCl, 500 µL) and mixed until fully dissolved. DTT was added (30 mg) and shaken for 30 min at RT, 500 rpm, to reduce disulfide bonds, before being removed via desalting into 1 mL of the same buffer (PD Minitrap G25, GE Healthcare). The resulting protein concentration was determined (Nanodrop) and immediately followed by the addition of DBHDA (60 eq from a freshly prepared 0.5 M DMSO stock) and subsequent shaking (500 rpm) at 25 °C for 45 min, then 37 °C for 2 h. The protein was desalted as before to remove the excess DBHDA and exchange into the desired buffer. Protein yield and concentration was determined by Nanodrop and conversion was determined by LC-MS analysis.

Calculated mass: 17686 Da

Observed mass: 17686 Da



Installation of AcLys on histone eH3.1-Dha18

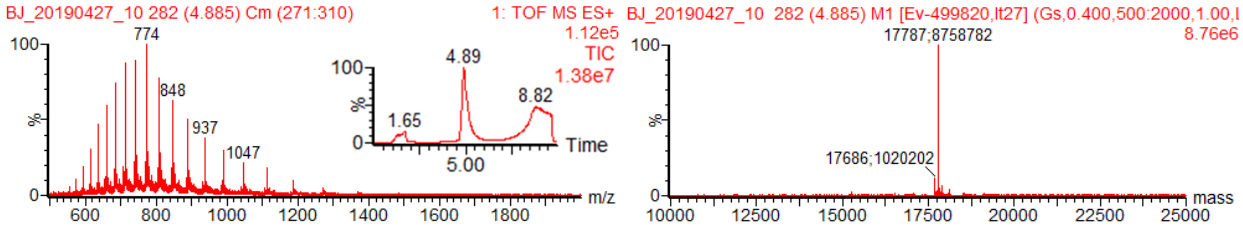


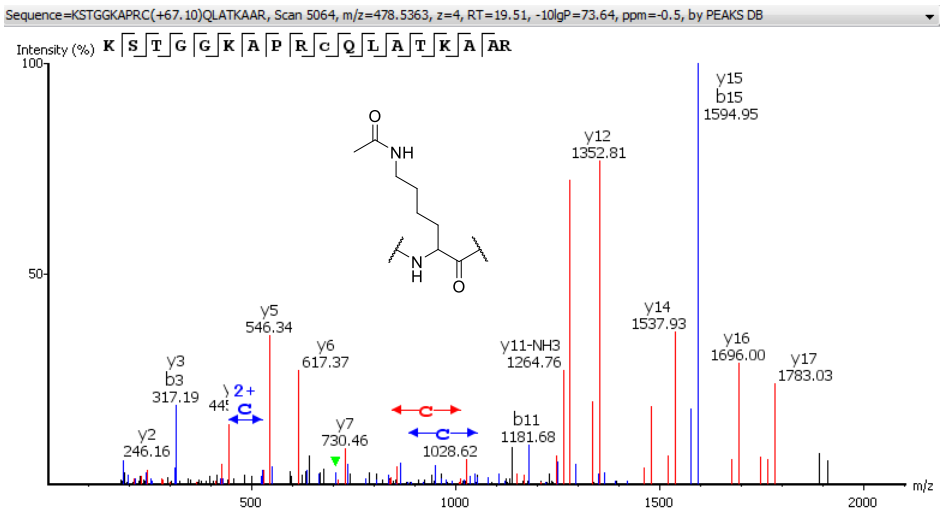
Human histone eH3.1-Dha18 **90% Conversion** Human histone eH3.1-AcLys18

In the glovebox, a glass HPLC vial was charged with NH_4OAc buffer (500 mM, pH 6, 3 M GdnHCl, 90 µL) containing Human histone eH3.1-Dha18 (100 µg, final concentration of 1 mg/mL, 56 µM). After the sequential addition of $Ru(bpm)_3Cl_2$ (1 µL of a 56 mM stock prepared fresh in water, 10 eq), catechol (1 µL of a 560 mM stock prepared fresh in water, 100 eq) and N-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)propyl)acetamide (5 µL of a 280 mM stock prepared fresh in buffer, 250 eq), the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 1 h. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17787 Da. Observed mass: 17787 Da

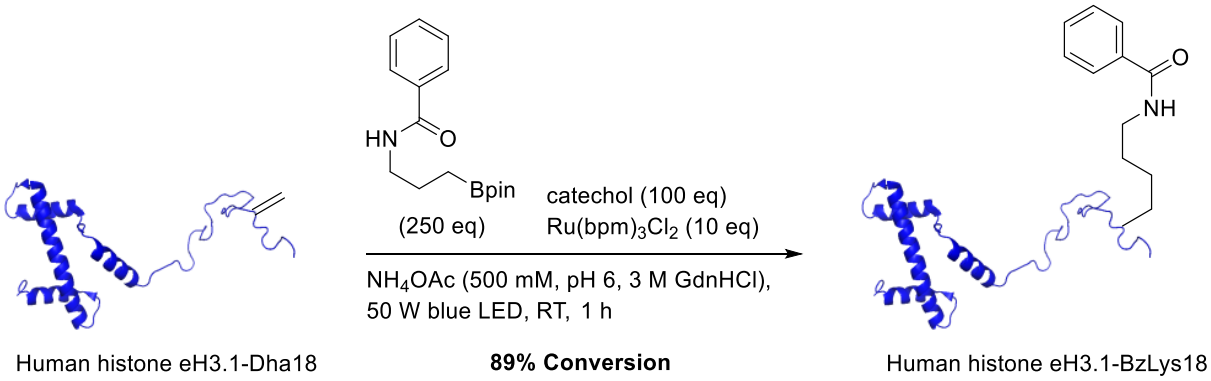
Calculated Dha mass: 17686 Da. Observed mass: 17686 Da





LC-MS/MS of histone eH3.1-AcLys18: ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

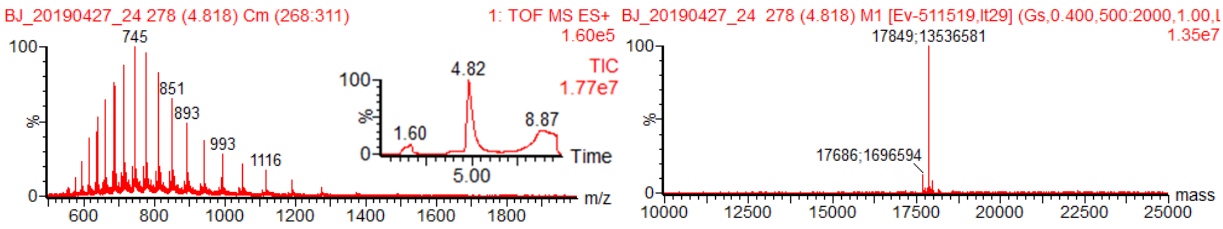
Installation of BzLys on histone eH3.1-Dha18

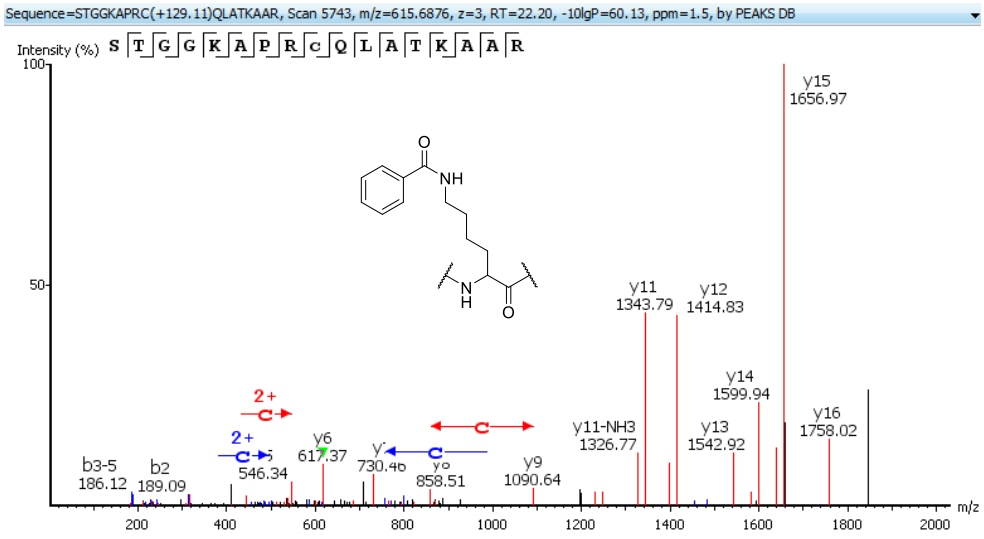


In the glovebox, a glass HPLC vial was charged with NH_4OAc buffer (500 mM, pH 6, 3 M GdnHCl, 90 µL) containing Human histone eH3.1-Dha18 (100 µg, final concentration of 1 mg/mL, 56 µM). After the sequential addition of $Ru(bpm)_3Cl_2$ (1 µL of a 56 mM stock prepared fresh in water, 10 eq), catechol (1 µL of a 560 mM stock prepared fresh in water, 100 eq) and N-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)propyl)benzamide (5 µL of a 280 mM stock prepared fresh in acetonitrile, 250 eq) the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 1 h. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17849 Da. Observed mass: 17849 Da

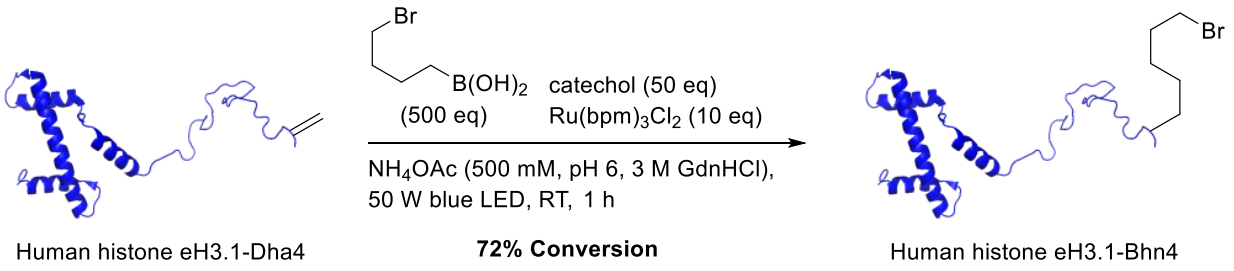
Calculated Dha mass: 17686 Da. Observed mass: 17686 Da





LC-MS/MS of histone eH3.1-BzLys18: ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

Installation of Bhn on histone eH3.1-Dha4

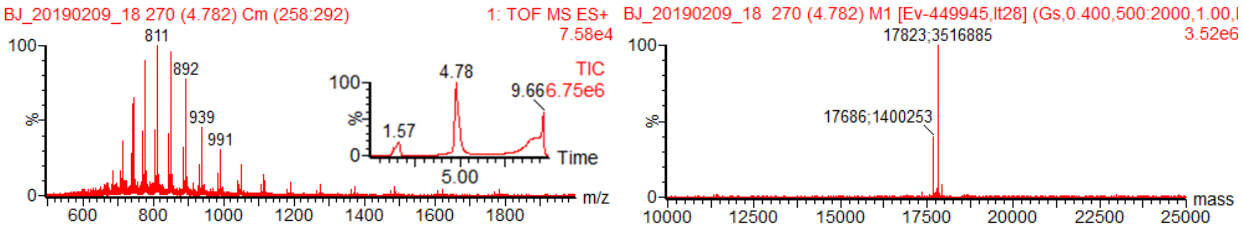


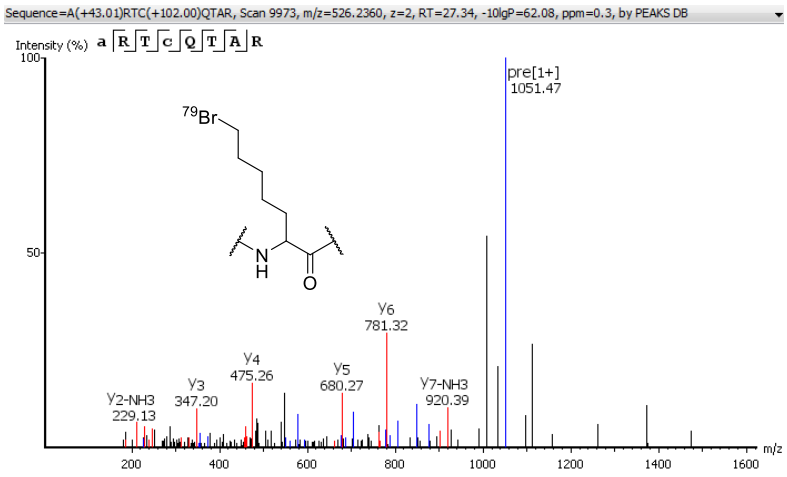
Human histone eH3.1-Dha4 **72% Conversion** Human histone eH3.1-Bhn4

In the glovebox, a glass HPLC vial was charged with NH_4OAc buffer (500 mM, pH 6, 3 M GdnHCl, 95 µL) containing Human histone eH3.1-Dha4 (100 µg, final concentration of 1 mg/mL, 56 µM). After the sequential addition of $Ru(bpm)_3Cl_2$ (1 µL of a 56 mM stock prepared fresh in water, 10 eq), catechol (0.5 µL of a 560 mM stock prepared fresh in water, 50 eq) and 4-bromobutylboronic acid (5 µL of a 560 mM stock prepared fresh in buffer, 500 eq), the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 1 h. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17822 Da. Observed mass: 17823 Da

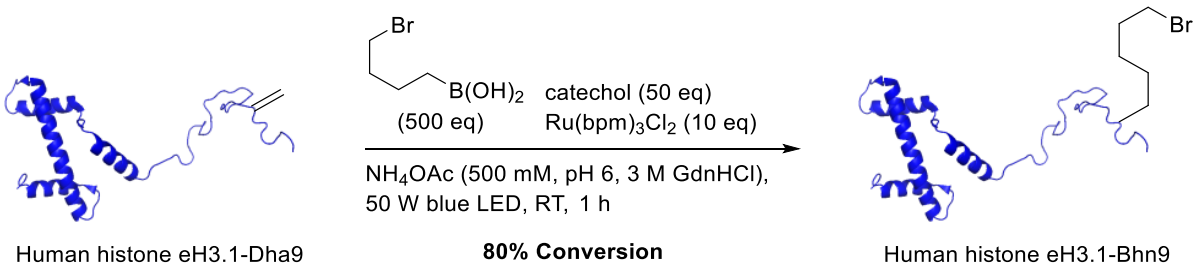
Calculated Dha mass: 17686 Da. Observed mass: 17686 Da





LC-MS/MS for histone eH3.1-Bhn4: ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

Installation of Bhn on histone eH3.1-Dha9

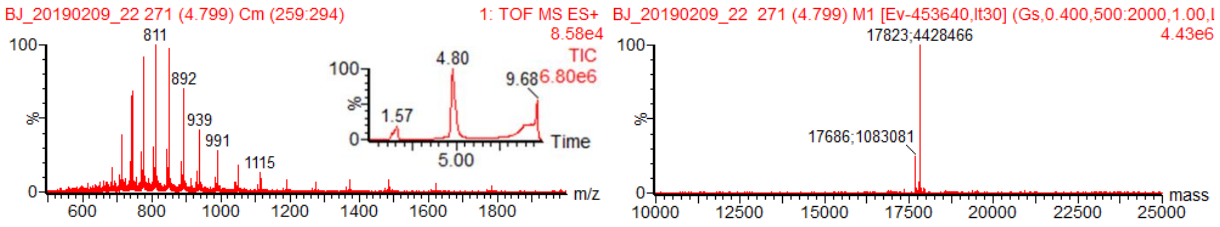


Human histone eH3.1-Dha9 **80% Conversion** Human histone eH3.1-Bhn9

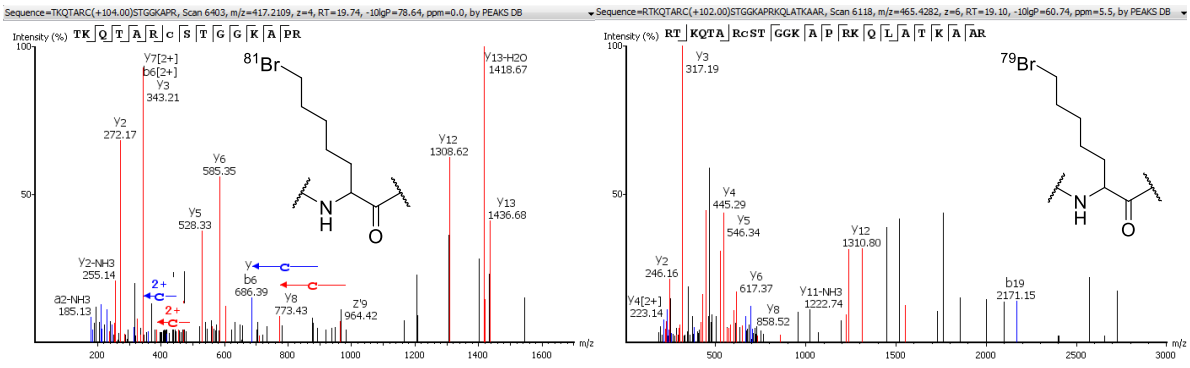
In the glovebox, a glass HPLC vial was charged with NH_4OAc buffer (500 mM, pH 6, 3 M GdnHCl, 95 µL) containing Human histone eH3.1-Dha9 (100 µg, final concentration of 1 mg/mL, 56 µM). After the sequential addition of $Ru(bpm)_3Cl_2$ (1 µL of a 56 mM stock prepared fresh in water, 10 eq), catechol (0.5 µL of a 560 mM stock prepared fresh in water, 50 eq) and 4-bromobutylboronic acid (5 µL of a 560 mM stock prepared fresh in buffer, 500 eq), the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 1 h. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17822 Da. Observed mass: 17823 Da

Calculated Dha mass: 17686 Da. Observed mass: 17686 Da

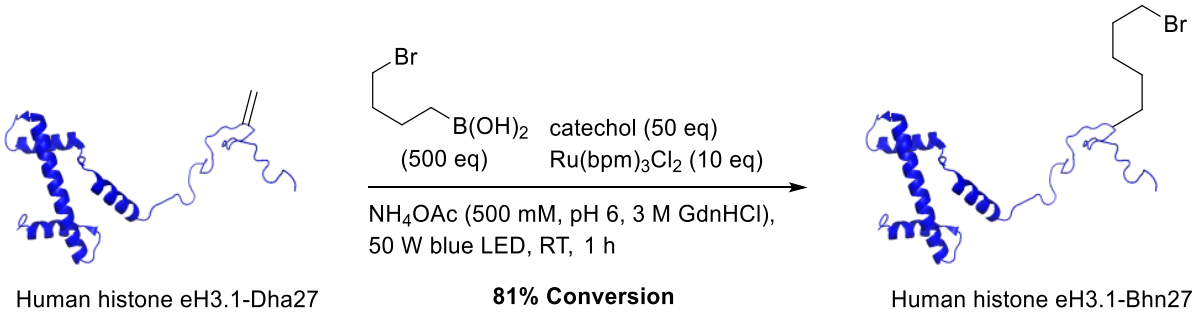


^{81}Br ^{79}Br



LC-MS/MS for histone eH3.1-Bhn9: ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

Installation of Bhn on histone eH3.1-Dha27

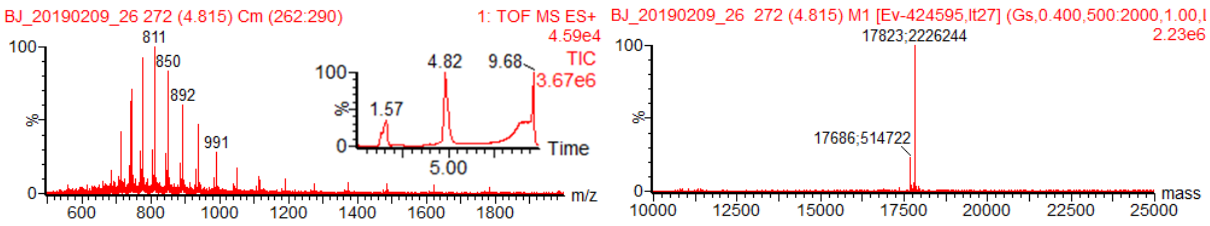


Human histone eH3.1-Dha27 **81% Conversion** Human histone eH3.1-Bhn27

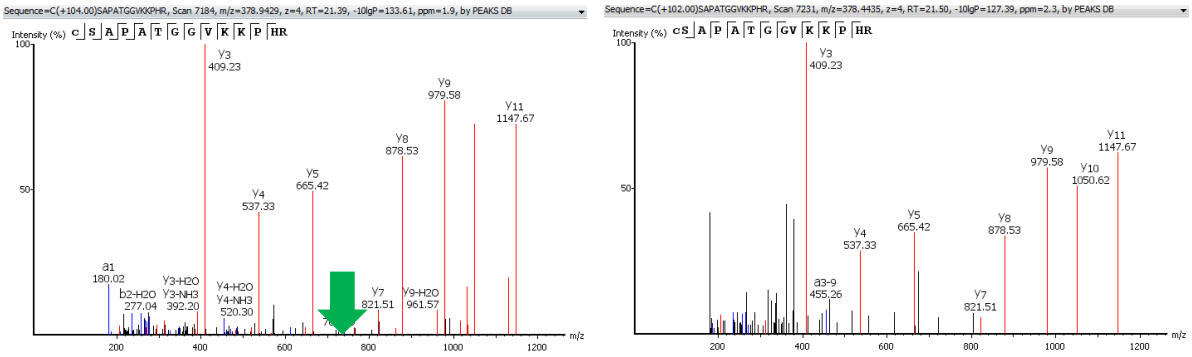
In the glovebox, a glass HPLC vial was charged with NH_4OAc buffer (500 mM, pH 6, 3 M GdnHCl, 95 µL) containing Human histone eH3.1-Dha27 (100 µg, final concentration of 1 mg/mL, 56 µM). After the sequential addition of $Ru(bpm)_3Cl_2$ (1 µL of a 56 mM stock prepared fresh in water, 10 eq), catechol (0.5 µL of a 560 mM stock prepared fresh in water, 50 eq) and 4-bromobutylboronic acid (5 µL of a 560 mM stock prepared fresh in buffer, 500 eq), the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 1 h. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17822 Da. Observed mass: 17823 Da

Calculated Dha mass: 17686 Da. Observed mass: 17686 Da

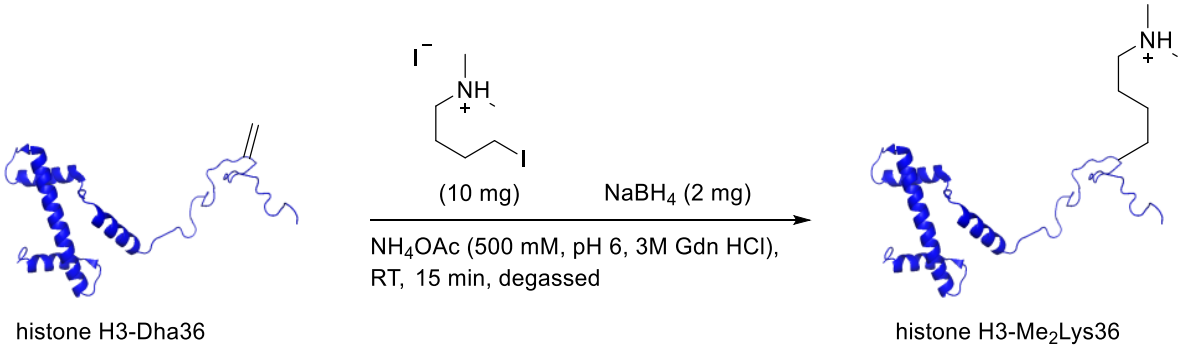


^{81}Br ^{79}Br



LC-MS/MS of histone eH3.1-Bhn27: ArgC digest leads to 100% sequence coverage (1% FDR). Q Exactive (HCD).

Installation of Me_2Lys on histone H3-Dha36

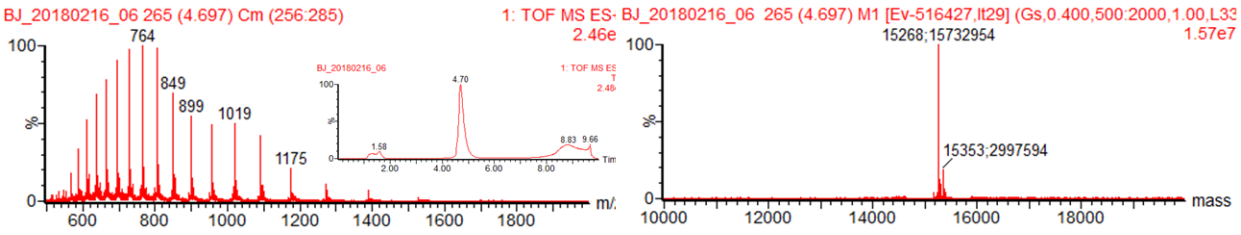


histone H3-Dha36 histone H3-Me_2Lys36

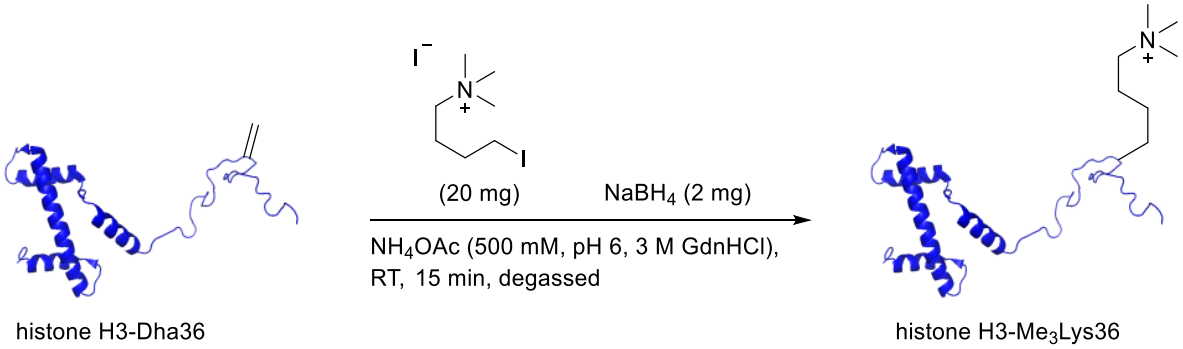
In the glovebox, an Eppendorf tube was charged with NH_4OAc buffer (500 mM, pH 6, 3 M Gdn HCl, 100 µL) containing histone H3-Dha36 (100 µg, final concentration of 1 mg/mL, 67 µM). After the addition of 3-iodopropyldimethylamine iodide (10 mg), $NaBH_4$ (2 mg) was quickly added and mixed, and the reaction incubated in the glovebox for 15 min at RT. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 15267 Da. Observed mass: 15268 Da

Calculated double addition mass: 15353 Da. Observed mass: 15353 Da

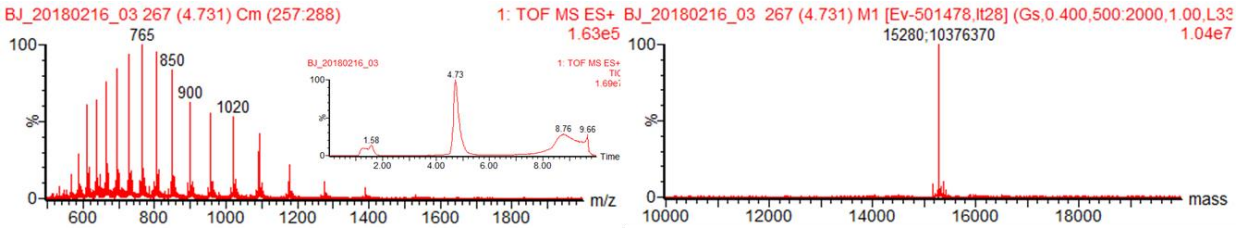


Installation of Me_3Lys on histone H3-Dha36

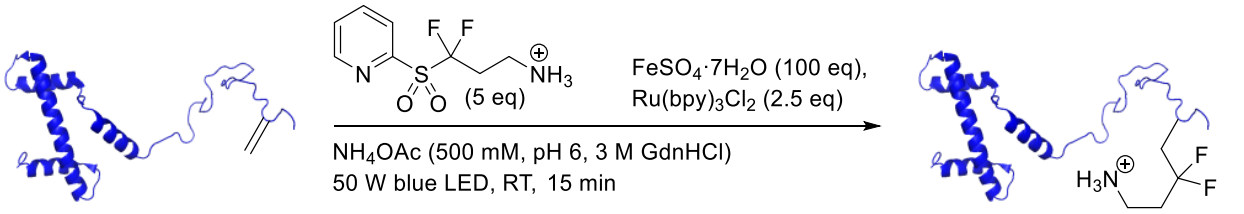


In the glovebox, an Eppendorf tube was charged with NH_4OAc buffer (500 mM, pH 6, 3 M Gdn HCl, 100 µL) containing histone H3-Dha36 (100 µg, final concentration of 1 mg/mL, 67 µM). After the addition of 3-iodopropyltrimethylamine iodide (20 mg), $NaBH_4$ (2 mg) was quickly added and mixed, and the reaction incubated in the glovebox for 15 min at RT. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 15280 Da. Observed mass: 15280 Da



Installation of γCF$_2$-Lys on histone eH3.1-Dha18



histone eH3.1-Dha18
1.5 mg protein in 0.5 mL buffer

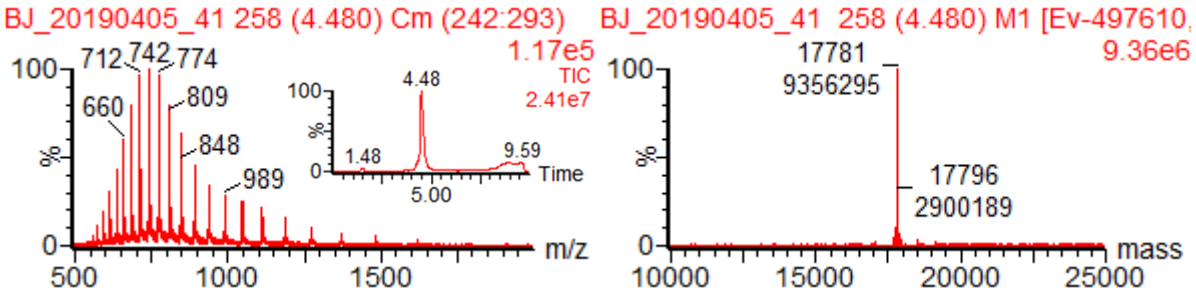
Full conversion
96% yield

histone eH3.1-γCF$_2$-Lys18
1.44 mg after purification

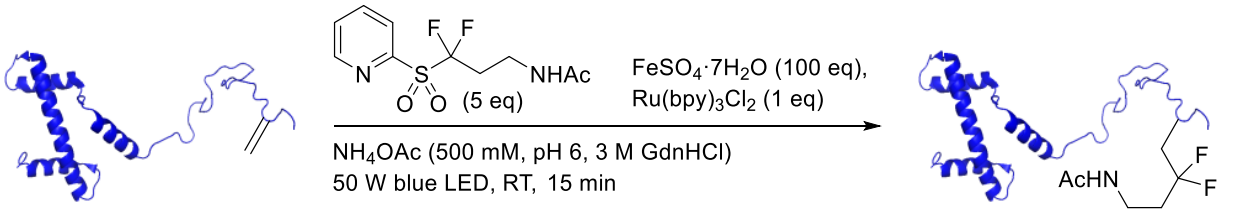
In the glovebox, a glass HPLC vial was charged with an aliquot of histone eH3.1-Dha18 (1.5 mg) and diluted with NH_4OAc (500 mM, pH 6, 3 M GdnHCl) to a final protein concentration of 3 mg/mL. After the addition of 3,3-difluoro-3-(pyridin-2-ylsulfonyl)propan-1-amine trifluoroacetate (5 eq from a DMSO 20 mM stock), $Ru(bpy)_3Cl_2$ (2.5 eq in 2 µL water) and $FeSO_4 \cdot 7H_2O$ (100 eq in 4 µL water) the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 15 min. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17781 Da. Observed mass: 17781 Da

Calculated Met Ox mass: 17797 Da. Observed Met Ox mass: 17796 Da



Installation of γCF$_2$-AcLys on histone eH3.1-Dha18



histone eH3.1-Dha18
1.5 mg protein in 1.5 mL buffer

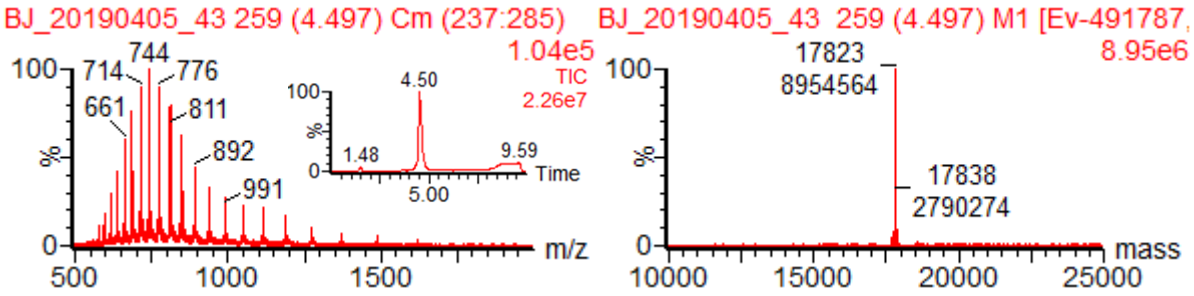
>95% conversion
95% yield

histone eH3.1-γCF$_2$-AcLys18
1.43 mg after purification

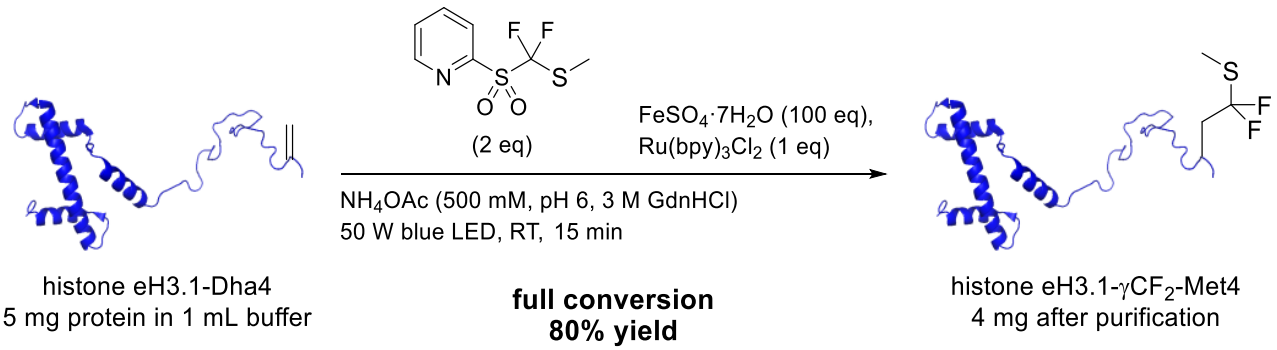
In the glovebox, a glass HPLC vial was charged with an aliquot of histone eH3.1-Dha18 (1.5 mg) and diluted with NH_4OAc (500 mM, pH 6, 3 M GdnHCl) to a final protein concentration of 3 mg/mL. After the addition of *N*-(3,3-difluoro-3-(pyridin-2-ylsulfonyl)propyl)acetamide (5 eq from a 20 mM DMSO stock), $Ru(bpy)_3Cl_2$ (1 eq in 2 µL water) and $FeSO_4 \cdot 7H_2O$ (100 eq in 4 µL water) the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 15 minutes. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17823 Da. Observed mass: 17823 Da

Calculated Met Ox mass: 17839 Da. Observed Met Ox mass: 17839 Da



Installation of γCF_2-Met on histone eH3.1-Dha4



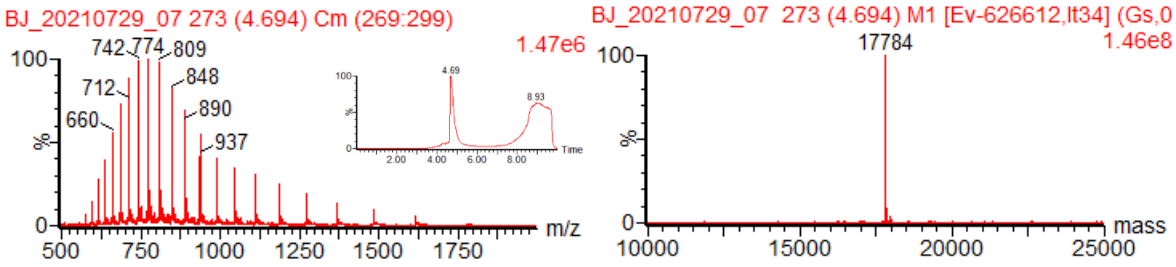
histone eH3.1-Dha4
5 mg protein in 1 mL buffer

full conversion
80% yield

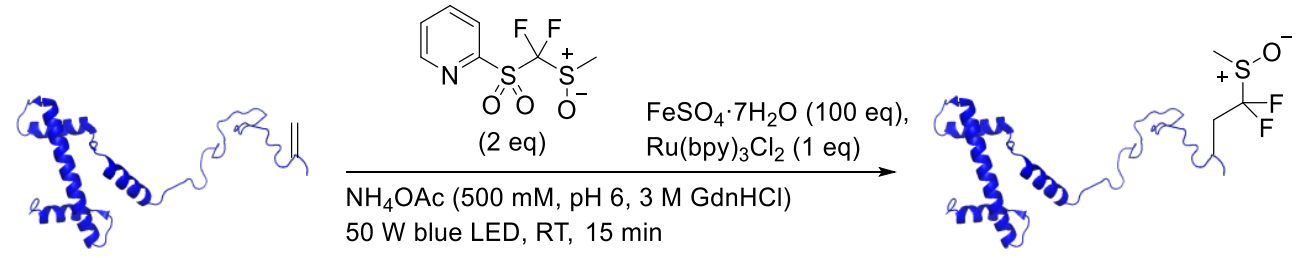
histone eH3.1-γCF_2-Met4
4 mg after purification

In the glovebox, a glass HPLC vial was charged with an aliquot of histone eH3.1-Dha4 (5 mg) and diluted with NH_4OAc (500 mM, pH 6, 3 M GdnHCl) to a final protein concentration of 5 mg/mL. After the addition of 2-((difluoro(methylthio)methyl)sulfonyl)pyridine (2 eq from a DMSO 20 mM stock), $Ru(bpy)_3Cl_2$ (1 eq in 2 µL water) and $FeSO_4·7H_2O$ (100 eq in 4 µL water) the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 15 min. The reaction was quenched with βME (10 µL) and immediately buffer exchanged for desired applications. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17784 Da. Observed mass: 17784 Da



Installation of γCF_2-Met(O) on histone eH3.1-Dha4



histone eH3.1-Dha4
5 mg protein in 1 mL buffer

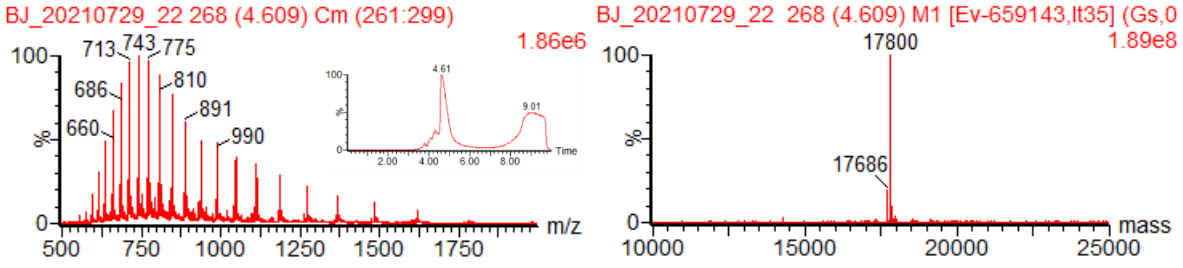
86% conversion
80% yield

histone eH3.1-γCF_2-Met(O)4
4 mg after purification

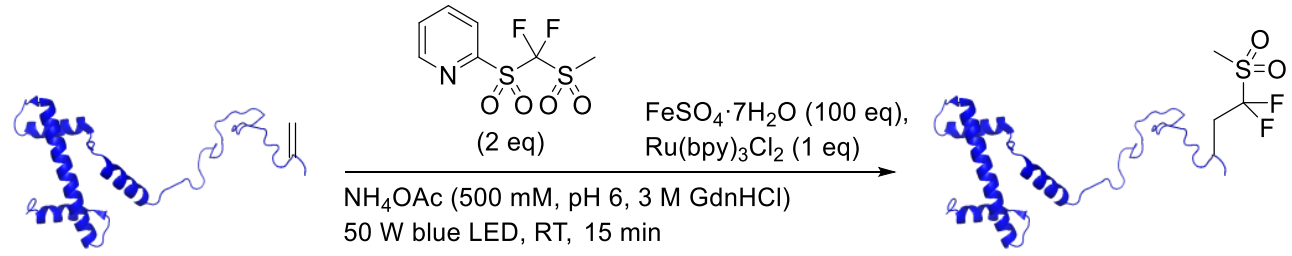
In the glovebox, a glass HPLC vial was charged with an aliquot of histone eH3.1-Dha4 (5 mg) and diluted with NH_4OAc (500 mM, pH 6, 3 M GdnHCl) to a final protein concentration of 5 mg/mL. After the addition of 2-((difluoro(methylsulfinyl)methyl)sulfonyl)pyridine (2 eq from a DMSO 20 mM stock), $Ru(bpy)_3Cl_2$ (1 eq in 2 µL water) and $FeSO_4 \cdot 7H_2O$ (100 eq in 4 µL water) the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 15 min. The reaction was quenched with βME (10 µL) and immediately buffer exchanged for desired applications. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17800 Da. Observed mass: 17800 Da

Calculated Dha mass: 17686 Da. Observed mass: 17686 Da



Installation of γCF$_2$-Met(O$_2$) on histone eH3.1-Dha4



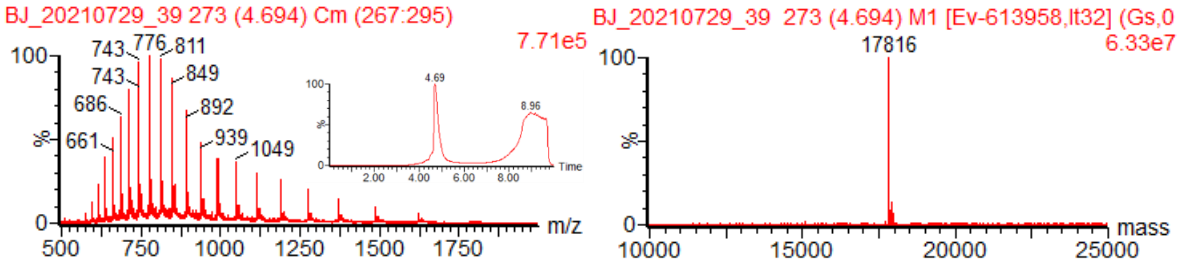
histone eH3.1-Dha4
5 mg protein in 1 mL buffer

full conversion
80% yield

histone eH3.1-γCF$_2$-Met(O$_2$)4
4 mg after purification

In the glovebox, a glass HPLC vial was charged with an aliquot of histone eH3.1-Dha4 (5 mg) and diluted with NH_4OAc (500 mM, pH 6, 3 M GdnHCl) to a final protein concentration of 5 mg/mL. After the addition of 2-((difluoro(methylsulfonyl)methyl)sulfonyl)pyridine (2 eq from a DMSO 20 mM stock), $Ru(bpy)_3Cl_2$ (1 eq in 2 µL water) and $FeSO_4 \cdot 7H_2O$ (100 eq in 4 µL water) the vial was sealed with a cap, transferred out of the glovebox and irradiated with blue LED light (50 W) for 15 min. The reaction was quenched with βME (10 µL) and immediately buffer exchanged for desired applications. Conversion was determined by analysis of an aliquot of the crude mixture by LC-MS.

Calculated mass: 17816 Da. Observed mass: 17816 Da



Chapter 3: Incorporation of Extracellular Histones into Chromatin and Exploring *In Vivo* Histone Modification Chemistry

“The zebrafish is just a bag for the reaction, it doesn’t matter if it’s a dead bag”

- Professor Benjamin Guy Davis

Introduction

Background on extracellular histone incorporation

As discussed thoroughly in the thesis **Introduction**, extracellular histones possess a unique ability to penetrate cell membranes and localize to the nucleus, although the mechanism of this, the effect on the cell, and the role/fate of the incorporated histones, are unclear and understudied[144]. Intact nucleosomes can enter cells too, but more slowly than free histones, resulting in localization in the cytosol[172]. Additionally, free histones in the blood trigger a robust inflammatory response which is not observed with intact nucleosomes, although there has been a traditional lack of distinction between the two in literature[141,142]. Overall, there is a general lack of characterization surrounding the roles of free histones and nucleosomes in circulation, with a complete omission of data on the fate of histones once they are localized in the nucleus.

This chapter aims to investigate, characterize, and leverage the ability of extracellular free histones to penetrate cells and enter the nucleus, and when appropriate, compare it with that of intact nucleosomes. The chemical protein modification methods developed and/or utilized in previous chapters[208] were used here to either install desired histone PTMs of interest prior to cellular incorporation, tag histones with a trackable ^{18}F-radiolabel to follow their fate in injected animals, or test potential PTM-installation chemical methods on Dha-tagged histones *in vivo*. Furthermore, conventional biochemical techniques such as western blotting, immunoprecipitation, immunohistochemistry (IHC), and RNA-seq were used to explore and characterize the roles of extracellular histones in a variety of model organisms, such as HeLa cells, tail vein or brain injected rodents, and microinjected zebrafish embryos and larvae (**Figure 49**).

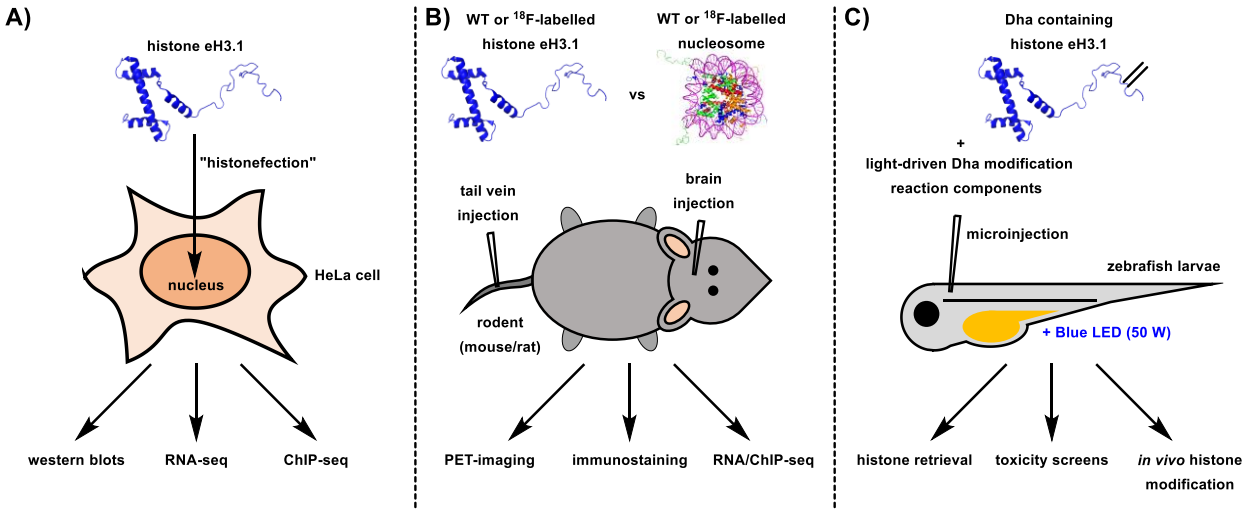


Figure 49: Investigating the incorporation and chemical modification of extracellular histones and/or nucleosomes in **(A)** HeLa cells, **(B)** brain and tail injected rodents, and **(C)** microinjected zebrafish embryos and larvae.

Initial investigations into histone uptake into HeLa cells

Choice of a "trackable" epitope tagged histone H3

To allow for the identification of extracellular histones after their incorporation into the nucleus, and possibly the chromatin, of target cells, they needed to be distinguishable from the native histones. To accomplish this, we decided to use an epitope tagged histone variant, named histone eH3.1 (**Figure 50**) in previous literature (e stands for epitope tag, a dual C-terminal FLAG-HA tag)[301-303].

This model histone was chosen for several reasons:

1) The C-terminal epitope tags were less likely to interfere with the histone's cell membrane penetrating ability that relies on the N-terminal disordered histone tail[148].

2) Histone eH3.1 has been successfully used in many *in vivo* experiments, with established protocols for its extraction[301,302], and evidence that the epitope tags do not interfere with nucleosome deposition or its native function[303].

3) The histone H3.1 variant, the most abundant H3 variant in cycling cells[301], is cell cycle dependent, being expressed preferentially during S-phase to provide enough histones for new nucleosome formation during DNA replication[304,305]. We reasoned that synchronizing cells to G1/S via a double thymidine block[306], followed by releasing them into S-phase upon adding extracellular histones would allow the cell penetrating eH3.1 histones to be efficiently incorporated into new nucleosomes by the cell's native machinery.

Lastly, to allow implementation of our various Dha modification chemical strategies, we mutated the two native Cys residues (Cys96 and Cys110) to Ala, so that Cys residues installed at sites of interest via genetic mutation could be selectively transformed to Dha and used for chemical modification[127,226,227]. These Cys mutations have been used extensively for Cys based histone modification experiments, and are believed to cause negligible effects on nucleosome structure or histone function[307]. Therefore, in this thesis, histone eH3.1-WT refers to a histone containing Cys96 and Cys110 to Ala mutations in all cases (**Figure 50**).

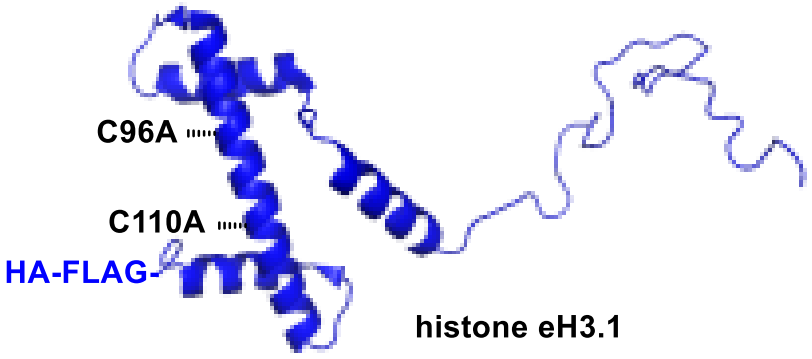


Figure 50: Portrayal of the model histone eH3.1 used throughout this thesis, highlighting Cys to Ala mutations and the C-terminal dual FLAG-HA epitope tags.

Initial investigations into histone H3 uptake into Hela cells

To begin our investigations, a simple model system of cultured WT HeLa cells was treated with conditions modeling one of the early reports of direct histone translocation across cell membranes[151]. WT HeLa cells (1 x 10^6 cells in a T75 flask) were incubated with a 1:1 equimolar mixture of histone eH3.1-WT and histone H4-WT (1 μM each, PBS, 37 °C, 15 min) followed by three PBS washes and recovery (1 or 24 h) in normal growth media and conditions (DMEM, FBS 10%, L-glutamine 2 mM, 37 °C, 5% CO_2 incubator) (**Figure 51A**). The mixture of histone eH3.1 and histone H4 was used to improve the solubility in PBS, as histone eH3.1 alone was observed to crash out over time. However, we further rationalized their use together both due to previous literature showing higher incorporation efficiencies with mixtures of the two[157], and the fact that histones H3 and H4 are chaperoned from the cytosol to the nucleus as a tetramer[139].

Following the WT HeLa cell histone treatment and recovery, the cells were harvested (via Trypsin-EDTA detachment), and fractionated to produce cytoplasm, nuclear lysate, and chromatin (insoluble pellet) fractions (**Figure 51A**). Western blots against the three cell fractions revealed that the incorporated histone eH3.1 (detected via the FLAG tag) was present primarily in the chromatin fraction of the 1 h recovery sample, but had experienced multiple cleavages, reducing its length (**Figure 51B, top**).

We reasoned that the observed histone eH3.1 cleavage was due to the constant use of Trypsin-EDTA for cell detachment during cell splitting and harvesting. Therefore, we repeated the experiment using only physical detachment methods (cell scraping). Pleasingly, this resulted in the presence of full-length histone eH3.1 (via FLAG tag detection) primarily in the chromatin fractions of both the 1 h and 24 h recovery conditions,

with no bands detected in the no histone control (**Figure 51B, second from top**). Furthermore, the control western blot against native histone H3 showed a faint band at higher molecular weight indicative of the incorporated histone eH3.1, distinguishable due to the added mass of the epitope tags (**Figure 51B, second from bottom, ***). This indicates that after 1 h, a substantial percentage of histone H3 in the HeLa cell chromatin fraction is composed of incorporated extracellular histones.

Inspired by these findings, we repeated the experiment with various unnaturally modified variants of histone eH3.1, and successfully detected histone eH3.1-Dha9 and the alkyl halide containing histone eH3.1-Bhn9 (via the FLAG tag) in WT HeLa cell chromatin as well (**Figure 51C**). Although admittedly these early attempts produced some messy western blot results, this still gave great promise for the incorporation of full length chemically modified histones for a variety of future epigenetics applications.

Following this, a histone eH3.1 concentration and recovery time screen was attempted on HeLa cells (0.1×10^6 cells in 6-well plates at start). Remarkably, even at nM incubation concentrations on <1 million cells, histone eH3.1 could be detected in WT HeLa cell chromatin up to 3 days after the histone exposure, surviving three cell divisions and two rounds of splitting (**Figure 51D**).

We reasoned that, due to the long-lived nature of histones once imbedded into nucleosomes and chromatin[308], this was evidence that our incorporated histones were becoming integrated into nucleosomes and being utilized by the HeLa cells, fulfilling their native function. This had enormous implications for the fate of extracellular histones, possibly PTM-laden[163], that are present in the blood[140,142,164], and also hinted at the possibility of incorporating pre-modified histones into the native chromatin of cells.

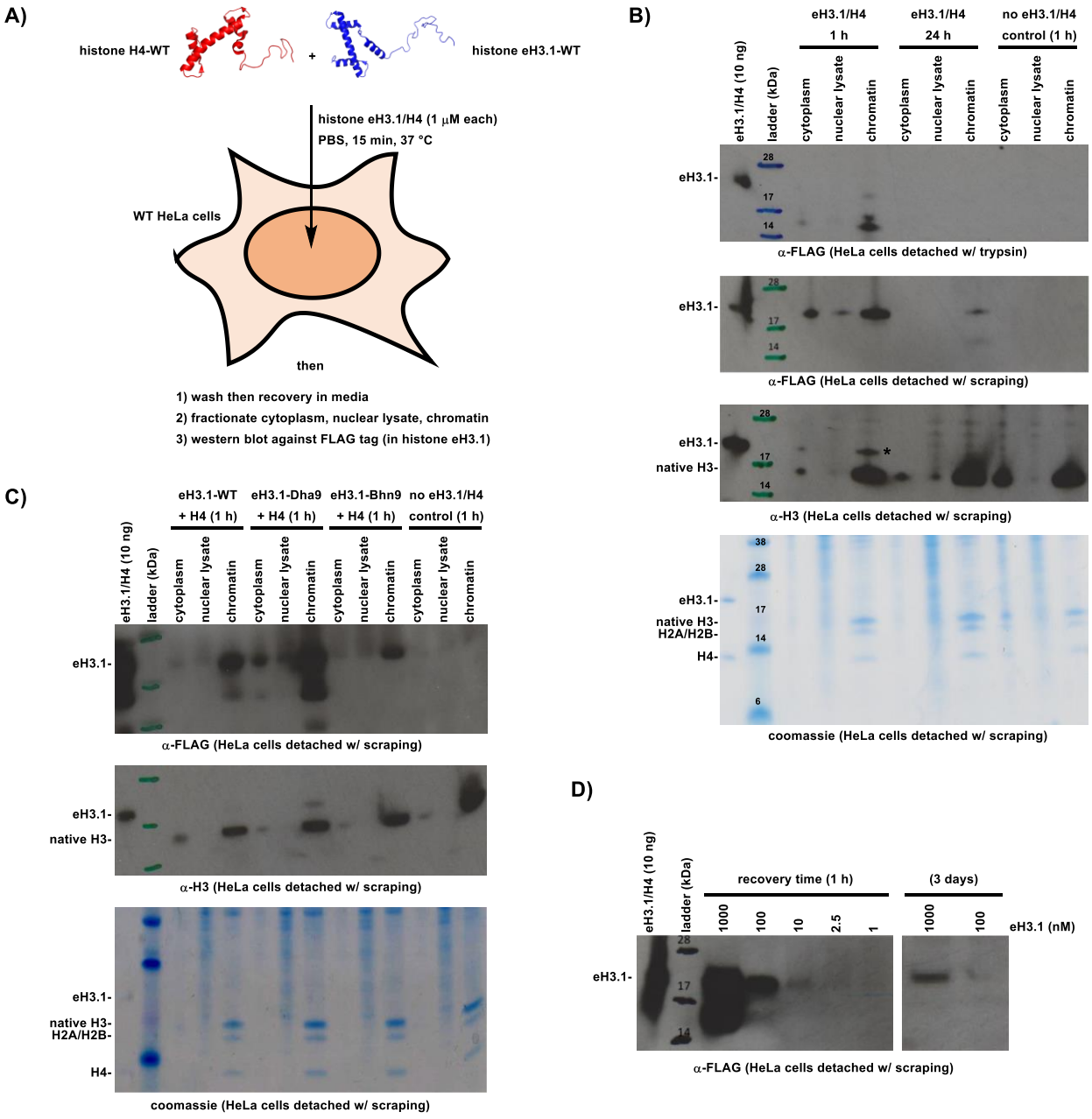


Figure 51: A) Experimental layout for histone incorporation into HeLa cells followed by cell fractionation and detection by western blot. **B)** Initial western blots found histone eH3.1-WT in the chromatin fraction after incubation with HeLa cells. **C)** Unnaturally modified histones could be incorporated. **D)** Incorporated histones could be detected at low initial incubation concentrations and after several days in HeLa cells.

Transcriptional response of HeLa cells to extracellular histone uptake assessed by RNA-seq

Histone incorporation, uptake validation, and RNA prep

Now that we had confirmed and characterized the ability of extracellular histones to penetrate HeLa cells and incorporate into chromatin, we next ventured to quantify the transcriptional response of the HeLa cells to this histone incorporation event. Furthermore, we were curious about the possible effects of incorporating PTM-containing histones, and therefore we synthesized four site-specific trimethyllysine modified histones with the light-driven radical Dha modification reaction from **Chapter 1**: the active euchromatin associated histone eH3.1-Me$_3$Lys4, facultative heterochromatin associated eH3.1-Me$_3$Lys9, relatively understudied eH3.1-Me$_3$Lys18, and constitutive heterochromatin associated eH3.1-Me$_3$Lys27 (**Figure 52A**)[68].

To help create a consistent transcriptional background between WT HeLa cells, they were synchronized to G1/S phase via a double thymidine block protocol[309]. This also created an environment where, once HeLa cells were released into S-phase, the incorporation of histone eH3.1 would mimic the native expression and accumulation of the histone H3.1 variant in the cytosol that occurs during S-phase[304,305]. Once G1/S synchronized, WT HeLa cells (~80% confluent in 12-well plates, 4 biological replicates per condition) were treated with a 1:1 equimolar mixture of histone eH3.1 and histone H4 (1 μM each, PBS, 15 min, 37 °C), washed 3x with PBS, and allowed to recover in normal growth media for 6 h (**Figure 52B**). Samples included a no histone control, histone eH3.1-WT, or histone eH3.1-Me$_3$Lys4/9/18/27 mixed with WT histone H4.

After recovery, the HeLa cells were harvested by physical detachment (cell scraping), and lysed with TRIzol reagent (Thermo Fisher, cat. no. 15596026). This allowed for the total RNA and protein content of the cells to be separated and purified (Direct-zol Miniprep Plus Kit). SDS-PAGE gels and western blots of both the starting material histone mixtures and the TRIzol reagent protein extracts revealed successful uptake of all the histone variants minus the no histone control (**Figure 52C**). However, it must be noted that the histone eH3.1-Me_3Lys18 starting material showed low histone concentrations compared the other histone samples, further evident by the respective extremely faint band seen for histone eH3.1-Me_3Lys18 in the TRIzol reagent protein extract. This is likely due to an error in preparation, and we decided to continue with RNA-seq experiments.

To prepare for RNA-seq, the mRNA was enriched from the total RNA (NEBNext Poly(A) mRNA Magnetic Isolation Module). Libraries were prepared (NEBNext Ultra II Directional RNA Library Prep Kit for Illumina), then multiplexed and quantified using a High-sensitivity d1000 TapeStation (Agilent), then sequenced using a NextSeq 500 (Illumina). Target sequencing depth was >20 million reads per sample library. Special thanks to Dr. Martin Philpott in the group of Professor Udo Oppermann for assistance during RNA-seq experiments.

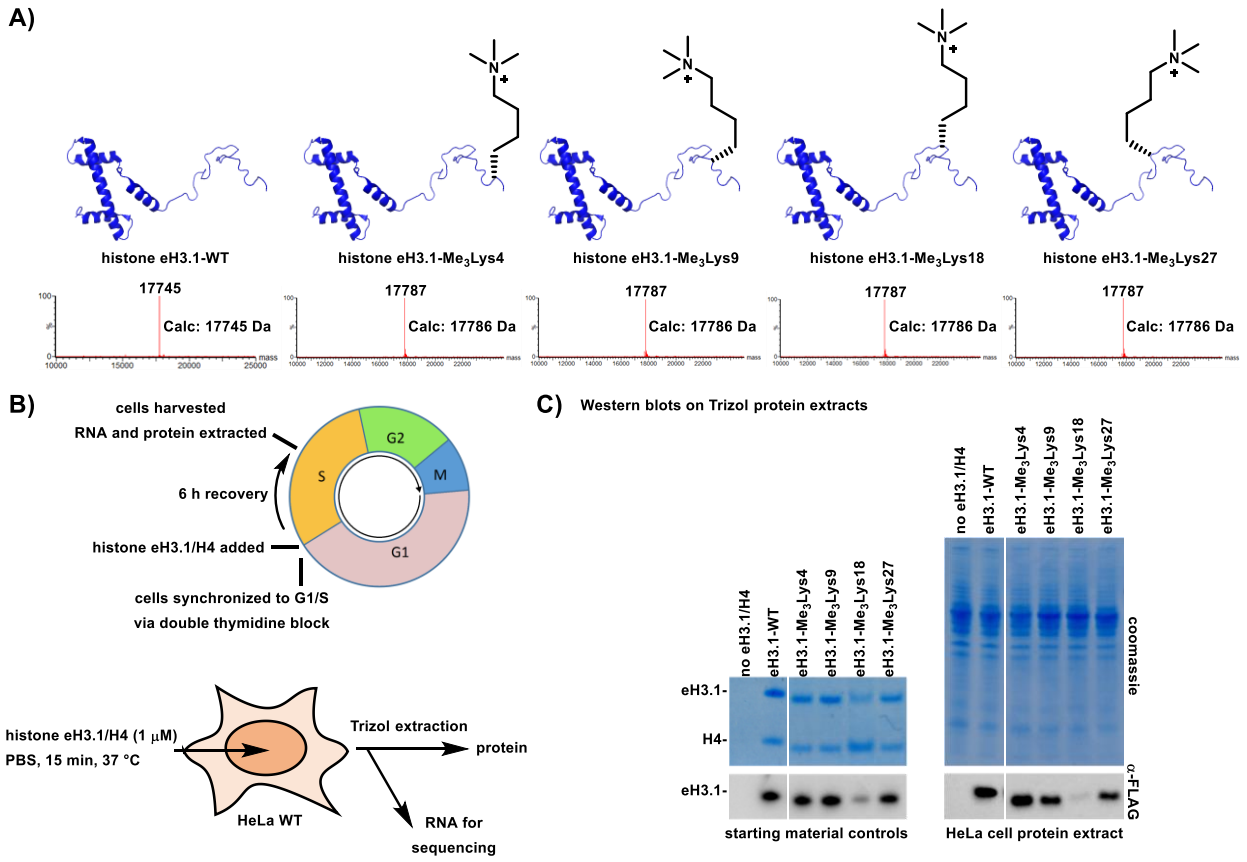


Figure 52: A) All of the histone PTMs used, with intact protein LC-MS deconvoluted spectra characterization. **B)** Experimental design with WT HeLa cell synchronization at G1/S phase via a double thymidine block followed by histone incorporation then RNA and protein extraction with TRIzol reagent 6 h later. **C)** Gels and western blots on TRIzol reagent protein extracts indicating successful histone incorporation.

RNA-seq results and analysis

With RNA sequencing results in hand, we first compared the condition of histone eH3.1-WT incorporation with that of the no histone control (**Figure 53A**). The data is portrayed as a volcano plot, with the y-axis representing statistical significance (p-value), and the x-axis giving the fold-change of transcription between histone eH3.1-WT and the no histone control for identified gene transcripts (Dr. Adam Cribbs performed sequencing analysis). Notably, several genes showed significantly heightened levels of transcription in response to histone eH3.1 incorporation, showing the effects on transcription when a cell is flooded with a “naked” unmodified histone.

Particular standouts were the FOS family proteins FOS and FOSB (regulators of cell proliferation, differentiation, and transformation as members of the activator protein 1 complex)[310,311], and EGR1 (early growth response protein 1), a transcriptional regulator that modulates the immune response[312,313]. Moreover, there were many genes that, while displaying lower log2 fold changes, showed highly significant changes (-log10(padj) >10). Some, such as PPP1R15A and ATF3 are known to either promote growth arrest and DNA damage[314] or are expressed upon physiological stress[315], respectively.

Cumulatively, this data provides the first quantitative analysis of the cellular response to extracellular histone uptake. Perhaps unsurprisingly, the results align with the observed effects of free histones in circulation, which provoke strong inflammatory and immune responses[140,141,164]. It is possible that the dramatic cellular response (one of growth arrest, damage response, and immune response) against extracellular histone uptake is a defense mechanism against the uptake of external histones (and their PTMs) from entering the cell and altering function.

Next, results from the four modified histones, eH3.1-Me$_3$Lys4, eH3.1-Me$_3$Lys9, eH3.1-Me$_3$Lys18 and eH3.1-Me$_3$Lys27, were compared to the unmodified histone eH3.1-WT and portrayed as volcano plots (**Figure 53B**). Surprisingly, there were no significant transcriptional changes due to any of the pre-installed PTMs. It should be noted that the lower fold changes of EGR1, FOS, FOSB and ATF3 seen with histone eH3.1- Me$_3$Lys18 are almost certainly due to the lower histone concentration inadvertently used during HeLa cell incubations (discussed above).

The lack of any transcriptional effects by the four pre-installed PTMs may have been due to several reasons. First, the incorporated extracellular histones might have been deposited randomly in the genome (speculation at this point), causing specific transcriptional effects to 'average out' when mRNA from thousands of cells was taken for sequencing. Second, the time scale of the experiment (6 h cell recovery after histone incorporation) may have been too short or too long to capture histone PTM specific transcriptional changes. Lastly, to have caused a general transcriptional change, the histone PTMs would need to have been recognized by specific readers that would cause a feedback loop leading to differential gene expression. This may not exist for the tested histone PTMs but may exist for others.

However, on a positive note, the ability to incorporate an epigenetically relevant PTM of interest on a full-length histone into living cells, with minimal effects on the transcriptome compared to that of the WT histone, means that specific incorporated histone PTMs could be studied in their native chromatin context, without altering cell behavior outside of the effects of the incorporated WT histone control.

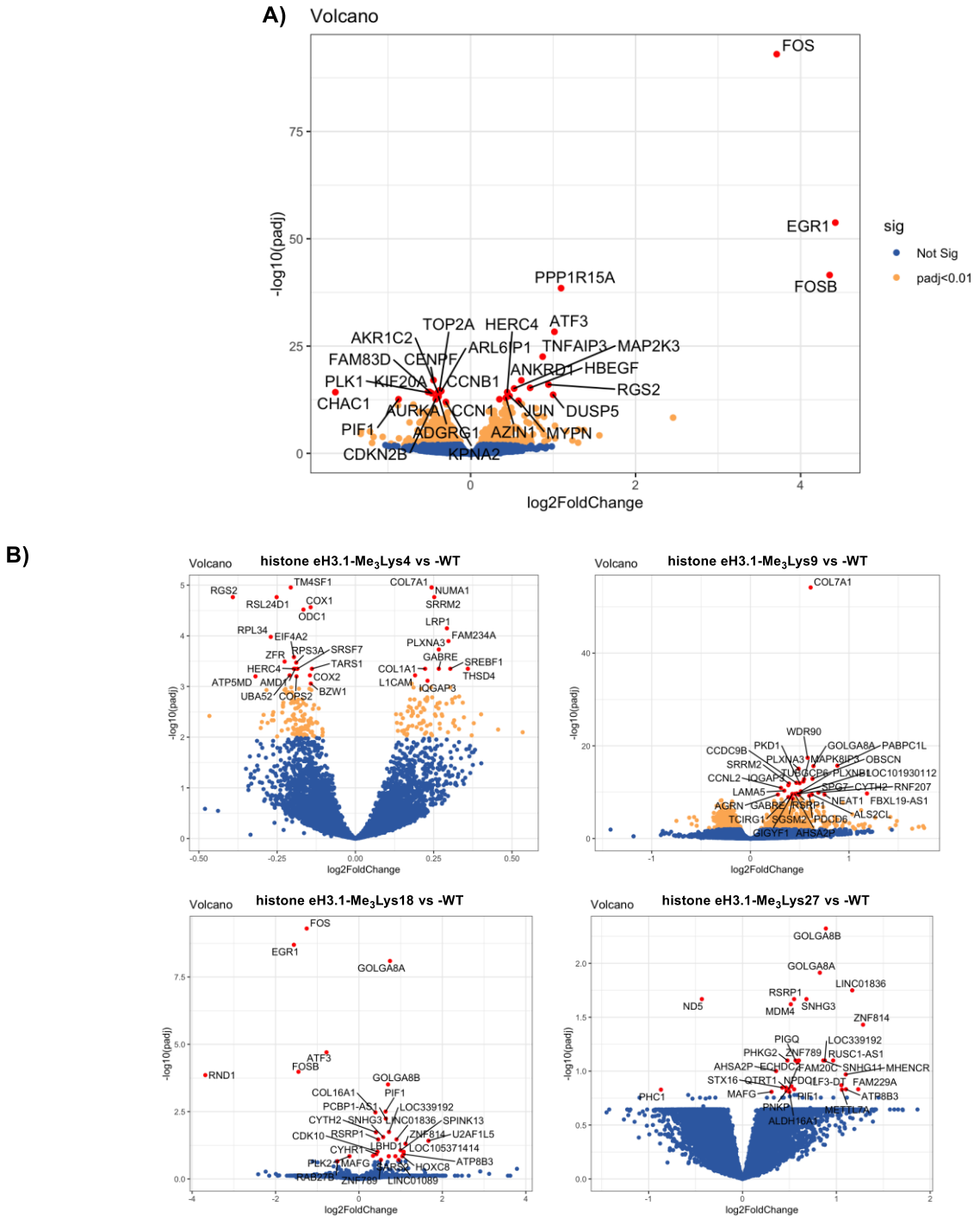


Figure 53: A) RNA-seq results of histone eH3.1-WT/H4 incorporation vs no histone control, displayed in a volcano plot. **B)** Volcano plots of PTM-containing histones vs WT showing lack of significant transcriptional changes in this experiment.

ChIP-seq and future plans

Many mysteries remained in our investigation on extracellular histone incorporation into HeLa cells. First, while the incorporated histones were always identified in the insoluble chromatin pellet, we had no solid evidence that they were fulfilling their native functions by being incorporated into nucleosomes. Second, if incorporated histones were imbedded into native nucleosomes, perhaps during S-phase DNA replication, were there preferential genomic loci where they would get incorporated, or would they be randomly dispersed throughout the genome.

To address these questions, we designed a ChIP-seq[88] experiment (**Figure 54A**), which would identify the localization of incorporated histone eH3.1 by sequencing the DNA from nucleosomes enriched via the HA epitope tag in histone eH3.1. Similar to the RNA-seq experiment, G1/S synchronized WT HeLa cells (50% confluent in 150 mm dishes, 3 biological replicates per condition) were treated with a 1:1 equimolar mixture of histone eH3.1 and histone H4 (1 μM each, PBS, 15 min, 37 °C) or no histone control, washed 3x with PBS, and allowed to recover in normal growth media for 36 h. The longer recovery time was to allow for cells to divide once and go through S-phase twice after histone treatment, giving incorporated histones extra time to be inserted into nucleosomes. Cells were then harvested by physical detachment (cell scraping), fixed with formaldehyde (to crosslink nucleosomes), quenched with glycine, and the chromatin extracted (Chromatin Extraction Kit-Flexible Format, Abcam, ab223876), then sheared via sonication to release soluble polynucleosome fragments.

The soluble chromatin was analyzed by western blot (**Figure 54B**) after first reversing the crosslinks (boiling for 1 h). Pleasingly, histone eH3.1 was clearly visible in the soluble

chromatin fraction, along with some suspected cleaved eH3.1, but not in the no histone control lane as expected. Native histone H3 was also present in the control western blot using an α-histone H3 antibody, showing successful chromatin extraction. These results provide indirect evidence that the incorporated histones are present in nucleosomes, as they are present in the soluble nucleosome fraction after chromatin shearing via sonication, but further validation should be undertaken.

Future work, to be done by colleague Adeline Poh, will use the chromatin samples extracted here for ChIP-seq utilizing the HA epitope tag on the incorporated histone eH3.1. These results should identify the genomic localization preference of extracellular histone eH3.1 after cellular uptake and chromatin incorporation.

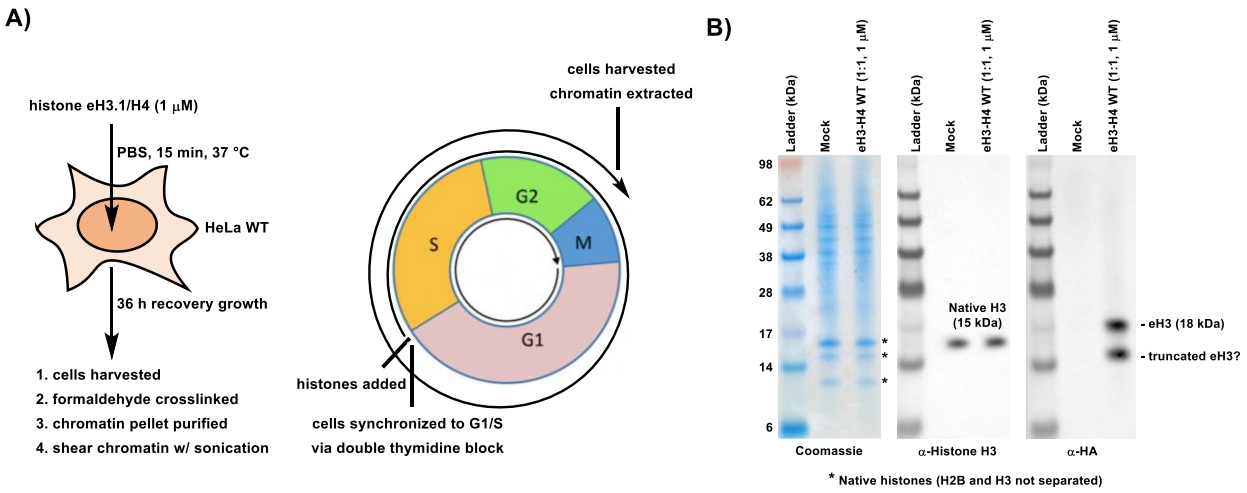


Figure 54: A) Experimental design for histone incorporation into G1/S synchronized WT HeLa cells (via double thymidine block) followed by chromatin extraction for eventual ChIP-seq (after 36 h incubation). **B)** Western blots confirm the presence of incorporated histone eH3.1 in the solubilized chromatin (sonicated).

Tracking native and radiolabelled histones injected intravenously or intracranially into rodents

Mouse tail vein injections of WT and ^{18}F-radiolabelled histones

We next desired to characterize the role of extracellular histones in circulation, both in terms of their ability to penetrate cells they reach through the blood, and by comparing their distribution and toxicity to that of extracellular nucleosomes. We hypothesized that the circulation system may provide extracellular histones with the vector needed to spread and become incorporated into the chromatin of the cells throughout the body, leading to the robust damage response seen both in previous literature[140-142,164] and revealed by our RNA-seq experiments (see above).

However, there was a general lack of strategies available for tracking extracellular histones vs nucleosomes in circulation[141], providing us the opportunity to showcase our newest chemical tool, an improvement on the light-driven radical Dha difluoroalkylation reaction that utilized a benzothiazole sulfone reagent to incorporate both ^{19}F- and ^{18}F-radiolabelled sidechains into proteins, including on both Dha containing histones and intact nucleosomes (reaction developed by Adeline Poh, **Figure 55A**). Radiolabeling proteins with ^{18}F allows for their real-time detection in the body via positron emission tomography (PET) imaging[316], which we planned to use to track the distribution of ^{18}F-radiolabelled histones and nucleosomes after tail vein injection in mice.

First, we attempted to recover “cold” (non-radiolabeled) histone eH3.1-WT in the chromatin of various mouse organs, revealed by α-FLAG or α-HA western blots, after tail vein injections (**Figure 55B**). Histone eH3.1-WT (2 mg/mL, 100 µL in PBS) was injected

into mice tail veins (performed by Dr. Abi Yates and Professor Daniel Anthony), along with a PBS control, and incubated for 2 or 24 h (2 biological replicates per condition). Then, 10 different organs were harvested (heart, plasma, buffy coat, brain, spleen, liver, lung, muscle, gut, kidney), and their chromatin extracted (Chromatin Extraction Kit-Flexible Format, Abcam, ab223876). However, extensive western blots failed to detect histone eH3.1 via its FLAG or HA epitope tag in any organ (data not shown).

We decided to proceed with ^{18}F-radiolabeled histone eH3.1-γC^{18}F-hAla4 injections, followed by time-course PET imaging (carried out by Sarah Able and Katherine Vallis). The desired histone was prepared (by Adeline Poh) and injected into mice tail veins (10 MBq in PBS), with PET images (overlayed on CT scans) taken every 2.5 min (**Figure 55C**). The PET images revealed rapid localization of the ^{18}F radiotracer to the liver, then kidney, then bladder, suggesting fast proteolysis of the radiolabeled histone, removing it from the blood before it could become incorporated into surrounding chromatin.

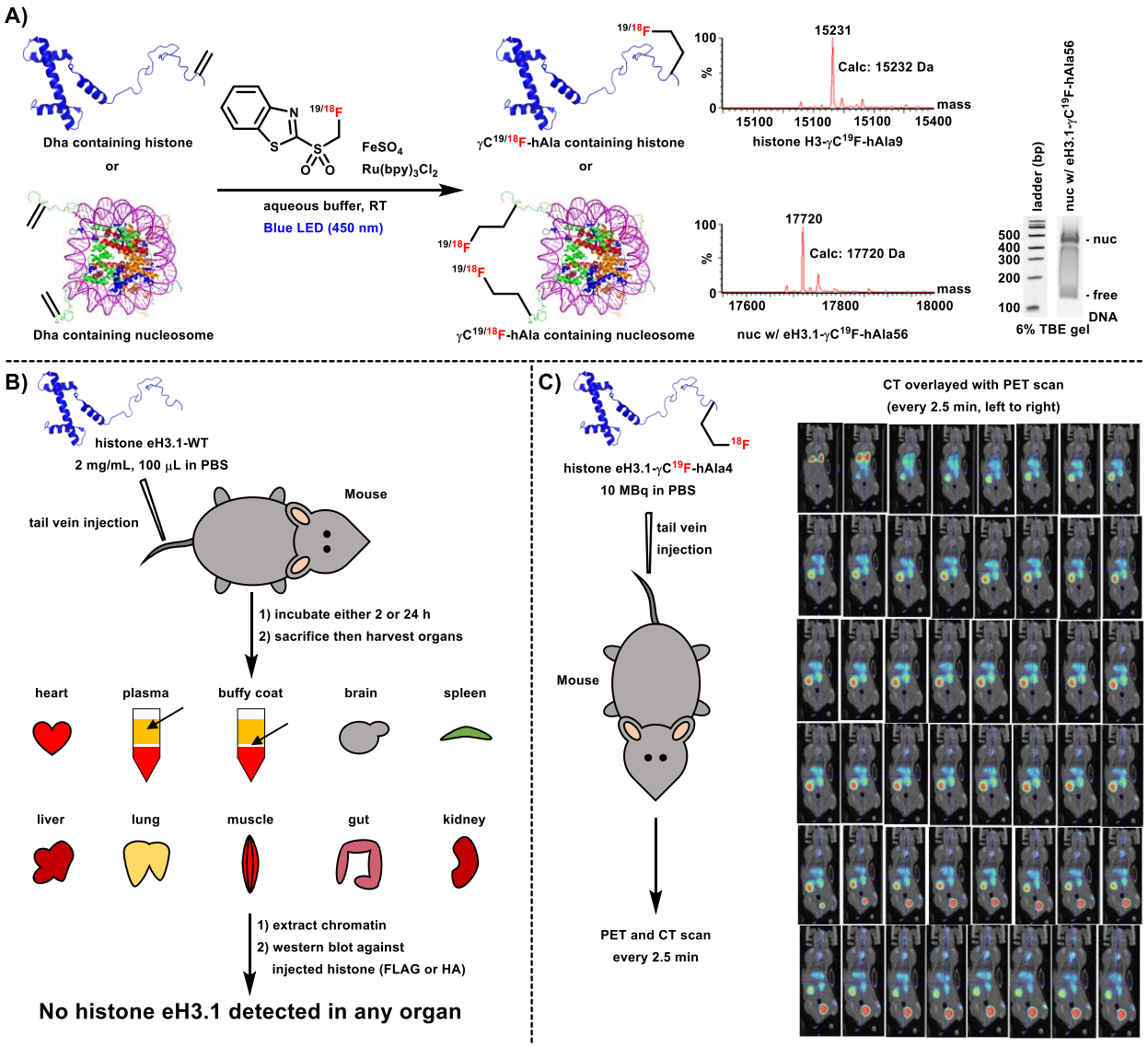


Figure 55: A) Strategy for ^{18}F-radiolabelling Dha-tagged histones and nucleosomes (nuc) with an ^{18}F-benzathiozole sulfone reagent. **B)** Tail vein injection of histone eH3.1-WT into mice followed by organ extractions and western blots revealed no accumulation of histone eH3.1 in any organ. **C)** Tail vein injection of ^{18}F-radiolabelled histone eH3.1 followed by time-course PET and CT scans showed accumulation of injected histones in lung, then liver, then kidney, then bladder, suggesting rapid free histone digestion and removal from the blood. Figure adapted with permission from data provided by Adeline Poh.

Stability of histones and nucleosomes in blood plasma and serum

To further investigate the rapid depletion of WT and ^{18}F-radiolabeled histone eH3.1 from the blood observed above, we incubated either free histone eH3.1-WT, a 1:1 equimolar mixture of histone eH3.1-WT and H4-WT, or intact nucleosomes containing histone eH3.1-WT (prepared by Alice Bedwell) in either PBS, blood plasma or serum (1:10 dilution, derived from rat) (**Figure 56**). Western blotting against the FLAG epitope tag in histone eH3.1 could therefore reveal proteolytic digestion.

Interestingly, while free histone eH3.1-WT and the histone eH3.1-WT and H4-WT mixture were immediately proteolyzed in blood serum (evident by the disappearance of bands in the α-HA western blot after the first 15 min time point), the intact nucleosome remained protected (evident by the retention of bands at the proper eH3.1 size in the α-HA western blot). Furthermore, incubations in blood plasma produced no such proteolytic digestion for any histone species.

We reasoned that the heparin found in blood plasma[317] may have bound to the disordered, basic histone tails due to its negative charge, protecting it from proteolytic digestion prevalent in serum, where histones are known to be targets for various proteases such as activated protein C[171]. A similar interaction between the negatively charged phosphate backbone of the nucleosome DNA and the histone tail may explain the protection afforded to histone eH3.1 within intact nucleosomes. Remarkably, the addition of heparin is known to reduce the toxicity of free histones in circulation[174]. We speculate that the distinction between free histone and nucleosome digestion in serum may be due to their difference in cell penetrating abilities, with free histones preferentially targeted as they can penetrate cells rapidly and incorporate into chromatin, while

nucleosomes do so slowly and only localize to the cytoplasm[172]. Heparin, while protecting histones from proteolytic digestion, may block their ability to penetrate cell membranes, although this is currently speculation. Altogether, our data supports the conjecture that there is a robust defense mechanism and strong immunological response, both within cells (from RNA-seq experiments described above) and in the circulation system, that protects against the cell penetrating and chromatin incorporating abilities of extracellular free histones specifically.

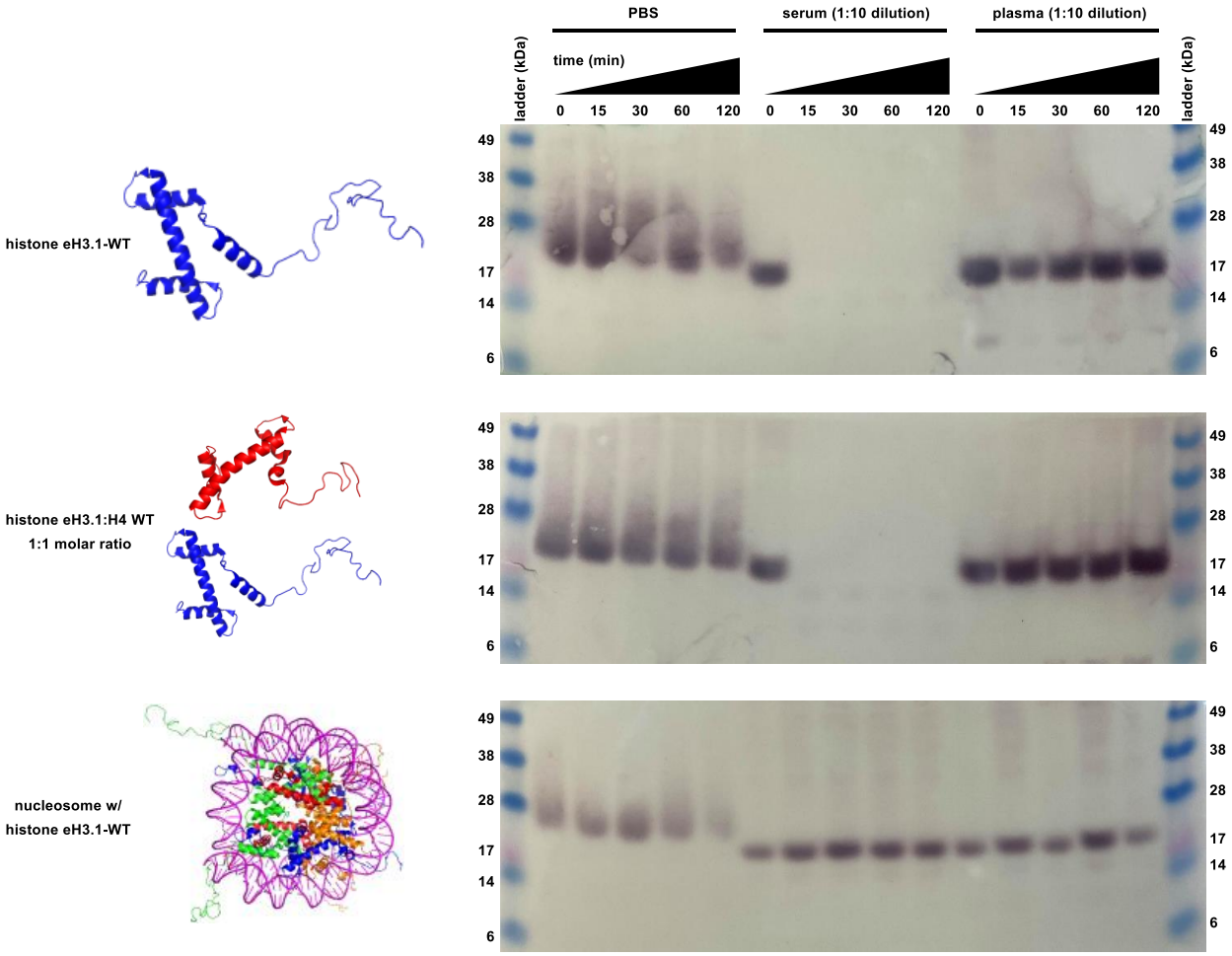


Figure 56: Western blot (α-HA) results from incubations of free histone eH3.1, histone eH3.1-H4 mixture, and eH3.1 containing nucleosomes (nuc) in PBS, plasma, or serum from rats. Figure adapted with permission from data provided by Adeline Poh.

Direct injection of free histones into rat brains followed by detection in chromatin

Due to the limitations described above in investigating free histones in circulation, which are susceptible to rapid proteolysis, we decided to bypass the circulation system and instead simulate organ injury, a known path to releasing free histones[140,142,164], by directly injecting free histones and nucleosomes intracranially in mice. Our main goal was to extract the chromatin from injected mice brains and detect the incorporation of histone eH3.1 by western blot, proving that histone uptake and chromatin localization happens in complex tissues and organisms in response to organ damage.

An equimolar mixture of histone eH3.1-WT and histone H4-WT was injected into mice brains (performed by Dr. Abi Yates, bilateral injection, 1 µg per side, in PBS) along with no histone control injections (PBS, 3 biological replicates per condition). After a 6 h recovery, mice were sacrificed, and brains were excised and fixed with formaldehyde before chromatin was extracted (Chromatin Extraction Kit-Flexible Format, Abcam, ab223876) and then sheared by sonication to produce a soluble polynucleosome fraction (**Figure 57A**). Notably, the mice brains injected with free histones (eH3.1 and H4 WT) displayed visible lesions, while mock injections showed no damage. Pleasingly, western blots against the histone eH3.1 HA epitope tag produced expected bands in the soluble chromatin fractions, with only faint bands in the lysate fraction and no bands in the mock condition (**Figure 57A, right**). This indicates that histone eH3.1 successfully penetrated cells and incorporated into the chromatin of mice brain cells.

We next compared the distribution of injected free histone eH3.1 and H4 WT (equimolar mixture) with that of intact nucleosomes in mice brains (**Figure 57B**). Mice were injected

with either the histone eH3.1 and H4 WT mixture or intact nucleosomes containing histone eH3.1 (unilaterally, 1 µg in PBS, 3 biological replicates per condition) and allowed to recover for 3 h (we chose to shorten the recovery time to reduce the amount of damage and cell death prior to staining). Mice were then sacrificed, and formaldehyde fixed brain slices were subjected to immunohistochemistry (α-HA, performed by Dr. Abi Yates) (**Figure 57C**). Intact nucleosome injected mice brains exhibited minimal damage, with nucleosomes widely diffused throughout the brain slice (**Figure 57C, left**). Free histone injected mice brains, in contrast, exhibited the same extensive lesion damage seen earlier, with limited histone diffusion, likely due to rapid cellular uptake in regions local to the injection site, leading to cell death (**Figure 57C, right**).

Future work (to be undertaken by Adeline Poh), will perform ChIP-seq experiments on the histone eH3.1 containing mice brain chromatin (**Figure 57A**) to determine where in the genome the incorporated histones preferentially localize. Furthermore, tissue from histone eH3.1 or intact nucleosome injected mice brains will be subjected to transcriptomic analysis to reveal the transcriptional response of cells exposed to extracellular histones (seen here to enter cells and incorporate into chromatin) or extracellular nucleosomes. This work will provide much needed quantitative analysis of the functions and consequences of extracellular histones vs nucleosomes, highlighting the ability of free histones to penetrate cells and incorporate into native chromatin, accompanied by a strong immunological response.

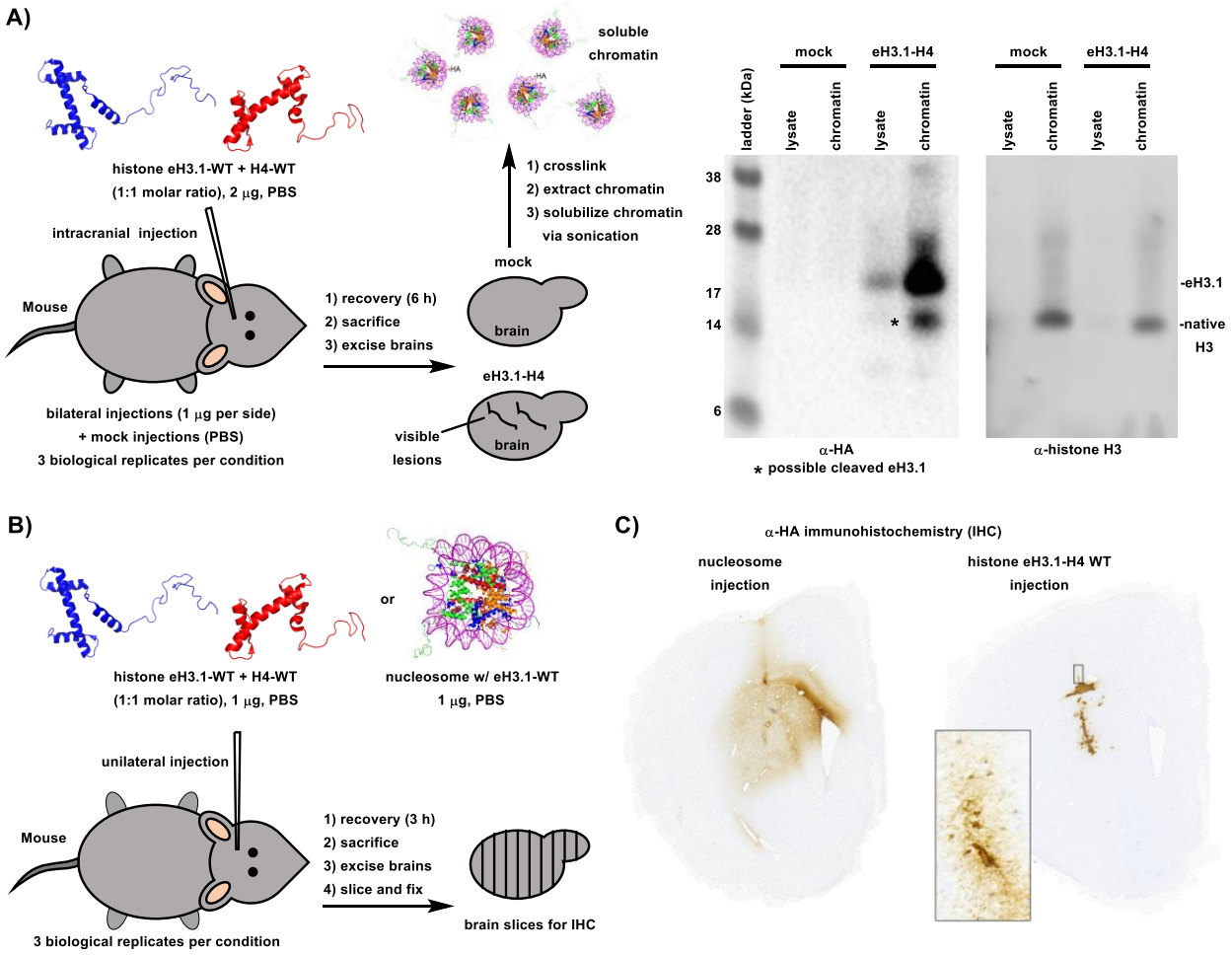


Figure 57: A) Experimental design for mice brain injection with histones eH3.1 and H4 WT, followed by chromatin extraction and western blot confirming the presence of histone eH3.1 in the chromatin. **B)** Experimental design for mice brain injection with histones eH3.1 and H4 WT or nucleosomes, followed by brain excision, slicing, and fixing for immunohistochemistry **(C)** against histone eH3.1 (α-HA). Figure adapted with permission with data provided by Dr. Abi Yates.

Initial progress at developing Zebrafish as a model organism for *in vivo* light-driven chemical protein modification

Zebrafish as a model organism for light-driven protein modification

Zebrafish are a popular model organism for many fields in biology due to a multitude of advantageous qualities. First and foremost, zebrafish are ideal model organisms for the study of vertebrate developmental biology[318,319]. Eggs, once fertilized, rapidly develop, with easily observable cell division and developmental stages occurring within minutes or hours, producing a hatched zebrafish larva in only 3 days (**Figure 58**)[320]. Furthermore, the relatively large size of the embryo (~1 mm diameter)[320] has led to the development of robust microinject techniques, allowing facile genetic manipulation[321].

Zebrafish also benefit from another quality, the translucence of their embryonic and larval stages. This allows not only simple visualization of their embryonic development, but the use of light as a reporter or tool for characterizing *in vivo* function in real-time. Robust strategies exist for light-triggered gene silencing through the use of photoactivatable DNA oligos called morpholinos that inhibit RNA splicing and translation[322], fluorescent imaging of glycan-PTMs through bio-orthogonal fluorophore conjugation to incorporated functionalized glycan derivatives[323], and optical control of cell signaling via photo-uncaging protected protein sidechains incorporated via genetic code expansion.

We envisioned that our light-driven radical Dha modification strategies[208], either developed in **Chapter 1** or by colleagues (Dr. Patrick Isenegger and Adeline Poh), could be utilized to install desired PTMs into proteins of interest in zebrafish, by microinjecting the necessary reaction components followed by illumination of the translucent zebrafish

embryos or larvae with blue light. If realized, this would provide a powerful tool for the site-specific and spatio-temporally controlled modification of proteins of interest inside living cells. This section explores this possibility, by testing the toxicity of the various reaction reagents on zebrafish embryo and larva stages, and establishing a platform for Dha containing histone microinjection, reaction, and retrieval that would allow for the desired modified histone product to be detected.

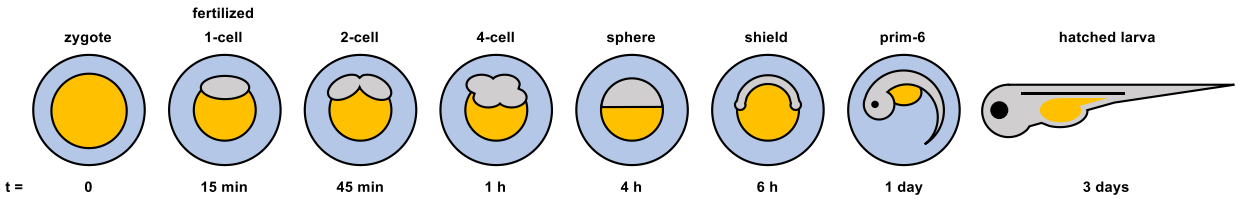


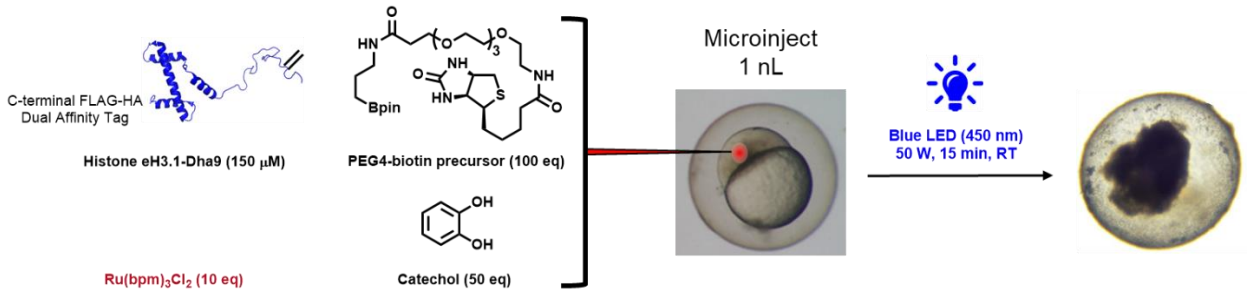
Figure 58: Portrayal of zebrafish embryonic developmental stages.

Initial attempts at *in vivo* protein modification and toxicity screens

We first attempted to modify microinjected histones with the light-driven radical Dha modification chemistry developed in **Chapter 1**. We chose to install a biotin containing sidechain onto histone eH3.1-Dha9, as the product of the reaction, a biotinylated FLAG-HA epitope tagged histone, would be possible to identify with respective antibodies in western blots of zebrafish protein extracts.

First, a reaction mixture was prepared containing histone eH3.1-Dha9 and WT histone H4 (150 µM each), a PEG4-biotin-boronic acid pinacol ester sidechain radical precursor (100 eq), catechol (50 eq) and $Ru(bpm)_3Cl_2$ (10 eq) in HEPES buffer (50 mM, pH 7.5), then KCl was added (from a 3 M stock, 300 mM final concentration). Then, newly fertilized zebrafish embryos (single cell stage, see **Figure 58, second from left**) were microinjected with the reaction mixture (1 nL per embryo) and exposed to blue light in our photobox (450 nm, 15 min, RT, 50 W) (**Figure 59A**). Unfortunately, shortly after light exposure, the embryos experienced rapid cell death evident by extensive protein aggregation. Efforts to solubilize the proteins and identify potential biotinylated histone reaction products failed. This indicated that either the reaction components or conditions were not biocompatible.

A)



B)

	Control		Low [ ]		Med [ ]		High [ ]	
Blue LED (50 W, 15 min)	+	-	+	-	+	-	+	-
$Ru(bpm)_3Cl_2$ (μM)	0		5		50		500	
Catechol (mM)	0		0.5		5		50	
R⌒Bpin (mM)	0		0.5		5		50	
Survival (2 h)	10/10	10/10	10/10	10/10	0/10	1/10	0/10	0/10

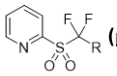
	Control		Low [ ]		Med [ ]		High [ ]	
Blue LED (50 W, 15 min)	+	-	+	-	+	-	+	-
$Ru(bpy)_3Cl_2$ (μM)	0		5		50		500	
$FeSO_4$ (μM)	0		50		500		5000	
 (μM)	0		5		50		500	
Survival (2 h)	10/10	10/10	10/10	10/10	10/10	10/10	1/10	3/10



Figure 59: A) Initial attempts at *in vivo* modification of histone eH3.1-Dha with light-driven radical Dha modification reagents and conditions microinjected into the first cell of a zebrafish embryo led to immediate death. **B)** Toxicity screens on zebrafish larvae (3 dpf) with reagents for two light-driven Dha modification reactions show Dha difluoroalkylation conditions are biocompatible at realistic reaction concentrations. Photo of the toxicity screen set up (right) showing live larva for Med [ ] (bottom left).

To investigate the cause of the rapid toxicity seen in our initial zebrafish microinjection attempts, we performed a toxicity screen with different mixture concentrations of the reaction components (**Figure 59B, top**). Furthermore, we tested reaction components for the light-driven radical Dha difluoroalkylation reaction, reasoning that the lower concentrations needed for that reaction may prove to be more biocompatible (**Figure 59B, bottom**). Zebrafish larvae, 3 days post-fertilization (3 dpf), were charged into glass vials (1 mL of E3 media, n = 10 per vial) and stock solutions containing mixtures of photocatalyst, additive, or sidechain radical precursor (4-bromobutylboronic acid or difluoromethyl 2-pyridyl sulfone, **Figure 59B**) were added. The vials were exposed to blue light while control vials were not exposed. The zebrafish were allowed to recover for 1 h, then live and active zebrafish larvae were counted.

The results revealed that the reagent concentrations required for the light-driven radical Dha modification from **Chapter 1** were lethal for zebrafish larvae (3 dpf) (**Figure 59B, top**). However, the light-driven radical Dha difluoroalkylation reaction was tolerated by zebrafish larvae, with all ten larvae surviving reagent concentrations more than sufficient for the reaction to occur (**Figure 59B, bottom and pictured right**). We decided to focus on light-driven radical Dha difluoroalkylation reactions from then on.

During these initial toxicity screens, an improvement on the light-driven radical Dha difluoroalkylation reaction was developed by colleague Adeline Poh (summarized in **Figure 55A**). This version made use of a benzothiazole sulfone reagent (called BtSOOF here) and required lower reagent concentrations and less time for successful reactivity. We therefore decided to conduct a more extensive zebrafish toxicity screen, testing each individual BtSOOF reaction component, as well as full reaction mixtures, on zebrafish

larvae followed by two days of survival monitoring (**Figure 60**). Zebrafish larvae (3 dpf, n = 5 per condition) were individually added to wells (200 µL of E3 media, 64 well plate), and exposed to two different concentrations of each reagent, plus a full reaction mixture (both with and without added iron sulfate), with or without blue light illumination. Then, zebrafish larvae were incubated in the reaction components for two days, frequently monitored for their survival by observing for activity.

The toxicity screen results showed the general biocompatibility of blue light, the $Ru(bpy)_3Cl_2$ photocatalyst, the BtSOOF-biotin reagent, and the additive $FeSO_4$ reaction components incubated individually with zebrafish larvae, with most or all larvae surviving the full two days (**Figure 60**). Pleasingly, the full reaction mixture minus $FeSO_4$ showed excellent biocompatibility, even at the higher tested concentrations and in the presence of blue light, with most zebrafish larvae surviving the full two days. Surprisingly, adding $FeSO_4$ to the full reaction mixture caused dramatic toxicity, accelerated with blue light illumination, even though $FeSO_4$ alone did not cause such effects. The role of $FeSO_4$ in this reaction, while not essential, is to speed up the reaction by initially reducing the $Ru(bpy)_3Cl_2$ photocatalyst to its required Ru(I) state. We reasoned that rapid accumulation of Ru(I), or the corresponding Fe(III), may be toxic, explaining the lethality seen when adding $FeSO_4$ to the full reaction mixture, although this is purely speculation at this point. Luckily, the reaction does not require the $FeSO_4$ additive, and the reaction mixture minus $FeSO_4$ was at sufficiently high concentrations to conduct the reaction successfully. Therefore, we decided to proceed with a new round of zebrafish microinjections, using the tolerated reaction concentrations minus $FeSO_4$ for installation of the BtSOOF-biotin reagent onto histone eH3.1-Dha27.

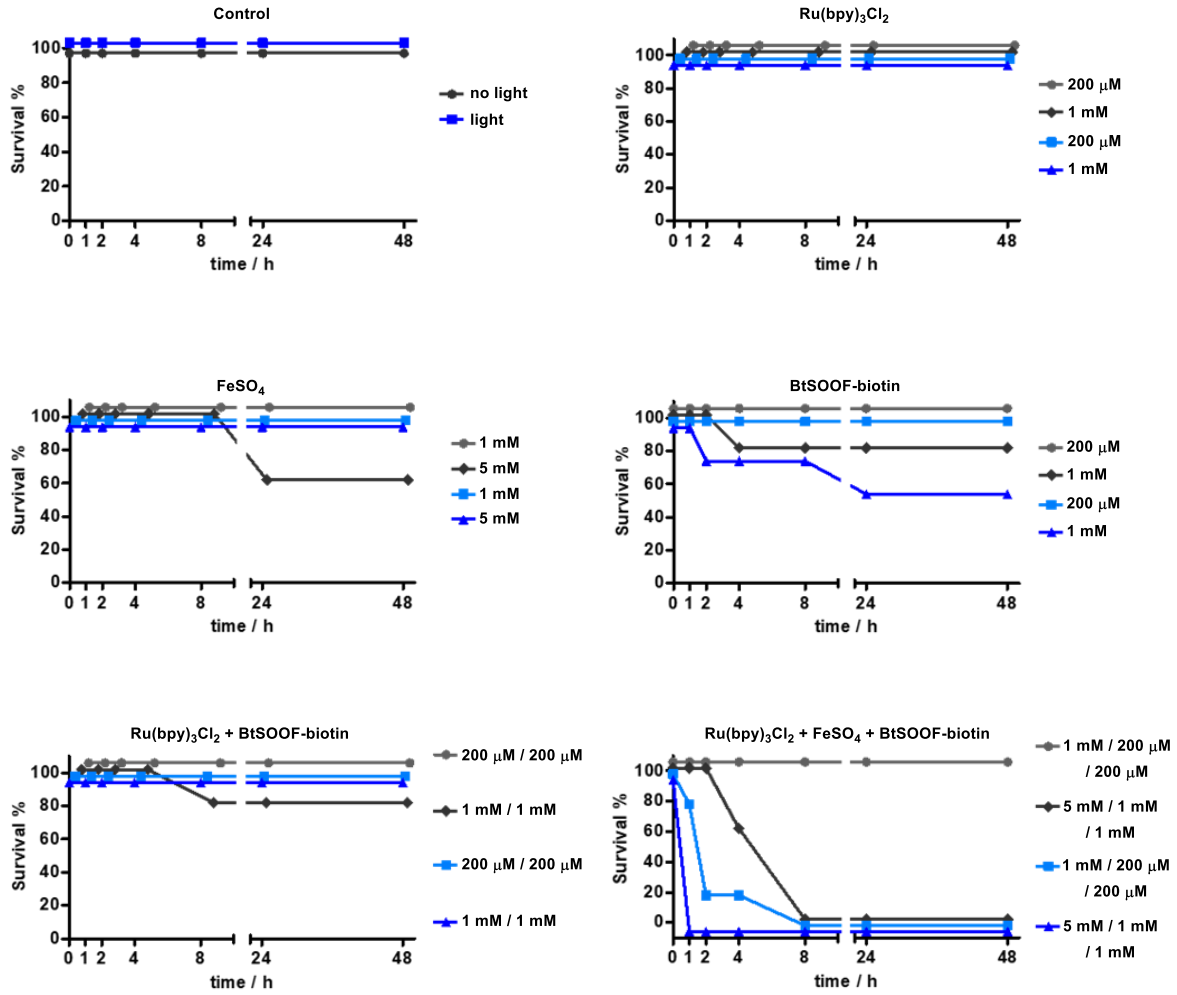


Figure 60: Time-course toxicity screen of reagents for the Dha difluoroalkylation reaction with the biotin benzothiazole sulfone reagent, carried out on zebrafish larva (3 dpf, n = 5). Blue indicates illumination of mixture with Blue LED (450 nm, 50 W, 15 min), darker lines indicate the higher tested concentration.

Development of a platform for histone microinjection, retrieval, and detection in zebrafish larva

With the toxicity screening results in hand, we decided to proceed with a new round of zebrafish microinjections, using the tolerated reaction concentrations minus $FeSO_4$ for installation of the BtSOOF-biotin reagent onto histone eH3.1-Dha27 (**Figure 61**). A reaction mixture of histone eH3.1-Dha27 (200 µM, with equimolar histone H4 WT for solubility), the BtSOOF-biotin reagent (1 eq), and the $Ru(bpy)_3Cl_2$ photocatalyst (1 eq) was prepared in HEPES buffer (50 mM, pH 7.5). Zebrafish larvae (3 dpf, n = 25 per condition) were microinjected (2 nL) at the base of the head and immediately exposed to blue light (450 nm, 50 W, 5 min, RT). Control conditions either were not exposed to light, did not contain the histone eH3.1-Dha27 protein, or were not injected. Most zebrafish larvae (20/25 per condition) were then deyolked, and total protein was extracted for western blots. A subset of the zebrafish (5/25 per condition) were placed in fresh E3 media and monitored for survival over two days.

Western blots of the zebrafish successfully detected the intact microinjected target histone eH3.1 protein via its HA epitope tag (**Figure 61, right**). However, no bands were present at the desired size in the α-biotin western blot (**Figure 61, right**), with higher molecular weight bands in all conditions likely due to endogenous biotinylated proteins in zebrafish[324]. Western blots against purified histone eH3.1-biotin27 (synthesized by Adeline Poh) showed low sensitivity of several antibodies to our biotinylated histone eH3.1 (data not shown). On a positive note, all the microinjected zebrafish, in every condition, survived for a further two days after microinjection, indicating that even though the reaction was not successful, the conditions and reagents were biocompatible.

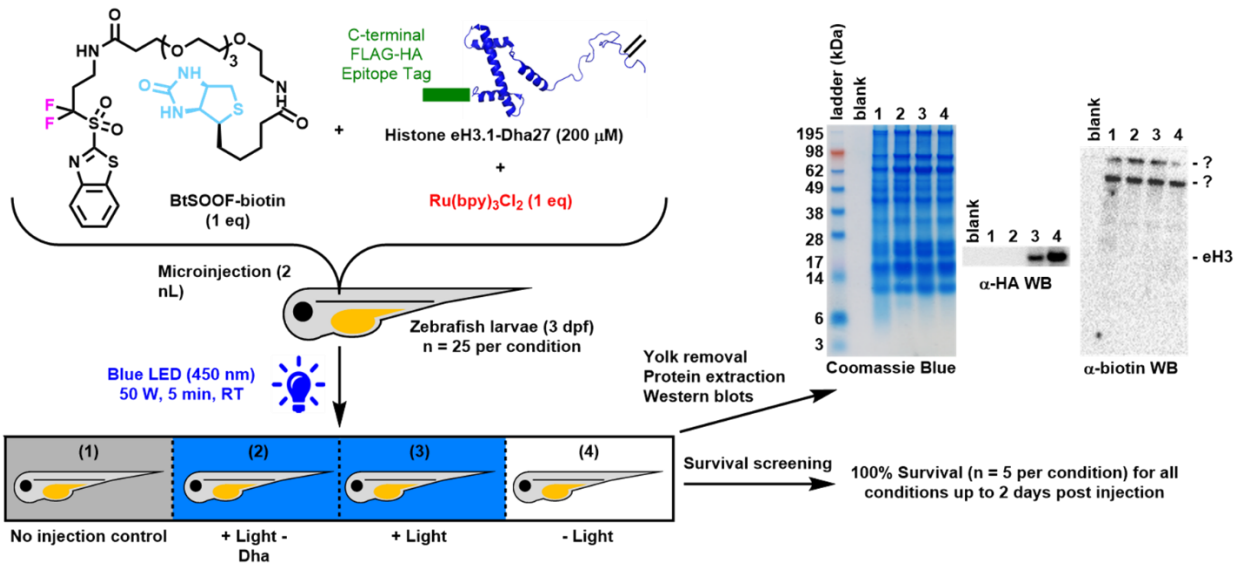


Figure 61: Experimental design for zebrafish larva (3 dpf) microinjection with reagents for Dha difluoroalkylation with BtSOOF-biotin, followed by illumination, protein extraction, and western blots against the injected histone eH3.1 target (α-HA) and the desired modification (α-biotin) revealing no desired product bands, but high survival rates.

We therefore decided to redo the microinjections with a different sidechain precursor, and further purify the reaction product via an α-HA IP. However, before conducting another round of zebrafish microinjections, we first decided to test our ability to detect the installed sidechain on the target Dha containing histone. We settled on installing histone eH3.1-γCF-Me_2Lys4, as the corresponding antibody easily detected the reaction product by western blot without the use of added $FeSO_4$, and importantly, gave no band for the histone eH3.1-Dha4 starting material (**Figure 62A**).

Encouraged by these results, we repeated microinjection experiments. A reaction mixture of histone eH3.1-Dha4 (200 μM, with equimolar histone H4 WT for solubility), the BtSOOF-γCF-Me_2Lys reagent (5 eq), and the $Ru(bpy)_3Cl_2$ photocatalyst (1 eq) was

prepared in HEPES buffer (50 mM, pH 7.5). Zebrafish larvae (4 dpf, n = 25 per condition) were microinjected (2-6 nL) at the base of the head and immediately exposed to blue light (450 nm, 50 W, 5 min, RT). Control conditions either were not exposed to light, did not contain the histone eH3.1-Dha27 protein, or were not injected. In addition, a condition was added that contained $FeSO_4$ (1 eq), to test if added iron could accelerate the reaction while being tolerated by the zebrafish. Most zebrafish larvae (20/25 per condition) were then deyolked, and total protein was extracted, then further purified via an α-HA IP for western blots. A subset of the zebrafish (5/25 per condition) was again placed in fresh E3 media and monitored for survival over two days (**Figure 62B**).

Disappointingly, the results mirrored our previous attempt (**Figure 61**), with no desired bands in the histone H3-Me_2Lys4 western blot outside of that of native histone H3 (**Figure 62B**). However, histone eH3.1 was easily detected via its HA epitope tag in all conditions where it was microinjected, showing that the injected histone variant could be easily extracted, purified by IP, and detected by western blot after zebrafish microinjection. This, couple with the 100% survival rate after two days exhibited both with and without added $FeSO_4$ gave us hope that future attempts could prove successful.

Currently, it is unknown why the reactions are not working in the zebrafish, with several possible explanations. Various reaction components, including the histone, may be rapidly sequestered from each other, inhibiting reactivity. Alternatively, biogenic compounds, such as glutathione[229], may be adding to Dha prior to the reaction starting, or potentially poisoning other components. Regardless, future efforts (to be undertaken by Adeline Poh) will investigate alternative installed sidechains, Dha containing proteins, and conditions.

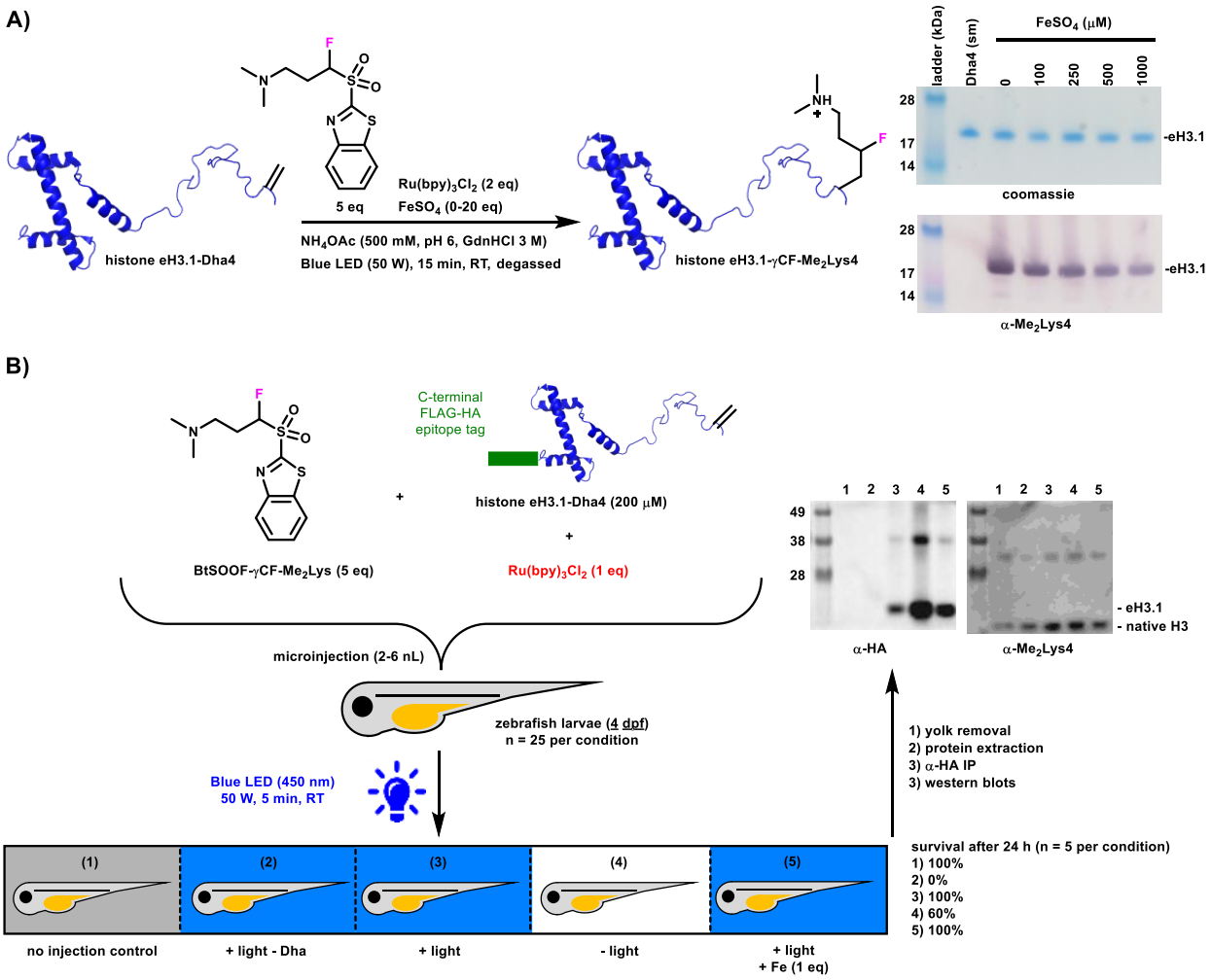


Figure 62: A) Test reaction installing the histone eH3.1-γCF-Me2Lys4 PTM, successfully identified by western blot (right). **B)** Experimental design for zebrafish larva (4 dpf) microinjection with reagents for installation of the histone eH3.1-γCF-Me2Lys4 PTM, followed by illumination, protein extraction, α-HA IP, and western blots against the injected histone eH3.1 target (α-HA) and the desired modification (α-Me2Lys4) revealing no desired product bands, but high survival rates.

Conclusion

Summary and advantages brought about by this chapter

This chapter began with investigating the incorporation of WT and PTM-containing histones into HeLa cells. Histones were tracked using the epitope tagged histone eH3.1 variant (C-terminal FLAG-HA tagged), allowing for its incorporation rate and localization to be characterized. Furthermore, a variety of PTM-containing histones were incorporated into HeLa cells, followed by RNA-seq, to quantify the transcriptional response of the cells to extracellular histone incorporation. Lastly, incorporated histones were successfully identified in the purified chromatin of HeLa cells, which will be used in future ChIP-seq experiments to identify the genomic localization of incorporated histones.

Next, WT or ^{18}F-radiolabeled histones were injected into the tail veins of mice, with the goal of tracking the histones through either their identification in the chromatin of various organs (via the FLAG-HA tag on WT histone eH3.1), or via positron emission tomography (PET) imaging and tracking of the ^{18}F-radiolabeled histones, giving insights into the stability of free histones vs intact nucleosomes in the blood. Then, to simulate free histone or nucleosome release during injury[141], WT histones or nucleosomes (containing histone eH3.1) were injected into mice brains, followed by IHC and chromatin extraction, to distinguish their different diffusion, incorporation, and localization qualities. Chromatin samples taken from mice brains showed the presence of incorporated histone eH3.1 and will be used in the future for ChIP-seq experiments, paired with transcriptomic analysis of the injected mice brain samples.

Lastly, Dha-tagged histones were injected into zebrafish embryos and larvae and used as targets for various light-driven radical Dha modification strategies[208]. These

experiments were not done with the goal of investigating the role of injected histones into zebrafish, but were undertaken instead to assess the biocompatibility, bio-orthogonality, and ability of various histone modification chemical techniques to modify injected Dha-containing histones *in vivo*, leveraging the advantageous translucent qualities of zebrafish needed for light-driven chemistry. While attempts were unsuccessful, a platform for histone injection, modification, and retrieval for analysis from zebrafish was established, and toxicity screens of various Dha modification reaction conditions and components showed promise for future endeavors.

Prospects for future work

The ability of histones to penetrate cell membranes and incorporate into chromatin is an intriguing biological phenomenon that has proved to be much too big for a single PhD thesis to thoroughly explore. The results shown here, quantifying HeLa cell transcriptional response to incorporated histones, characterizing free histone and nucleosome stability in the circulatory system, and identifying incorporated histones in the chromatin of intracranially injected mice, only begin to chip away at characterizing this unexplored ability. Immediate future work will identify the genomic locations of incorporated extracellular histones through ChIP-seq experiments on chromatin already collected from experiments described above. Further work should investigate effects of specific PTMs on the ability of histones to penetrate cells, and what their effects are on the chromatin of the cells they enter.

In addition, the ability to incorporate modified histones into the chromatin of living cells opens an avenue for studying specific histone PTMs in native chromatin context. There are a wide range of chemical methodologies for introducing desired PTMs into histones[94,95], but most of them must happen *in vitro*, require substantial engineering of the cellular system of interest, or are limited in the sidechains they can introduce into histones. Histones can now be synthesized *in vitro* with desired PTMs, then incorporated into the chromatin of cells of interest via simple incubation, allowing their effects in the chromatin to be studied.

Furthermore, there are a variety of diseases and cancers promoted by the accumulation of mutated histones or dysregulated histone PTMs[50,295,296]. While currently ambitious, one could envision treatments where necessary histone can be used as a therapeutic to flood

the organ or tissue of interest, promoting its incorporation in the chromatin and restoring native epigenetic function. Alternatively, the understudied role of free histones in the blood, and strong immune response against their presence, hints at potential roles of extracellular histones released from diseased or damaged cells incorporating into the surrounding healthy tissue. This could be a yet uncharacterized mechanism of disease phenotypic spreading, or a potential circulatory communication pathway characterized by the profile of released PTM containing histones.

Intriguingly, some human associated pathogens (with no histones or chromatin) have been found to express SET containing proteins with high similarity to Eukaryotic histone methyltransferases. This was puzzling as there were no described mechanisms for those enzymes to be inserted into host human cells[325,326]. It is possible that bacterial pathogens express histone modifying enzymes that modify histones in the circulatory system, which when incorporated into human cells, affect gene expression advantageously for the pathogens.

Lastly, while recent attempts at modifying Dha containing histones microinjected into zebrafish were unsuccessful, they remain promising targets for applying our various light-driven Dha modification techniques. Zebrafish were surprisingly easy to work with, and easily tolerated the reaction components and conditions, with high rates of survival after being injected with reagent concentrations that should allow for reactivity. The platform established in this thesis work should allow for future work to easily attempt many variations of our reactions in zebrafish, with a robust method for retrieving and enriching the microinjected target Dha containing histones.

Statement on collaboration

The work outlined in this thesis chapter was done with the assistance of many others, who assisted with various aspects throughout the course of the work. Their contributions are explicitly stated in the text when appropriate and outlined in full here.

RNA-seq experiments were carried out in the lab of Professor Udo Oppermann, under the direction of Dr. Martin Philpott. Dr. Adam Cribbs carried out processing and analysis of the RNA-seq data.

All rodent studies with injected histones and nucleosomes were performed by Dr. Abi Yates in the group of Professor Daniel Anthony. Adeline Poh developed the histone ^{18}F-radiolabelling chemical method, and was a partner throughout all injection experiments, helping make required materials and carrying out western blots and other characterization work. Alice Bedwell provided the nucleosome samples used for injections. Sarah Able and Katherine Vallis carried out PET imaging experiments. Adeline Poh is carrying out future work on this project.

We collaborated with the group of Professor Mathilda Mommersteeg for zebrafish experiments, who supplied zebrafish embryos, and provided equipment and training on how to work with zebrafish larvae, with special thanks to William Stockdale. Adeline Poh was again a constant research partner throughout the later stages of the zebrafish work and is carrying on with future attempts to modify proteins *in vivo*.

Experimental Details

General Methods

Details for general experimental procedures, intact protein LC-MS, production (expression, purification, and Dha formation) of histone proteins, α-HA IPs and western blotting are given in the **Experimental Details** section of **Chapter 1**.

Cell culture

Adherent WT HeLa cells (Acc No: 93021013, Lot: 17A001, P: +5) were grown as monolayers in T-75 flasks filled with 10 mL of Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), and L-glutamine (2 mM).

HeLa cell histone incorporation protocol

WT HeLa cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS), 2 mM L-glutamine in a 5% CO_2 37 °C incubator until they reached 60-80% confluency in T-75 flasks. They were washed with PBS, then incubated with a 1:1 molar ratio of WT histone H4 and the respective histone eH3.1 sample, each at 1 μM, for 15 min in PBS in a 5% CO_2 37 °C incubator. The cells were then washed three times with PBS and allowed to recover for variable times (experiment dependent) in normal media and conditions (see above) before being washed twice with PBS and harvested with a cell scraper. Trypan blue was used to check cell viability and the cells were centrifuged in PBS for 5 min 800 g before being placed on ice and proceeding to cell fractionation.

HeLa cell fine fractionation protocol

All steps were carried out either on ice or at 4 °C and all fractions were stored at -80 °C. Approximately 1 x 10^6 HeLa cells were resuspended in Hypotonic Buffer (500 µL of 10 mM Tris base, 15 mM sodium chloride, 1.5 mM magnesium chloride, pH 7.6, with Protease Inhibitor Cocktail Roche, cat. no. 11873580001) and allowed to lyse for 15 min. Nuclei were pelleted by centrifugation at 400 g for 5 min and the supernatant, designated the cytoplasm fraction, was removed. Nuclei were resuspended again in Hypotonic Buffer (500 µL) and subjected to 10 strokes with a loose pestle Dounce homogenizer to lyse residual intact cells and cytoplasmic organelles before repeating the above centrifugation and removing the supernatant. The nuclei pellet was then resuspended in Hypotonic Buffer containing 1% Triton X-100 (500 µL) and lysed with 10 strokes from a tight pestle Dounce homogenizer. The nucleoplasm fraction was separated from the chromatin fraction by pelleting the latter with centrifugation at 10,000 g for 10 min, and the chromatin pellet was washed once with Hypotonic Buffer containing 1% Triton X-100 (500 µL) before being resuspended in RIPA-SDS Lysis Buffer (500 µL RIPA buffer + 1% SDS) with the help of a 5 s sonication at 35% amplitude.

Double thymidine block protocol

HeLa WT cells were cultured (DMEM + 10% FBS + 2 mM GlutaMax) and divided into 12 well plates, with 4 wells for each of the 12 sample conditions in the previous page (48 total). Once cells were at roughly 40% confluency, thymidine blocking solution was added (PBS stock, final concentration 2 mM per 1 mL media) and the cells were incubated for 16 h. Cells were then washed 2x with PBS, fresh media was added, and cells were allowed to grow for 9 h. The thymidine blocking was repeated, and cells were incubated for a further 14 h. Media was removed, and cells were washed 2x with PBS. This treatment synchronizes all cells to G1/S phase, to be released into S phase where the genome is replicated, allowing the best chance of externally added histones to be incorporated into the new chromatin.

HeLa cell histone incorporation protocol for RNA-seq

Immediately after the last wash of the double thymidine block procedure above, cells were incubated in PBS containing an equimolar mixture of modified Histones eH3.1 and WT H4 (1 μM each) for 15 min at 37 °C. After the histone treatment, the cells were washed 3x with PBS and incubated with normal media for 6 h. Media was then removed, and 350 μL of Trizol reagent was added to lyse cells and stabilize RNA for future purification and sequencing.

Total RNA and protein purification from HeLa cells for RNA-seq

Total RNA was removed from the Trizol reagent and purified using the Direct-zol Miniprep Plus Kit (Zymo Research) following the manufacturer's instructions. After an equal volume of ethanol was added to the Trizol reagent mixture and placed on the Zymo-Spin IIICG Column, the flow through containing the total protein extract was collected and stored at -20 °C for later use. Protein was later precipitated by adding 4 volumes of ice-cold acetone and incubating for 30 min. The precipitant was pelleted at 13,000 rpm, 5 min, 4 °C, washed with 400 µL of 95% ethanol, and air dried for 5 min. The pellet was then dissolved in 0.5 mL Sample Buffer (NuPAGE LDS Sample buffer containing 10% βME) by shaking at 37 °C for 30 min. It was then stored at -20 °C for downstream applications such as SDS-PAGE and western blot analysis.

mRNA enrichment and library preparation

The mRNA was enriched from the total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490) following manufacturer's instructions. Libraries were prepared with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB #E7760). Libraries were then multiplexed, quantified using a High-sensitivity d1000 TapeStation (Agilent), then sequenced using a NextSeq 500 (Illumina). Target sequencing depth was >20 million reads per sample library.

HeLa cell histone incorporation protocol for ChIP-seq

Immediately after the last wash of the double thymidine block procedure, WT HeLa cells (50% confluent in 150 mm dishes, 3 biological replicates per condition) were incubated in PBS containing an equimolar mixture of modified Histones eH3.1 and WT H4 (1 μM each) for 15 min at 37 °C. After the histone treatment, the cells were washed 3x with PBS and incubated with normal media for 36 h. Media was then removed, and cells were harvest by physical detachment (cell scraping), immediately preceding chromatin extraction.

Chromatin extraction and shearing (from HeLa cells and mice tissue samples)

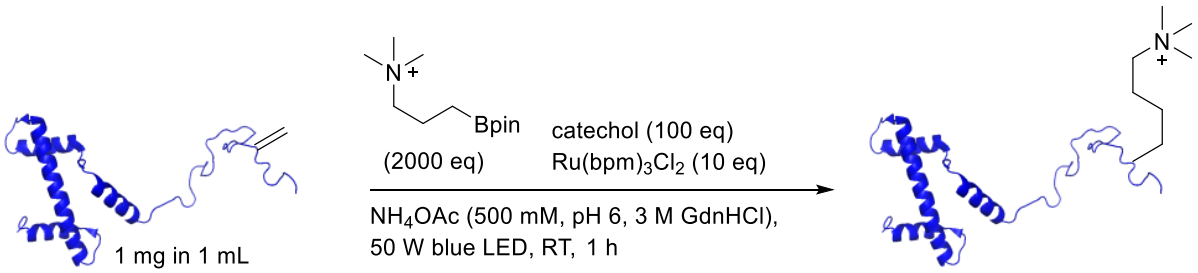
Chromatin was extracted according to the Chromatin Extraction Kit-Flexible Format (Abcam, ab223876). Briefly, HeLa cells were collected by scraping, washed with PBS, crosslinked with 1% formaldehyde (15 min in PBS), and quenched with the kit provided formaldehyde quenching solution (5 min in PBS) before 3x PBS washes. Mice tissue samples were diced with a scalpel, washed twice with ice-cold PBS, crosslinked in 1% formaldehyde (15 min in PBS) with frequent vortexing, and quenched with the kit provided formaldehyde quenching solution (5 min in PBS) before 3x PBS washes. Crosslinked HeLa cells or mice tissue were resuspended in 500 μL of Extraction and Lysis Buffer supplemented with protease inhibitors, then treated with 10 passes of both the loose and tight Dounce homogenizer settings, before being incubated for 30 min at RT with 1000 rpm shaking. Chromatin pellets spun down (10,000 rpm, 15 min) and washed once with milliQ water. Chromatin was then sheared by sonication with a Bioruptor Pico (30s on/30s off, 15 min for cells, 30 min for tissue), insoluble material was spun down (10,000 g, 10 min) and soluble sheared chromatin was stored at -80 °C.

Zebrafish microinjections

Zebrafish embryos and larvae were grown in petri dishes containing E3 media, incubated at 28 °C. Microinjections were performed with a PicoSpritzer III microinjector, either into the first cell of the embryo, or at the base of the head of 3-4 dpf larvae. The injection mixtures were prepared by premixing equimolar amounts of histone eH3.1-Dha and WT histone H4 in HEPES buffer (50 mM, pH 7.5), then adding 1% v/v of 100x stocks of all reagents (prepared in HEPES buffer) to give the desired final concentration. The pH was checked after the addition of all reagents, and the reaction mixture was sterile filtered prior to zebrafish microinjections. After the injections, the embryos or larvae were deyolked by rapid pipetting with a 200 µL pipet tip. The remaining tissue, separated by centrifuging (500 g, 5 min) was lysed with a tissue homogenizer in 5 M Urea, shaken at 600 rpm 37 °C for 1 h, and centrifuged (10 min 15k rpm) to remove debris. The total protein extracts were often further purified by α-HA immunoprecipitations (described in **Chapter 2**) but were first diluted 40x in the IP buffer to diminish urea concentrations.

Protein modification details

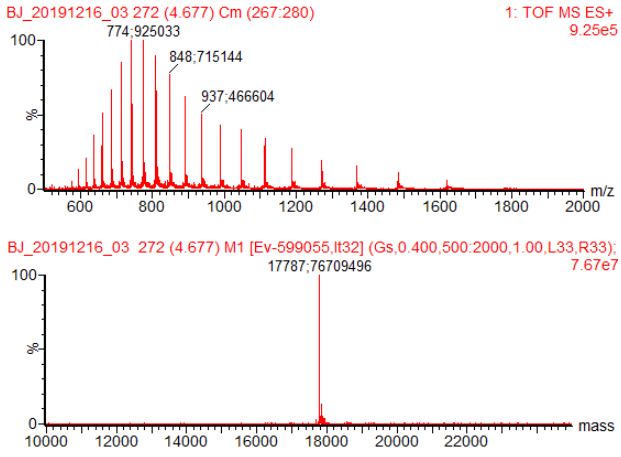
Installation of trimethyllysine (Me_3Lys) onto histone eH3.1-Dha4/9/18/27



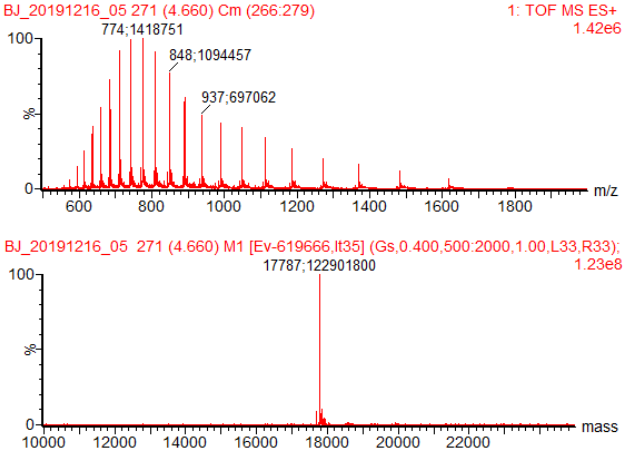
histone eH3.1-Dha4/9/18/27 **>90% conversion** histone eH3.1-Me_3Lys4/9/18/27

Calculated mass: 17786 Da. Observed mass: 17787 Da

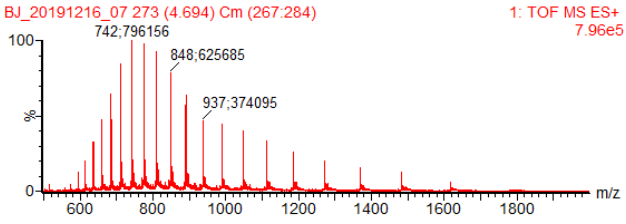
histone eH3.1-Me_3Lys4

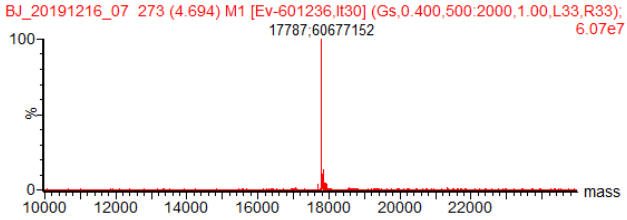


histone eH3.1-Me_3Lys9

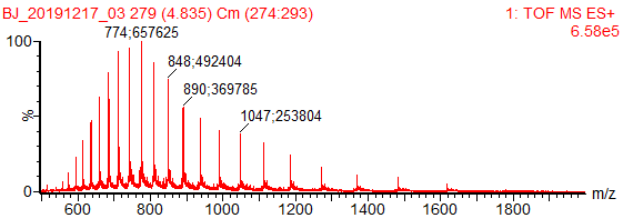


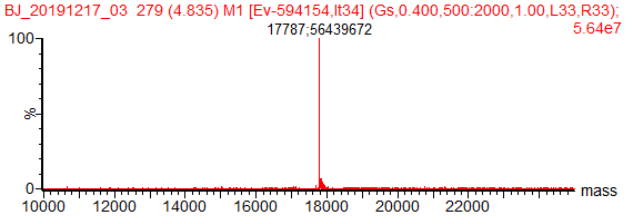
histone eH3.1-Me_3Lys18





histone eH3.1-Me_3Lys27





Concluding Remarks

“First step, put samples on the mass spec, next step, Nobel Prize”

- Dr. Patrick Isenegger, on a Saturday morning

Outlook and lessons learned

During my time in the Professor Ben Davis group, I have been a part of either developing or applying many new Dha modification chemical methods, always with the goal of learning more about histone PTMs and epigenetics. This was accomplished either by installing sidechains of interest directly into histones for study (such as AcLys, BzLys, or Me_3Lys PTMs), or utilizing unnatural or reactive sidechains (such as alkyl halide containing Bhn for crosslinking, γCF_2 containing sidechains for ^{19}F-NMR, or Bal for DICE NMR), to enable the study of epigenetic processes in new ways. Throughout the process of seeing this small piece of the field of protein modification grow (**Figure 63**), several lessons became apparent about how to conduct research as a chemical biologist.

I gained a new appreciation for the need to constantly be developing and testing the limits of new protein modification methods. Coming into the group, my background was in biochemistry and genetics, and I reasoned that the biggest impact I would have would be through using known and established protein modification methods to investigate histone PTMs through careful planning and application. However, in practice, most of the applications described in this thesis were not planned from the outright, but often serendipitously discovered as we explored the capabilities and limitations of new protein modification techniques.

For example, the light-driven radical Dha modification reaction (**Chapter 1**) was never planned to allow for the installation of electrophilic sidechains, but during substrate scope investigations, this capability was revealed, allowing for the follow-up application of proximity-driven protein-protein crosslinking between Bhn-containing histones and the histone modifying protein KDM4A (**Chapter 2**).

In another example, the Dha difluoroalkylation reaction, developed by Dr. Patrick Isenegger, utilized the fluorine atoms initially as a component for promoting more efficient radical initiation. It was only during a discussion with Patrick at a pub on a Friday evening that we realized the γCF_2 group could be used to track PTM processing further up the sidechain, initiating a frenzied trip back to the lab, and resulting in conclusive ^{19}F-NMR results of the deacetylation of histone H3-γCF_2-AcLys18 before the end of the weekend. This was a testament to how fit for purpose the Dha difluoroalkylation reaction was for this exact application.

Over time I realized that this philosophy of “following the chemistry” was a cornerstone of the success of our group. At times this was frustrating, as even though a project may have seemed complete and ready to publish, Professor Davis would insist on more and more applications, each more ambitious than the last. I realize now he was “following the chemistry,” and would only see a project as complete once it finally ran up against the wall that was the true limitations of the method. This is an important consideration in the field of protein modification and chemical biology, as new methods will either not be employed, trusted, or read by the biologists that we hope will use our methods, unless they are tested in complex biological settings and shown to tackle relevant biological problems. Too often, protein modification publications skip this step, and either show simple biological applications of their method that do not offer any improvement over other methods, or do not attempt any biological applications at all.

Lastly, this philosophy of “following the chemistry” works with biology as well. In chemical biology, we are often thinking as engineers, and I have often fallen into the trap of designing a biological experiment and then getting frustrated when it leads to an

unexpected outcome. My previous boss, Professor Michael Freitag, an epigenetics and microbiology expert at Oregon State University, gave me some great advice on this. He said, “if you are investigating an important biological question, and you design a good experiment, then even if it fails, it will tell you something important.”

This certainly was the case when I started researching the ability of histones to penetrate cells and incorporate into chromatin. I was initially thinking like an engineer, with a desired outcome that modified histones could be injected into the blood and incorporate into chromatin in cells around the body, a potential route to both study histone PTMs *in vivo*, and a treatment for various histone PTM dysregulation diseases. The data gathered in this thesis showed that histones were rapidly degraded in the blood, and when they did penetrate cells, they caused a strong immunological response. While initially disappointed, I soon realized that I was trying to engineer a system that I did not fully understand, and that the results, which I had viewed as a failure, were telling me something much more important. I shifted my focus to characterizing the effects on extracellular histones and intact nucleosomes on cells and tissues, work which is still ongoing, and has revealed a robust and fascinating system where the degradation of free histones in the circulatory system and the strong damage response of cells to histone incorporation together hints at a defense mechanism against the incorporation of extracellular histones into chromatin.

To conclude, while it is important to have plans and desired outcomes, real breakthroughs, success, and fulfillment in research comes from being curious. I am grateful for my time in this group, content with what I have accomplished, and thankful for to my colleagues and friends who have continually fostered and stimulated my curiosity.

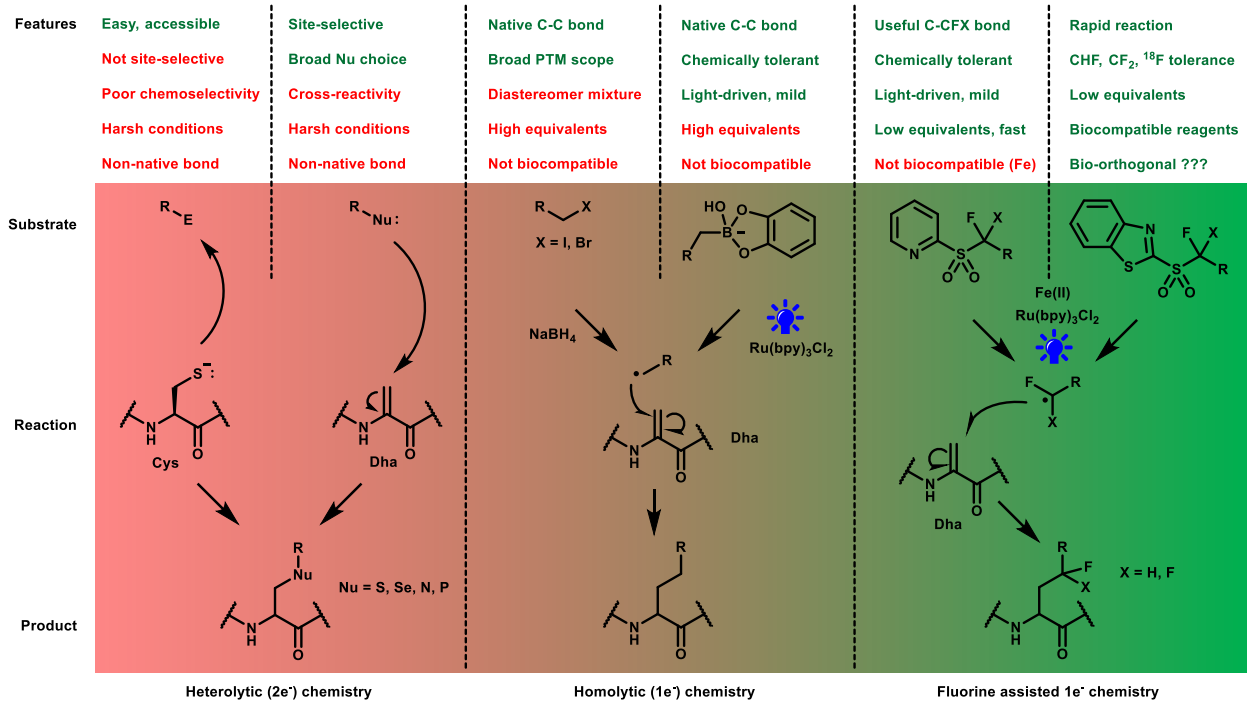


Figure 63: Progression of Dha based protein modification chemistry.

References

1 Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563, doi:10.1038/227561a0 (1970).

2 Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences* **83**, 6233, doi:10.1073/pnas.83.17.6233 (1986).

3 Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8, (2016).

4 Cavalli, A., Salvatella, X., Dobson, C. M. & Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences* **104**, 9615, (2007).

5 Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **159**, 995-1014, (2014).

6 Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78-85, (2016).

7 Miga, K. H. *et al.* Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**, 79-84, (2020).

8 International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945, (2004).

9 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, (2001).

10 Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304-1351, (2001).

11 Claverie, J.-M. What If There Are Only 30,000 Human Genes? *Science* **291**, 1255-1257, (2001).

12 Black, D. L. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry* **72**, 291-336, (2003).

13 Matlin, A. J., Clark, F. & Smith, C. W. J. Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology* **6**, 386-398, (2005).

14 Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693-705, (2007).

15 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45, (2000).

16 Driscoll, D. M. & Copeland, P. R. Mechanism and regulation of selenoprotein synthesis. *Annual Review of Nutrition* **23**, 17-40, (2003).

17 Krzycki, J. A. The direct genetic encoding of pyrrolysine. *Current Opinion in Microbiology* **8**, 706-712, (2005).

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

19 Beck-Sickinger, A. G. & Mörl, K. Posttranslational Modification of Proteins. Expanding Nature's Inventory. By Christopher T. Walsh. *Angewandte Chemie International Edition* **45**, 1020-1020, (2006).
20 Wang, Y.-C., Peterson, S. E. & Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Research* **24**, 143-160, (2014).
21 Jensen, O. N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* **8**, 33-41, (2004).
22 Ryan, B. J., Nissim, A. & Winyard, P. G. Oxidative post-translational modifications and their involvement in the pathogenesis of autoimmune diseases. *Redox Biol* **2**, 715-724, (2014).
23 Clarke, S. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. *International Journal of Peptide and Protein Research* **30**, 808-821, (1987).
24 Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med* **40**, 271-280, (2017).
25 Ubersax, J. A. & Ferrell Jr, J. E. Mechanisms of specificity in protein phosphorylation. *Nature Reviews Molecular Cell Biology* **8**, 530-541, (2007).
26 Ali, I., Conrad, R. J., Verdin, E. & Ott, M. Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics. *Chemical Reviews* **118**, 1216-1252, (2018).
27 Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature Reviews Molecular Cell Biology* **15**, 536-550, (2014).
28 Komaniecki, G. & Lin, H. Lysine Fatty Acylation: Regulatory Enzymes, Research Tools, and Biological Function. *Frontiers in Cell and Developmental Biology* **9**, (2021).
29 Luo, M. Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chemical Reviews* **118**, 6656-6705, (2018).
30 Guccione, E. & Richard, S. The regulation, functions and clinical relevance of arginine methylation. *Nature Reviews Molecular Cell Biology* **20**, 642-657, (2019).
31 Hastie, C. J., McLauchlan, H. J. & Cohen, P. Assay of protein kinases using radiolabeled ATP: a protocol. *Nat Protoc* **1**, 968-971, (2006).
32 Pande, S. V. & Caramancion, M. N. A simple radioisotopic assay of acetylcarnitine and acetyl-CoA at picomolar levels. *Anal Biochem* **112**, 30-38, (1981).
33 Feng, Y., Li, M., Wang, B. & Zheng, Y. G. Discovery and Mechanistic Study of a Class of Protein Arginine Methylation Inhibitors. *Journal of Medicinal Chemistry* **53**, 6028-6039, (2010).
34 Fra, A., Yoboue, E. D. & Sitia, R. Cysteines as Redox Molecular Switches and Targets of Disease. *Front Mol Neurosci* **10**, 167-167, (2017).
35 Kim, G., Weiss, S. J. & Levine, R. L. Methionine oxidation and reduction in proteins. *Biochim Biophys Acta* **1840**, 901-905, (2014).

36 Leahy, D. *et al.* Discrimination of Methionine Sulfoxide and Sulfone by Human Neutrophil Elastase. *Molecules* **26**, 5344, (2021).
37 Ahsan, H. 3-Nitrotyrosine: A biomarker of nitrogen free radical species modified proteins in systemic autoimmunogenic conditions. *Human Immunology* **74**, 1392-1399, (2013).
38 Tiwari, N. *et al.* The SrrAB two-component system regulates Staphylococcus aureus pathogenicity through redox sensitive cysteines. *Proc Natl Acad Sci U S A* **117**, 10989-10999, (2020).
39 Moskovitz, J. Methionine sulfoxide reductases: ubiquitous enzymes involved in antioxidant defense, protein regulation, and prevention of aging-associated diseases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1703**, 213-219, (2005).
40 Stephenson, R. C. & Clarke, S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J Biol Chem* **264**, 6164-6170 (1989).
41 Clarke, S. Aging as war between chemical and biochemical processes: Protein methylation and the recognition of age-damaged proteins for repair. *Ageing Research Reviews* **2**, 263-285, (2003).
42 McCudden, C. R. & Kraus, V. B. Biochemistry of amino acid racemization and clinical application to musculoskeletal disease. *Clin Biochem* **39**, 1112-1130, (2006).
43 Stabler, T. V., Byers, S. S., Zura, R. D. & Kraus, V. B. Amino acid racemization reveals differential protein turnover in osteoarthritic articular and meniscal cartilages. *Arthritis Res Ther* **11**, R34-R34, (2009).
44 Luisa Di Gioia, M. *et al.* N-Methylated Amino Acids And Peptides: Synthesis And Biological Activity. *Mini Reviews in Medicinal Chemistry* **16**, 683-690 (2016).
45 Song, H. & Naismith, J. H. Enzymatic methylation of the amide bond. *Current Opinion in Structural Biology* **65**, 79-88, (2020).
46 Varland, S., Osberg, C. & Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **15**, 2385-2401, (2015).
47 Nuijens, T. *et al.* Enzymatic C-terminal amidation of amino acids and peptides. *Tetrahedron Letters* **53**, 3777-3779, (2012).
48 Clarke, S. & Tamanoi, F. Fighting cancer by disrupting C-terminal methylation of signaling proteins. *J Clin Invest* **113**, 513-515, (2004).
49 Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes & Development* **23**, 781-783, (2009).
50 Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nature Biotechnology* **28**, 1057-1068, (2010).
51 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research* **21**, 381-395, (2011).
52 Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics* **14**, 100-112, (2013).
53 Bowman, G. D. & Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chemical reviews* **115**, 2274-2295, (2015).

54 Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293**, 1074-1080, (2001).
55 Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3-3, (2015).
56 Gilbert, N. *et al.* Formation of facultative heterochromatin in the absence of HP1. *EMBO J* **22**, 5540-5550, (2003).
57 Trojer, P. & Reinberg, D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* **28**, 1-13, (2007).
58 Gräff, J. & Tsai, L.-H. Histone acetylation: molecular mnemonics on the chromatin. *Nature Reviews Neuroscience* **14**, 97-111, (2013).
59 Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nature Reviews Molecular Cell Biology* **18**, 90-101, (2017).
60 Bannister, A. J., Schneider, R. & Kouzarides, T. Histone Methylation: Dynamic or Static? *Cell* **109**, 801-806, (2002).
61 Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* **13**, 343-357, (2012).
62 Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation. *Epigenetics* **7**, 1098-1108, (2012).
63 Sakabe, K., Wang, Z. & Hart, G. W. β-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences* **107**, 19915-19920, (2010).
64 Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proceedings of the National Academy of Sciences* **104**, 60-65, (2007).
65 Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications – writers that read. *EMBO reports* **16**, 1467-1481, (2015).
66 Allshire, R. C. & Karpen, G. H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nature Reviews Genetics* **9**, 923-937, (2008).
67 Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91**, 313-321, (1985).
68 Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903, (2008).
69 Beck, D. B., Oda, H., Shen, S. S. & Reinberg, D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev* **26**, 325-337, (2012).
70 Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361-1373, (2006).
71 Kuo, A. J. *et al.* The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* **484**, 115-119, (2012).
72 Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**, 1251-1262, (2004).
73 Fischle, W. *et al.* Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116-1122, (2005).

74 Justin, N. *et al.* Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nature Communications* **7**, 11316, (2016).

75 Poepsel, S., Kasinath, V. & Nogales, E. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nature Structural & Molecular Biology* **25**, 154-162, (2018).

76 Huang, H. *et al.* Lysine benzoylation is a histone mark regulated by SIRT2. *Nature Communications* **9**, 3374, (2018).

77 Farrelly, L. A. *et al.* Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**, 535-539, (2019).

78 Cheng, Y. *et al.* Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduction and Targeted Therapy* **4**, 62, (2019).

79 Jackson, V. Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. *Cell* **15**, 945-954, (1978).

80 Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends in Biochemical Sciences* **25**, 99-104, (2000).

81 Britton, L.-M. P., Gonzales-Cope, M., Zee, B. M. & Garcia, B. A. Breaking the histone code with quantitative mass spectrometry. *Expert Review of Proteomics* **8**, 631-643, (2011).

82 Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chemical reviews* **115**, 2376-2418, (2015).

83 Alabert, C. *et al.* Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nature Cell Biology* **16**, 281-291, (2014).

84 Johnson David, S., Mortazavi, A., Myers Richard, M. & Wold, B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* **316**, 1497-1502, (2007).

85 Pepke, S., Wold, B. & Mortazavi, A. Computation for ChIP-seq and RNA-seq studies. *Nature Methods* **6**, S22-S32, (2009).

86 Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* **2010**, 5384-5384, (2010).

87 de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes & Development* **26**, 11-24, (2012).

88 Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* **10**, 669-680, (2009).

89 Duan, Q., Chen, H., Costa, M. & Dai, W. Phosphorylation of H3S10 blocks the access of H3K9 by specific antibodies and histone methyltransferase. Implication in regulating chromatin dynamics and epigenetic inheritance during mitosis. *The Journal of biological chemistry* **283**, 33585-33590, (2008).

90 Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274-276, (2015).

91 Egelhofer, T. A. *et al.* An assessment of histone-modification antibody quality. *Nature Structural & Molecular Biology* **18**, 91-93, (2011).

92 Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends in Genetics* **27**, 389-396, (2011).
93 Su, Z. & Denu, J. M. Reading the Combinatorial Histone Language. *ACS Chemical Biology* **11**, 564-574, (2016).
94 Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational modification of histones. *Current Opinion in Chemical Biology* **45**, 35-47, (2018).
95 Müller, M. M. & Muir, T. W. Histones: At the Crossroads of Peptide and Protein Chemistry. *Chemical Reviews* **115**, 2296-2349, (2015).
96 Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annual Review of Biochemistry* **83**, 379-408, (2014).
97 Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment – Expanding the chemistry in biology. *Chemical Science* **6**, 50-69, (2015).
98 Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53-60, (2017).
99 Chen, H., Venkat, S., McGuire, P., Gan, Q. & Fan, C. Recent Development of Genetic Code Expansion for Posttranslational Modification Studies. *Molecules* **23**, 1662, (2018).
100 Elsässer, S. J., Ernst, R. J., Walker, O. S. & Chin, J. W. Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nature Methods* **13**, 158-164, (2016).
101 Ji, Y. *et al.* Genetically encoding ε-N-benzoyllysine in proteins. *Chemical Communications* **57**, 1798-1801, (2021).
102 Park, H.-S. *et al.* Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science* **333**, 1151-1154, (2011).
103 Zhang, M. S. *et al.* Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nature methods* **14**, 729-736, (2017).
104 Hoppmann, C. *et al.* Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nature Chemical Biology* **13**, 842-844, (2017).
105 Schwark, D. G., Schmitt, M. A. & Fisk, J. D. Dissecting the Contribution of Release Factor Interactions to Amber Stop Codon Reassignment Efficiencies of the Methanocaldococcus jannaschii Orthogonal Pair. *Genes (Basel)* **9**, 546, (2018).
106 Nguyen, D. P., Alai, M. M. G., Virdee, S. & Chin, J. W. Genetically Directing ε-N, N-Dimethyl-l-Lysine in Recombinant Histones. *Chemistry & Biology* **17**, 1072-1076, (2010).
107 Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically Encoding Nϵ-Methyl-l-lysine in Recombinant Histones. *Journal of the American Chemical Society* **131**, 14194-14195, (2009).
108 Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154, (1963).
109 Coin, I., Beyermann, M. & Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nature Protocols* **2**, 3247-3256, (2007).

110 Dawson Philip, E., Muir Tom, W., Clark-Lewis, I. & Kent Stephen, B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **266**, 776-779, (1994).

111 Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proceedings of the National Academy of Sciences* **95**, 6705, (1998).

112 Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nature Reviews Chemistry* **2**, 0122, (2018).

113 Stevens, A. J. *et al.* Design of a Split Intein with Exceptional Protein Splicing Activity. *Journal of the American Chemical Society* **138**, 2162-2165, (2016).

114 Stevens, A. J. *et al.* A promiscuous split intein with expanded protein engineering applications. *Proceedings of the National Academy of Sciences* **114**, 8538, (2017).

115 Burton, A. J., Haugbro, M., Parisi, E. & Muir, T. W. Live-cell protein engineering with an ultra-short split intein. *Proceedings of the National Academy of Sciences* **117**, 12041, (2020).

116 David, Y., Vila-Perelló, M., Verma, S. & Muir, T. W. Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nature Chemistry* **7**, 394-402, (2015).

117 Chalker, J. M. & Davis, B. G. Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Current Opinion in Chemical Biology* **14**, 781-789, (2010).

118 Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nature Communications* **5**, 4740, (2014).

119 Rayment, I. in *Methods in Enzymology* Vol. 276 171-179 (Academic Press, 1997).

120 Imiołek, M. *et al.* Residue-Selective Protein C-Formylation via Sequential Difluoroalkylation-Hydrolysis. *ACS Central Science* **7**, 145-155, (2021).

121 Sato, S. *et al.* Site-selective protein chemical modification of exposed tyrosine residues using tyrosine click reaction. *Bioconjugate Chemistry* **31**, 1417-1424, (2020).

122 Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science* **355**, 597-602, (2017).

123 Taylor, M. T., Nelson, J. E., Suero, M. G. & Gaunt, M. J. A protein functionalization platform based on selective reactions at methionine residues. *Nature* **562**, 563-568, (2018).

124 Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nature Chemistry* **10**, 205-211, (2018).

125 Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128**, 1003-1012, (2007).

126 Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A “Tag-and-Modify” Approach to Site-Selective Protein Modification. *Accounts of Chemical Research* **44**, 730-741, (2011).

127 Dadová, J., Galan, S. R. G. & Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Current Opinion in Chemical Biology* **46**, 71-81, (2018).

128 Bernardes, G. J. L., Chalker, J. M., Errey, J. C. & Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *Journal of the American Chemical Society* **130**, 5052-5053, (2008).
129 Guo, J., Wang, J., Lee, J. S. & Schultz, P. G. Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angewandte Chemie International Edition* **47**, 6399-6401, (2008).
130 Lin, Y. A. *et al.* Rapid cross-metathesis for reversible protein modifications via chemical access to Se-allyl-selenocysteine in proteins. *Journal of the American Chemical Society* **135**, 12156-12159 (2013).
131 He, P.-Y. *et al.* Late-stage peptide and protein modifications through phospha-Michael addition reaction. *Chemical Communications* **56**, 12632-12635, (2020).
132 Freedy, A. M. *et al.* Chemoselective installation of amine bonds on proteins through aza-Michael ligation. *Journal of the American Chemical Society* **139**, 18365-18375 (2017).
133 Dadová, J. *et al.* Precise probing of residue roles by post-translational β, γ-C, N aza-Michael mutagenesis in enzyme active sites. *ACS central science* **3**, 1168-1173 (2017).
134 Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science* **354**, 623-626, (2016).
135 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354** (2016).
136 Wright, T. H. & Davis, B. G. Post-translational mutagenesis for installation of natural and unnatural amino acid side chains into recombinant proteins. *nature protocols* **12**, 2243-2250 (2017).
137 Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260, (1997).
138 Huang, R.-c. C. & Bonner, J. Histone, a suppresor of chromosomal RNA synthesis. *Proceedings of the National Academy of Sciences* **48**, 1216-1222, (1962).
139 Hammond, C. M., Strømme, C. B., Huang, H., Patel, D. J. & Groth, A. Histone chaperone networks shaping chromatin function. *Nature Reviews Molecular Cell Biology* **18**, 141-158, (2017).
140 Abrams, S. T. *et al.* Circulating histones are mediators of trauma-associated lung injury. *Am J Respir Crit Care Med* **187**, 160-169, (2013).
141 Marsman, G., Zeerleder, S. & Luken, B. M. Extracellular histones, cell-free DNA, or nucleosomes: differences in immunostimulation. *Cell Death & Disease* **7**, e2518-e2518, (2016).
142 Silk, E., Zhao, H., Weng, H. & Ma, D. The role of extracellular histone in organ injury. *Cell death & disease* **8**, e2812-e2812, (2017).
143 Baake, M., Doenecke, D. & Albig, W. Characterisation of nuclear localisation signals of the four human core histones. *Journal of cellular biochemistry* **81**, 333-346 (2001).
144 Kaouass, M., Beaulieu, R. & Balicki, D. Histonefection: Novel and potent non-viral gene delivery. *Journal of Controlled Release* **113**, 245-254, (2006).

145 Han, H. *et al.* A comprehensive review on histone-mediated transfection for gene therapy. *Biotechnology Advances* **37**, 132-144, (2019).
146 Ryser, H.-P. Uptake of Protein by Mammalian Cells: An Underdeveloped Area: The penetration of foreign proteins into mammalian cells can be measured and their functions explored. *Science* **159**, 390-396 (1968).
147 Ryser, H. J.-P. & Hancock, R. Histones and basic polyamino acids stimulate the uptake of albumin by tumor cells in culture. *Science* **150**, 501-503 (1965).
148 Balicki, D., Putnam, C. D., Scaria, P. V. & Beutler, E. Structure and function correlation in histone H2A peptide-mediated gene transfer. *Proceedings of the National Academy of Sciences* **99**, 7467-7471 (2002).
149 Fritz, J. D., Herweijer, H., Zhang, G. & Wolff, J. A. Gene transfer into mammalian cells using histone-condensed plasmid DNA. *Human gene therapy* **7**, 1395-1404 (1996).
150 Puebla, I. *et al.* A recombinant H1 histone-based system for efficient delivery of nucleic acids. *Journal of biotechnology* **105**, 215-226 (2003).
151 Hariton-Gazal, E., Rosenbluh, J., Graessmann, A., Gilon, C. & Loyter, A. Direct translocation of histone molecules across cell membranes. *Journal of Cell Science* **116**, 4577-4586 (2003).
152 Gao, X.-z. & Huang, L.-b. Cationic liposome-mediated gene transfer. *Gene therapy* **2**, 710-722 (1995).
153 Niidome, T., Wakamatsu, M., Wada, A., Hirayama, T. & Aoyagi, H. Required structure of cationic peptide for oligonucleotide-binding and-delivering into cells. *Journal of peptide science: an official publication of the European Peptide Society* **6**, 271-279 (2000).
154 Zaitsev, S. *et al.* Histone H1-mediated transfection: role of calcium in the cellular uptake and intracellular fate of H1-DNA complexes. *Acta histochemica* **104**, 85-92 (2002).
155 Murphy, R. F., Jorgensen, E. D. & Cantor, C. R. Kinetics of histone endocytosis in Chinese hamster ovary cells. A flow cytofluorometric analysis. *Journal of Biological Chemistry* **257**, 1695-1701 (1982).
156 Rosenbluh, J., Singh, S. K., Gafni, Y., Graessmann, A. & Loyter, A. Non-endocytic penetration of core histones into petunia protoplasts and cultured cells: a novel mechanism for the introduction of macromolecules into plant cells. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1664**, 230-240 (2004).
157 Rosenbluh, J. *et al.* Translocation of histone proteins across lipid bilayers and Mycoplasma membranes. *Journal of molecular biology* **345**, 387-400 (2005).
158 Mi, Z., Mai, J., Lu, X. & Robbins, P. D. Characterization of a Class of Cationic Peptides Able to Facilitate Efficient Protein Transduction in Vitro and in Vivo. *Molecular Therapy* **2**, 339-347, (2000).
159 Fuchs, S. M. & Raines, R. T. Pathway for polyarginine entry into mammalian cells. *Biochemistry* **43**, 2438-2444, (2004).
160 Matsui, H., Tomizawa, K., Lu, Y. F. & Matsushita, M. Protein Therapy: in vivo protein transduction by polyarginine (11R) PTD and subcellular targeting delivery. *Curr Protein Pept Sci* **4**, 151-157, (2003).

161 Wagstaff, K. M., Glover, D. J., Tremethick, D. J. & Jans, D. A. Histone-mediated Transduction as an Efficient Means for Gene Delivery. *Molecular Therapy* **15**, 721-731, (2007).
162 Luhrs, K. A. *et al.* In vivo histone H1 migration from necrotic to viable tissue. *Oncotarget* **8**, 16275-16292, (2017).
163 Sulkowski, P. L. *et al.* Horizontal Transfer of Histone H3 by Mammalian Cells. *bioRxiv*, 2021.2006.2030.450524, (2021).
164 Allam, R., Kumar, S. V. R., Darisipudi, M. N. & Anders, H.-J. Extracellular histones in tissue injury and inflammation. *Journal of Molecular Medicine* **92**, 465-472, (2014).
165 Ribon, M., Mussard, J., Semerano, L., Singer, B. B. & Decker, P. Extracellular Chromatin Triggers Release of Soluble CEACAM8 Upon Activation of Neutrophils. *Frontiers in Immunology* **10**, (2019).
166 Bauden, M. *et al.* Circulating nucleosomes as epigenetic biomarkers in pancreatic cancer. *Clinical Epigenetics* **7**, 106, (2015).
167 Ungerer, V., Bronkhorst, A. J., Van den Ackerveken, P., Herzog, M. & Holdenrieder, S. Serial profiling of cell-free DNA and nucleosome histone modifications in cell cultures. *Scientific Reports* **11**, 9460, (2021).
168 Sadeh, R. *et al.* ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. *Nature Biotechnology* **39**, 586-598, (2021).
169 Abrams, S. T. *et al.* Human CRP defends against the toxicity of circulating histones. *The Journal of Immunology* **191**, 2495-2502 (2013).
170 Gauthier, V. J., Tyler, L. N. & Mannik, M. Blood clearance kinetics and liver uptake of mononucleosomes in mice. *The Journal of Immunology* **156**, 1151-1156 (1996).
171 Xu, J. *et al.* Extracellular histones are major mediators of death in sepsis. *Nature medicine* **15**, 1318-1321 (2009).
172 Wang, H. *et al.* Cellular uptake of extracellular nucleosomes induces innate immune responses by binding and activating cGMP-AMP synthase (cGAS). *Scientific Reports* **10**, 15385, (2020).
173 Monestier, M., Fasy, T. M., Losman, M. J., Novick, K. E. & Muller, S. Structure and binding properties of monoclonal antibodies to core histones from autoimmune mice. *Molecular immunology* **30**, 1069-1075 (1993).
174 Wildhagen, K. C. *et al.* Nonanticoagulant heparin prevents histone-mediated cytotoxicity in vitro and improves survival in sepsis. *Blood, The Journal of the American Society of Hematology* **123**, 1098-1101 (2014).
175 Allam, R. *et al.* Histones from dying renal cells aggravate kidney injury via TLR2 and TLR4. *Journal of the American Society of Nephrology* **23**, 1375-1388 (2012).
176 Semeraro, F. *et al.* Extracellular histones promote thrombin generation through platelet-dependent mechanisms: involvement of platelet TLR2 and TLR4. *Blood, The Journal of the American Society of Hematology* **118**, 1952-1961 (2011).
177 Xu, J., Zhang, X., Monestier, M., Esmon, N. L. & Esmon, C. T. Extracellular histones are mediators of death through TLR2 and TLR4 in mouse fatal liver injury. *The Journal of Immunology* **187**, 2626-2631 (2011).

178 Ekaney, M. L. *et al.* Impact of plasma histones in human sepsis and their contribution to cellular injury and inflammation. *Critical care* **18**, 1-9 (2014).
179 Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angewandte Chemie International Edition* **48**, 6974-6998 (2009).
180 Imiołek, M. *et al.* Selective radical trifluoromethylation of native residues in proteins. *Journal of the American Chemical Society* **140**, 1568-1571 (2018).
181 Dorman, G. & Prestwich, G. D. Benzophenone Photophores in Biochemistry. *Biochemistry* **33**, 5661-5673, (1994).
182 Farrell, I. S., Toroney, R., Hazen, J. L., Mehl, R. A. & Chin, J. W. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nature Methods* **2**, 377-384, (2005).
183 Dondoni, A., Massi, A., Nanni, P. & Roda, A. A New Ligation Strategy for Peptide and Protein Glycosylation: Photoinduced Thiol–Ene Coupling. *Chemistry – A European Journal* **15**, 11444-11449, (2009).
184 Griffiths, R. C. *et al.* Site-Selective Modification of Peptides and Proteins via Interception of Free-Radical-Mediated Dechalcogenation. *Angewandte Chemie* **132**, 23867-23875 (2020).
185 Rees, J. S., Li, X. W., Perrett, S., Lilley, K. S. & Jackson, A. P. Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters. *Curr Protoc Protein Sci* **80**, 19.27.11-19.27.18, (2015).
186 Kim, J. *et al.* Site-Selective Functionalization of Methionine Residues via Photoredox Catalysis. *Journal of the American Chemical Society* **142**, 21260-21266, (2020).
187 Yu, Y. *et al.* Chemoselective Peptide Modification via Photocatalytic Tryptophan β-Position Conjugation. *Journal of the American Chemical Society* **140**, 6797-6800, (2018).
188 Liu, X. & Gao, W. Precision Conjugation: An Emerging Tool for Generating Protein–Polymer Conjugates. *Angewandte Chemie International Edition* **60**, 11024-11035, (2021).
189 Messina, M. S., Messina, K. M. M., Bhattacharya, A., Montgomery, H. R. & Maynard, H. D. Preparation of biomolecule-polymer conjugates by grafting-from using ATRP, RAFT, or ROMP. *Progress in Polymer Science* **100**, 101186, (2020).
190 Song, H.-X., Han, Q.-Y., Zhao, C.-L. & Zhang, C.-P. Fluoroalkylation reactions in aqueous media: a review. *Green Chemistry* **20**, 1662-1731, (2018).
191 Ichiishi, N. *et al.* Protecting group free radical C–H trifluoromethylation of peptides. *Chemical Science* **9**, 4168-4175, (2018).
192 Imiołek, M. *et al.* Selective Radical Trifluoromethylation of Native Residues in Proteins. *Journal of the American Chemical Society* **140**, 1568-1571, (2018).
193 Kee, C. W. *et al.* 18F-Trifluoromethanesulfinate Enables Direct C–H 18F-Trifluoromethylation of Native Aromatic Residues in Peptides. *Journal of the American Chemical Society* **142**, 1180-1185, (2020).

194 Rahimidashaghoul, K. *et al.* Reductant-Induced Free Radical Fluoroalkylation of Nitrogen Heterocycles and Innate Aromatic Amino Acid Residues in Peptides and Proteins. *Chemistry – A European Journal* **25**, 15779-15785, (2019).
195 Noisier, A. F. M. *et al.* Late-Stage Functionalization of Histidine in Unprotected Peptides. *Angew Chem Int Ed Engl* **58**, 19096-19102, (2019).
196 Viehe, H. G., Janousek, Z., Merenyi, R. & Stella, L. The captodative effect. *Accounts of Chemical Research* **18**, 148-154, (1985).
197 Bottecchia, C. & Noël, T. Photocatalytic Modification of Amino Acids, Peptides, and Proteins. *Chemistry – A European Journal* **25**, 26-42, (2019).
198 King, T. A., Mandrup Kandemir, J., Walsh, S. J. & Spring, D. R. Photocatalytic methods for amino acid modification. *Chemical Society Reviews* **50**, 39-57, (2021).
199 Yang, T., Li, X.-M., Bao, X., Fung, Y. M. E. & Li, X. D. Photo-lysine captures proteins that bind lysine post-translational modifications. *Nature Chemical Biology* **12**, 70-72, (2016).
200 Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences* **99**, 11020-11024, (2002).
201 Gautier, A., Deiters, A. & Chin, J. W. Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells. *Journal of the American Chemical Society* **133**, 2124-2127, (2011).
202 Prier, C. K., Rankic, D. A. & MacMillan, D. W. C. Visible Light Photoredox Catalysis with Transition Metal Complexes: Applications in Organic Synthesis. *Chemical Reviews* **113**, 5322-5363, (2013).
203 Gao, X.-F., Du, J.-J., Liu, Z. & Guo, J. Visible-Light-Induced Specific Desulfurization of Cysteinyl Peptide and Glycopeptide in Aqueous Solution. *Organic Letters* **18**, 1166-1169, (2016).
204 Sato, S., Morita, K. & Nakamura, H. Regulation of Target Protein Knockdown and Labeling Using Ligand-Directed Ru(bpy)3 Photocatalyst. *Bioconjugate Chemistry* **26**, 250-256, (2015).
205 Aycock, R. A., Vogt, D. B. & Jui, N. T. A practical and scalable system for heteroaryl amino acid synthesis. *Chemical Science* **8**, 7998-8003, (2017).
206 Hohnstedt, L. F., Miniatas, B. O. & Waller, S. M. C. Aqueous Sodium Borohydride Chemistry. The Colnage Metals, Copper, Silver, and Gold. *Analytical Chemistry* **37**, 1163-1164, (1965).
207 Teegardin, K., Day, J. I., Chan, J. & Weaver, J. Advances in Photocatalysis: A Microreview of Visible Light Mediated Ruthenium and Iridium Catalyzed Organic Transformations. *Organic Process Research & Development* **20**, 1156-1163, (2016).
208 Josephson, B. *et al.* Light-driven post-translational installation of reactive protein side chains. *Nature* **585**, 530-537, (2020).
209 Matsui, J. K., Lang, S. B., Heitz, D. R. & Molander, G. A. Photoredox-Mediated Routes to Radicals: The Value of Catalytic Radical Generation in Synthetic Methods Development. *ACS Catalysis* **7**, 2563-2575, (2017).
210 Molander, G. A. Organotrifluoroborates: Another Branch of the Mighty Oak. *The Journal of Organic Chemistry* **80**, 7837-7848, (2015).

211 Primer, D. N., Karakaya, I., Tellis, J. C. & Molander, G. A. Single-Electron Transmetalation: An Enabling Technology for Secondary Alkylboron Cross-Coupling. *Journal of the American Chemical Society* **137**, 2195-2198, (2015).

212 Povie, G. *et al.* Catechols as Sources of Hydrogen Atoms in Radical Deiodination and Related Reactions. *Angewandte Chemie International Edition* **55**, 11221-11225, (2016).

213 Tyson, E. L., Niemeyer, Z. L. & Yoon, T. P. Redox Mediators in Visible Light Photocatalysis: Photocatalytic Radical Thiol–Ene Additions. *The Journal of Organic Chemistry* **79**, 1427-1436, (2014).

214 Ghosh, T. *et al.* Mixed-ligand complexes of ruthenium(II) containing new photoactive or electroactive ligands: synthesis, spectral characterization and DNA interactions. *J Biol Inorg Chem* **10**, 496-508, (2005).

215 Lin, Q., Li, Q., Batchelor-McAuley, C. & Compton, R. G. Two-Electron, Two-Proton Oxidation of Catechol: Kinetics and Apparent Catalysis. *The Journal of Physical Chemistry C* **119**, 1489-1495, (2015).

216 Springsteen, G. & Wang, B. A detailed examination of boronic acid–diol complexation. *Tetrahedron* **58**, 5291-5300, (2002).

217 Suzuki, J., Tanigawa, M., Inagi, S. & Fuchigami, T. Electrochemical Oxidation of Organotrifluoroborate Compounds. *ChemElectroChem* **3**, 2078-2083, (2016).

218 Musacchio, J. M. in *Biochemistry of Biogenic Amines* (eds Leslie L. Iversen, Susan D. Iversen, & Solomon H. Snyder) 1-35 (Springer US, 1975).

219 Haslam, E. & Cai, Y. Plant polyphenols (vegetable tannins): gallic acid metabolism. *Natural Product Reports* **11**, 41-66, (1994).

220 Boerjan, W., Ralph, J. & Baucher, M. Lignin Biosynthesis. *Annual Review of Plant Biology* **54**, 519-546, (2003).

221 Mollner, T. A. *et al.* Post-translational insertion of boron in proteins to probe and modulate function. *Nature Chemical Biology* **17**, 1245-1261, (2021).

222 Green, N. M. Avidin and streptavidin. *Methods Enzymol* **184**, 51-67, (1990).

223 Cigler, M. *et al.* Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo. *Angew Chem Int Ed Engl* **56**, 15737-15741, (2017).

224 Xiang, Z. *et al.* Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids. *Angew Chem Int Ed Engl* **53**, 2190-2193, (2014).

225 Yang, B. *et al.* Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. *Nat Commun* **8**, 2240, (2017).

226 Jones, L. H. Dehydroamino acid chemical biology: an example of functional group interconversion on proteins. *RSC Chemical Biology* **1**, 298-304, (2020).

227 Zhang, M., He, P. & Li, Y. Contemporary Approaches to α,β-Dehydroamino Acid Chemical Modifications. *Chemical Research in Chinese Universities* **37**, 1044-1054, (2021).

228 Griffiths, R. C. *et al.* Site-Selective Installation of Nϵ-Modified Sidechains into Peptide and Protein Scaffolds via Visible-Light-Mediated Desulfurative C–C Bond Formation. *Angewandte Chemie International Edition* **61**, e202110223, (2022).

229 Lai, K.-Y. *et al.* LanCLs add glutathione to dehydroamino acids generated at phosphorylated sites in the proteome. *Cell* **184**, 2680-2695.e2626, (2021).

230 Siewert, B. *et al.* Turning on the red phosphorescence of a [Ru(tpy)(bpy)(Cl)]Cl complex by amide substitution: self-aggregation, toxicity, and cellular localization of an emissive ruthenium-based amphiphile. *Chemical Communications* **53**, 11126-11129, (2017).
231 Hananya, N., Daley, S. K., Bagert, J. D. & Muir, T. W. Synthesis of ADP-Ribosylated Histones Reveals Site-Specific Impacts on Chromatin Structure and Function. *Journal of the American Chemical Society* **143**, 10847-10852, (2021).
232 Al Temimi, A. H. K. *et al.* Fine-tuning of lysine side chain modulates the activity of histone lysine methyltransferases. *Scientific Reports* **10**, 21574, (2020).
233 Temimi, A. H. K. A. *et al.* Lysine Possesses the Optimal Chain Length for Histone Lysine Methyltransferase Catalysis. *Scientific Reports* **7**, 16148, (2017).
234 Proietti, G., Wang, Y., Rainone, G. & Mecinović, J. Effect of lysine side chain length on histone lysine acetyltransferase catalysis. *Scientific Reports* **10**, 13046, (2020).
235 Xiao, R. *et al.* Toward an Orthogonal Protein Lysine Acylation and Deacylation System. *ChemBioChem* **n/a**, e202100551.
236 Huang, R. *et al.* Site-Specific Introduction of an Acetyl-Lysine Mimic into Peptides and Proteins by Cysteine Alkylation. *Journal of the American Chemical Society* **132**, 9986-9987, (2010).
237 Pieters, B. J. G. E. *et al.* Installation of Trimethyllysine Analogs on Intact Histones via Cysteine Alkylation. *Bioconjugate Chemistry* **30**, 952-958, (2019).
238 Al Temimi, A. H. K. *et al.* Lysine Ethylation by Histone Lysine Methyltransferases. *Chembiochem* **21**, 392-400, (2020).
239 Kleiner, R. E., Hang, L. E., Molloy, K. R., Chait, B. T. & Kapoor, T. M. A Chemical Proteomics Approach to Reveal Direct Protein-Protein Interactions in Living Cells. *Cell Chem Biol* **25**, 110-120.e113, (2018).
240 Nguyen, T.-A., Cigler, M. & Lang, K. Expanding the Genetic Code to Study Protein–Protein Interactions. *Angewandte Chemie International Edition* **57**, 14350-14361, (2018).
241 Binda, O. *et al.* A Chemical Method for Labeling Lysine Methyltransferase Substrates. *ChemBioChem* **12**, 330-334, (2011).
242 Dyer, P. N. *et al.* in *Methods in Enzymology* Vol. 375 23-44 (Academic Press, 2003).
243 Maeshima, K. *et al.* Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *EMBO J* **35**, 1115-1132, (2016).
244 Rogge, R. A. *et al.* Assembly of nucleosomal arrays from recombinant core histones and nucleosome positioning DNA. *J Vis Exp*, 50354, (2013).
245 Dhall, A., Weller, C. E., Chu, A., Shelton, P. M. M. & Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1–CoREST Complex. *ACS Chemical Biology* **12**, 2275-2280, (2017).
246 Leonen, C. J. A., Upadhyay, E. & Chatterjee, C. Studies of biochemical crosstalk in chromatin with semisynthetic histones. *Current Opinion in Chemical Biology* **45**, 27-34, (2018).
247 Amamoto, Y. *et al.* Synthetic Posttranslational Modifications: Chemical Catalyst-Driven Regioselective Histone Acylation of Native Chromatin. *Journal of the American Chemical Society* **139**, 7568-7576, (2017).

248 Fujiwara, Y. *et al.* Live-cell epigenome manipulation by synthetic histone acetylation catalyst system. *Proceedings of the National Academy of Sciences* **118**, e2019554118, (2021).
249 Faulkner, S. *et al.* Intact nucleosomal context enables chromodomain reader MRG15 to distinguish H3K36me3 from -me2. *bioRxiv*, 2020.2004.2030.070136, (2020).
250 Chen, X.-H. *et al.* Genetically Encoding an Electrophilic Amino Acid for Protein Stapling and Covalent Binding to Native Receptors. *ACS Chemical Biology* **9**, 1956-1961, (2014).
251 Wang, Y. *et al.* Cofactor Biogenesis in Cysteamine Dioxygenase: C−F Bond Cleavage with Genetically Incorporated Unnatural Tyrosine. *Angewandte Chemie International Edition* **57**, 8149-8153, (2018).
252 Xiang, Z. *et al.* Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nature Methods* **10**, 885-888, (2013).
253 Wang, N. *et al.* Genetically Encoding Fluorosulfate-l-tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo. *Journal of the American Chemical Society* **140**, 4995-4999, (2018).
254 Hoppmann, C. *et al.* Genetically Encoding Photoswitchable Click Amino Acids in Escherichia coli and Mammalian Cells. *Angewandte Chemie International Edition* **53**, 3932-3936, (2014).
255 Furman, J. L. *et al.* A Genetically Encoded aza-Michael Acceptor for Covalent Cross-Linking of Protein–Receptor Complexes. *Journal of the American Chemical Society* **136**, 8411-8417, (2014).
256 Xuan, W., Li, J., Luo, X. & Schultz, P. G. Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins. *Angewandte Chemie International Edition* **55**, 10065-10068, (2016).
257 Xuan, W., Shao, S. & Schultz, P. G. Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angewandte Chemie International Edition* **56**, 5096-5100, (2017).
258 St. Amant, A. H. *et al.* A Diene-Containing Noncanonical Amino Acid Enables Dual Functionality in Proteins: Rapid Diels–Alder Reaction with Maleimide or Proximity-Based Dimerization. *Angewandte Chemie International Edition* **58**, 8489-8493, (2019).
259 Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nature Methods* **2**, 261-268, (2005).
260 Xie, X. *et al.* Genetically Encoded Photoaffinity Histone Marks. *Journal of the American Chemical Society* **139**, 6522-6525, doi:10.1021/jacs.7b01431 (2017).
261 Hancock, R. L. *et al.* Lysine-241 Has a Role in Coupling 2OG Turnover with Substrate Oxidation During KDM4-Catalysed Histone Demethylation. *ChemBioChem* **19**, 917-921, (2018).
262 Sekirnik, R. *et al.* Inhibition of the histone lysine demethylase JMJD2A by ejection of structural Zn(ii). *Chemical Communications*, 6376-6378, (2009).
263 Siriwardana, K. *et al.* Studying the Effects of Cysteine Residues on Protein Interactions with Silver Nanoparticles. *J Phys Chem C Nanomater Interfaces* **119**, 2910-2916, (2015).

264 Freedman, H. H. & Dubois, R. A. An improved Williamson ether synthesis using phase transfer catalysis. *Tetrahedron Letters* **16**, 3251-3254, (1975).
265 Mandal, S. *et al.* A review on the advancement of ether synthesis from organic solvent to water. *RSC Advances* **6**, 69605-69614, (2016).
266 Levin, M., Stark, M. & Assaraf, Y. G. The JmjN domain as a dimerization interface and a targeted inhibitor of KDM4 demethylase activity. *Oncotarget* **9**, 16861-16882, (2018).
267 Shin, S. & Janknecht, R. Diversity within the JMJD2 histone demethylase family. *Biochem Biophys Res Commun* **353**, 973-977, (2007).
268 Jemmis, E. D. & Jayasree, E. G. Analogies between Boron and Carbon. *Accounts of Chemical Research* **36**, 816-824, (2003).
269 O'Neill, M. A., Ishii, T., Albersheim, P. & Darvill, A. G. Rhamnogalacturonan II: Structure and Function of a Borate Cross-Linked Cell Wall Pectic Polysaccharide. *Annual Review of Plant Biology* **55**, 109-139, (2004).
270 Kinder, D. H. & Ames, M. M. Synthesis of 2-amino-3-boronopropionic acid: A boron-containing analog of aspartic acid. *The Journal of Organic Chemistry* **52**, 2452-2454 (1987).
271 Kinder, D. H., Frank, S. K. & Ames, M. M. Analogs of carbamyl aspartate as inhibitors of dihydroorotase: preparation of boronic acid transition-state analogs and a zinc chelator carbamylhomocysteine. *Journal of medicinal chemistry* **33**, 819-823 (1990).
272 Clore, G. M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical reviews* **109**, 4108-4139 (2009).
273 Zheng, C. & Hayes, J. J. Structures and interactions of the core histone tail domains. *Biopolymers: Original Research on Biomolecules* **68**, 539-546 (2003).
274 Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proceedings of the National Academy of Sciences* **109**, 17857-17862 (2012).
275 Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. *Nature structural & molecular biology* **26**, 3-13 (2019).
276 Wei, Y., Mizzen, C. A., Cook, R. G., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. *Proceedings of the National Academy of Sciences* **95**, 7480-7484 (1998).
277 Prigent, C. & Dimitrov, S. Phosphorylation of serine 10 in histone H3, what for? *Journal of cell science* **116**, 3677-3685 (2003).
278 Sawicka, A. & Seiser, C. Histone H3 phosphorylation–a versatile chromatin modification for different occasions. *Biochimie* **94**, 2193-2201 (2012).
279 Dormann, H. L., Tseng, B. S., Allis, C. D., Funabiki, H. & Fischle, W. Dynamic regulation of effector protein binding to histone modifications: the biology of HP1 switching. *Cell cycle* **5**, 2842-2851 (2006).
280 Terada, Y. Aurora-B/AIM-1 Regulates the Dynamic Behavior of HP1α at the G2–M Transition. *Molecular Biology of the Cell* **17**, 3232-3241, (2006).

281 Mateescu, B., England, P., Halgand, F., Yaniv, M. & Muchardt, C. Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. *EMBO Rep* **5**, 490-496, (2004).
282 Olsen, C. A. Expansion of the Lysine Acylation Landscape. *Angewandte Chemie International Edition* **51**, 3755-3756, (2012).
283 Faulkner, S., Maksimovic, I. & David, Y. A chemical field guide to histone nonenzymatic modifications. *Current Opinion in Chemical Biology* **63**, 180-187, (2021).
284 Finnin, M. S., Donigian, J. R. & Pavletich, N. P. Structure of the histone deacetylase SIRT2. *Nat Struct Biol* **8**, 621-625, (2001).
285 Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J* **27**, 406-420, (2008).
286 Zhu, X. *et al.* Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat Genet* **46**, 287-293, (2014).
287 Newbold, R. F. & Mokbel, K. Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis. *Anticancer Res* **30**, 3309-3311 (2010).
288 Bleuyard, J.-Y. *et al.* MRG15-mediated tethering of PALB2 to unperturbed chromatin protects active genes from genotoxic stress. *Proceedings of the National Academy of Sciences* **114**, 7671-7676, (2017).
289 Hayakawa, T. *et al.* MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. *J Cell Sci* **123**, 1124-1130, (2010).
290 Hayakawa, T. *et al.* RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. *Genes Cells* **12**, 811-826, (2007).
291 Gautam, D., Johnson, B. A., Mac, M. & Moody, C. A. SETD2-dependent H3K36me3 plays a critical role in epigenetic regulation of the HPV31 life cycle. *PLOS Pathogens* **14**, e1007367, (2018).
292 Zheng, Y., Huang, X. & Kelleher, N. L. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Current opinion in chemical biology* **33**, 142-150, (2016).
293 Rothbart, S. B. *et al.* An Interactive Database for the Assessment of Histone Antibody Specificity. *Mol Cell* **59**, 502-511, (2015).
294 Claridge, T. D. W. in *High-Resolution NMR Techniques in Organic Chemistry (Third Edition)* (ed Timothy D. W. Claridge) 421-455 (Elsevier, 2016).
295 Bagert, J. D. *et al.* Oncohistone mutations enhance chromatin remodeling and alter cell fates. *Nat Chem Biol* **17**, 403-411, (2021).
296 Nacev, B. A. *et al.* The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* **567**, 473-478, (2019).
297 Lewis, P. W. *et al.* Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857-861, (2013).
298 Lim, J. C., You, Z., Kim, G. & Levine, R. L. Methionine sulfoxide reductase A is a stereospecific methionine oxidase. *Proceedings of the National Academy of Sciences* **108**, 10472-10477, (2011).

299 Jang, Y. *et al.* H3.3K4M destabilizes enhancer H3K4 methyltransferases MLL3/MLL4 and impairs adipose tissue development. *Nucleic Acids Res* **47**, 607-620, (2019).
300 Raj, R., Lercher, L., Mohammed, S. & Davis, B. G. Synthetic Nucleosomes Reveal that GlcNAcylation Modulates Direct Interaction with the FACT Complex. *Angewandte Chemie International Edition* **55**, 8918-8922, (2016).
301 Campos, Eric I. *et al.* Analysis of the Histone H3.1 Interactome: A Suitable Chaperone for the Right Event. *Molecular Cell* **60**, 697-709, (2015).
302 Campos, E. I. *et al.* The program for processing newly synthesized histones H3.1 and H4. *Nature Structural & Molecular Biology* **17**, 1343-1351, (2010).
303 Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell* **116**, 51-61, (2004).
304 Wu, R. S. & Bonner, W. M. Separation of basal histone synthesis from S-phase histone synthesis in dividing cells. *Cell* **27**, 321-330 (1981).
305 Wu, R. S., Tsai, S. & Bonner, W. M. Patterns of histone variant synthesis can distinguish GO from G1 cells. *Cell* **31**, 367-374 (1982).
306 Bostock, C. J., Prescott, D. M. & Kirkpatrick, J. B. An evaluation of the double thymidine block for synchronizing mammalian cells at the G1-S border. *Experimental Cell Research* **68**, 163-168, (1971).
307 Wilkinson, A. W. & Gozani, O. Histone-binding domains: Strategies for discovery and characterization. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1839**, 669-675, (2014).
308 Toyama, B. H. *et al.* Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**, 971-982, (2013).
309 Feldman, M., Vershinin, Z., Goliand, I., Elia, N. & Levy, D. The methyltransferase SETD6 regulates Mitotic progression through PLK1 methylation. *Proc Natl Acad Sci U S A* **116**, 1235-1240, (2019).
310 Ameyar, M., Wisniewska, M. & Weitzman, J. B. A role for AP-1 in apoptosis: the case for and against. *Biochimie* **85**, 747-752, (2003).
311 Karin, M., Liu, Z.-g. & Zandi, E. AP-1 function and regulation. *Current Opinion in Cell Biology* **9**, 240-246, (1997).
312 Desterke, C., Bennaceur-Griscelli, A. & Turhan, A. G. EGR1 dysregulation defines an inflammatory and leukemic program in cell trajectory of human-aged hematopoietic stem cells (HSC). *Stem Cell Res Ther* **12**, 419, (2021).
313 Gómez-Martín, D., Díaz-Zamudio, M., Galindo-Campos, M. & Alcocer-Varela, J. Early growth response transcription factors and the modulation of immune response: Implications towards autoimmunity. *Autoimmunity Reviews* **9**, 454-458, (2010).
314 Hollander, M. C., Zhan, Q., Bae, I. & Fornace, A. J., Jr. Mammalian GADD34, an Apoptosis- and DNA Damage-inducible Gene *. *Journal of Biological Chemistry* **272**, 13731-13737, (1997).
315 Ku, H.-C. & Cheng, C.-F. Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Frontiers in Endocrinology* **11**, (2020).

316 Alauddin, M. M. Positron emission tomography (PET) imaging with (18)F-based radiotracers. *Am J Nucl Med Mol Imaging* **2**, 55-76 (2012).
317 Lum, G. & Gambino, S. R. A comparison of serum versus heparinized plasma for routine chemistry tests. *Am J Clin Pathol* **61**, 108-113, (1974).
318 Dooley, K. & Zon, L. I. Zebrafish: a model system for the study of human disease. *Current Opinion in Genetics & Development* **10**, 252-256, (2000).
319 Veldman, M. B. & Lin, S. Zebrafish as a Developmental Model Organism for Pediatric Research. *Pediatric Research* **64**, 470-476, (2008).
320 Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Developmental Dynamics* **203**, 253-310, (1995).
321 Rosen, J. N., Sweeney, M. F. & Mably, J. D. Microinjection of zebrafish embryos to analyze gene function. *J Vis Exp*, 1115, (2009).
322 Shestopalov, I. A., Sinha, S. & Chen, J. K. Light-controlled gene silencing in zebrafish embryos. *Nature Chemical Biology* **3**, 650-651, (2007).
323 Agarwal, P., Beahm, B. J., Shieh, P. & Bertozzi, C. R. Systemic Fluorescence Imaging of Zebrafish Glycans with Bioorthogonal Chemistry. *Angew Chem Int Ed Engl* **54**, 11504-11510, (2015).
324 Xiong, Z. *et al.* In vivo proteomic mapping through GFP-directed proximity-dependent biotin labelling in zebrafish. *eLife* **10**, e64631, (2021).
325 Murata, M. *et al.* Chlamydial SET domain protein functions as a histone methyltransferase. *Microbiology* **153**, 585-592 (2007).
326 Hamon, M. A. & Cossart, P. Histone Modifications and Chromatin Remodeling during Bacterial Infections. *Cell Host & Microbe* **4**, 100-109, (2008).