

Post-translational mutagenesis: radical methods for protein modification

A thesis submitted to the board of the faculty of physical sciences in partial fulfillment for the degree of:

Doctor of Philosophy (D.Phil)
at the University of Oxford

By

Thomas Wright

St. Hilda's College, University of Oxford

May 2017

Abstract

The naturally occurring post-translational modification of proteins expands the chemical and structural diversity available for protein function. Enzymatic processes are known to modify proteins after translation, and in many cases these modifications have been demonstrated to be critical for biological function. However, the paucity of methods for introduction of site-specific modifications to proteins remains a key obstacle to their biochemical study. Access to these naturally modified proteins has long been complicated by their low abundance, difficulties in purifying homogeneous samples from cellular samples and the intractability of bacterial expression systems towards post-translational modification.

In this thesis, we report an important step towards circumventing these issues by using a novel 'tag and modify' chemical strategy, termed 'post-translational mutagenesis'. We describe the development and application of $C(sp^3)–C(sp^3)$ bond-forming reactions on proteins under biocompatible conditions, exploiting unusual free-radical chemistry to form a wide diversity of protein side-chains. Natural, unnatural, posttranslationally-modified (methylated, glycosylated, phosphorylated, hydroxylated) and labeled (fluorinated, isotopically-labeled) side-chains have been installed on a number of proteins. Extensive characterisation of modification site, connectivity and modified protein structure and function confirm the benign nature of our protocols. The modified proteins thus produced have been used to investigate novel aspects of epigenetic regulation of histone proteins, in the context of free histones and in nucleosome particles, and as novel 'isotope PTMs' for protein characterisation by NMR.

We believe this approach, a form of 'protein editing', will have wide application in biochemistry and synthetic biology for accessing diverse modified proteins for research.

Publication list

1 Wright, T. H., Vallée, M. R. J. & Davis, B. G. From Chemical Mutagenesis to Post‐Expression Mutagenesis: A 50 Year Odyssey. *Angewandte Chemie International Edition* **55**, 5896-5903 (2016).

2 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

3 Wright, T. H. & Davis, B. G. Post-translational mutagenesis for installation of natural and unnatural amino acids into recombinant proteins. *Nature Protocols 12 (10), 2243-2250*

Acknowledgements

I would like to thank Prof. Ben Davis for giving me the opportunity to spend 3.5 memorable years in his group – a formative period for me in both science and life.

I would also like to thank Prof. Andy Baldwin for interest and patience in my sporadically timed enthusiasms for protein NMR.

This thesis, and the broader project from which it sprang, were truly collaborative efforts. I would like to thank Rafal Wiewiora, Ben Bower, Justin Chalker, Ritu Raj, Gogulan Karunanithy and Wai-Lung Ng in particular for their assistance with key experiments. Special mention to Marie-Laetitia Thezenas and Benedikt Kessler for endless help with MS/MS.

I would also like to thank Bala, Tom P and Seb for day-to-day advice on lab and living.

Finally, thanks to all members of BGD group, past and present, for all the good times and support. Gustaf Hemberg, Costi Giurgiu and Leanne Minall get a special mention for co-habitation.

Abbreviations

^{1}H NMR	proton nuclear magnetic resonance
^{13}C NMR	carbon-13 nuclear magnetic resonance
°C	degree Celsius
[M]	MS (see MS) molecular peak
Amp.	ampicillin
aq.	aqueous
Ac	acetyl
ACN	acetonitrile
Ar	argon
calc.	calculated
CD_3OD	deuterated methanol
$CDCl_3$	deuterated chloroform
COSY	correlated spectroscopy
D_2O	deuterated water
Da	dalton
DCM	dichloromethane
Dha	dehydroalanine
DMAP	4-dimethylamino pyridine
DNA	2'-deoxyribose nucleic acid
EPL	expressed protein ligation
eq.	equivalent
ES	electrospray
Et	ethyl
et al.	et alia
Et_2O	diethyl ether
EtOAc	ethyl acetate
FA	formic acid
g	gram
GlcNAc	N-acetylglucosamine
H3	Histone H3
HMBC	hydrogen multiple bond correlation

HMQC	hydrogen multiple quantum correlated spectroscopy
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
Hz	hertz
h	hour
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	NMR coupling constant (Hz)
Kan	kanamycin
kbp	kilo-base pair
kDa	kilodalton
LB	Luria Bertani
LC-MS	liquid chromatography-mass spectrometry
NCL	native chemical ligation
NH_4Cl	ammonium chloride
NH_4OAc	ammonium acetate
PTM	post-translational modification
SBL	subtilisin from *Bacillus lentus*

Experiments performed in collaboration

Chapter 3

Initial results reported in the introduction are from Justin Chalker.
Initial experiments with SBL and metal-mediated systems were performed with Rafal Wiewiora and Robert Vallee.

Chapter 4

EPR experiments were performed by Ben Bower and Will Myers.

Chapter 5

NMR studies were performed with Gogulan Karuthanithy.
Interaction proteomics experiments were performed by Ritu Raj.

Chapter 6

Western blot of H3-S10ph mimetics was performed by Wai-Lung Ng.
Alphascreen assay was performed by Wai-Lung Ng and Maola Khan.
Preliminary NMR studies were performed by Wai-Lung Ng and Tim Claridge.
Synthesis of model compound **6** was performed by Mateusz Imiolek.
Final NMR studies were performed with Gogulan Karuthanithy and Prof. Andrew Baldwin.

Table of contents

Introduction: Proteins and their modifications 2

Total chemical synthesis of proteins 4

Biological methods for the incorporation of post-translationally modified and unnatural amino acids into proteins 15

Hybrid approaches – the 'tag and modify' strategy 21

'Chemical mutagenesis' 30

Goals of this thesis 32

Chapter 2: General experimental procedures 51

Chapter 3: Development of a low-valent metal approach to post-translational mutagenesis 63

Introduction 63

Results 68

Experimental section: Chapter 3 85

Chapter 4: The discovery of borohydride-mediated 'chemical mutation' 127

Introduction 127

Results 128

Experimental section: Chapter 4 156

Chapter 5: Post-translational mutagenesis for the investigation of histone methylation 222

Introduction 222

Results 231

Experimental section: Chapter 5 252

Chapter 6: Post-translational mutagenesis for protein phosphorylation 288

Introduction 288

Results 296

Experimental section: Chapter 6 309

Chapter 7: 'Photoredox mutagenesis' for radical- mediated protein modification 330

Introduction 330

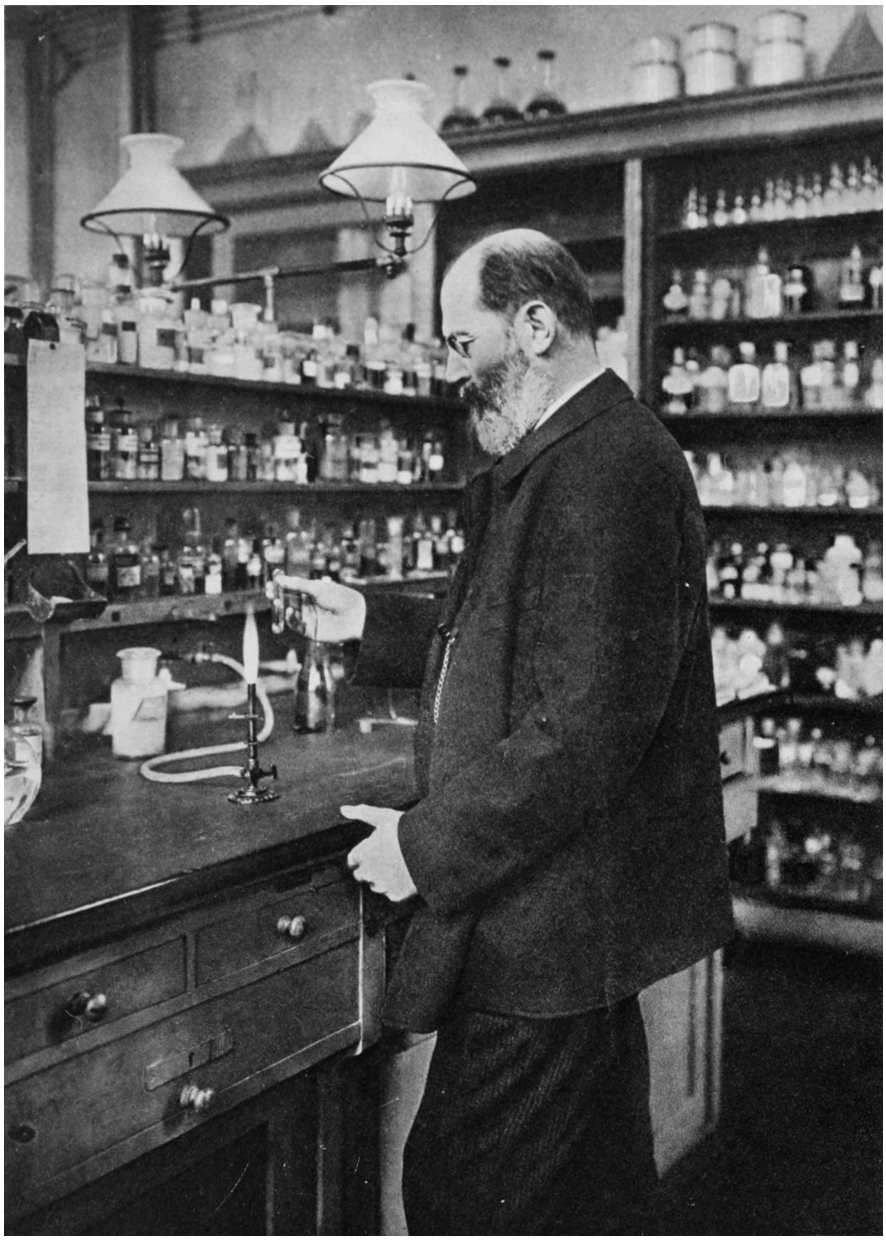
Results 332

Experimental section: Chapter 7 343

Conclusion: Prospects for chemical mutation of proteins 377

“Whereas cautious professional colleagues fear that rational study of this class of compounds *[proteins]*, because of their complicated structure and their highly inconvenient physical characteristics, would today still uncover insurmountable difficulties, other optimistically endowed observers, among which I would count myself, are inclined to the view that an attempt should at least be made to besiege this virgin fortress with all the expedients of the present; because only through this hazardous affair can the limitations of the ability of our methods be ascertained.”

Emil Fischer



Introduction: Proteins and their modifications

Proteins, as the 'workhorse' molecules of life, have long attracted interest from synthetic chemists. The sheer size and chemical and structural complexity of proteins, not to mention their fundamental relevance to biology, have made proteins enticing targets for both total synthesis and chemical modification. The chemical modification of proteins has been used to both unravel a given protein's native biological functions, and to extend and enhance function for biotechnology applications. In recent years, efforts towards 'bio-orthogonal'[1] chemistry have enabled the modification of proteins in a cellular, and even whole organism context, through the use of highly selective and rapid chemical reactions. The clinical success of antibody-drug conjugates[2], the further development of intact protein vaccine platforms[3], and the increased interest in enzymes for industrial biotechnology[4-6], have also stimulated recent intense research efforts in chemical protein modification.

In a natural context, proteins are synthesized by the ribosomal translation machinery which catalyses the key amide bond formation between a polypeptide and an incoming amino acid. However, in all organisms examined to date, extensive post-translational modification of proteins expands the chemical diversity available for protein function[7] (**Figure 1a**). Both enzymatic and strictly chemical processes are known to modify proteins after translation, and in many cases these modifications have been demonstrated to be critical for biological function. For example, the methylation[8] and acetylation[9] of histone proteins in the eukaryotic nucleus has long been recognized as a key mechanism in the regulation of gene expression.

Access to these naturally modified proteins has long been complicated by their low abundance, difficulties in purifying homogeneous samples from cellular samples and the intractability of bacterial expression systems towards post-translational modification. In recent years, intense effort has

begun to make headway in overcoming these challenges, using both strictly chemical[10], strictly biological[11], and hybrid approaches[12]. However, important challenges remain and most progress to date has been piecemeal, with most studies targeting a single modification for incorporation. A facile and general route to any modified protein of interest is still lacking.

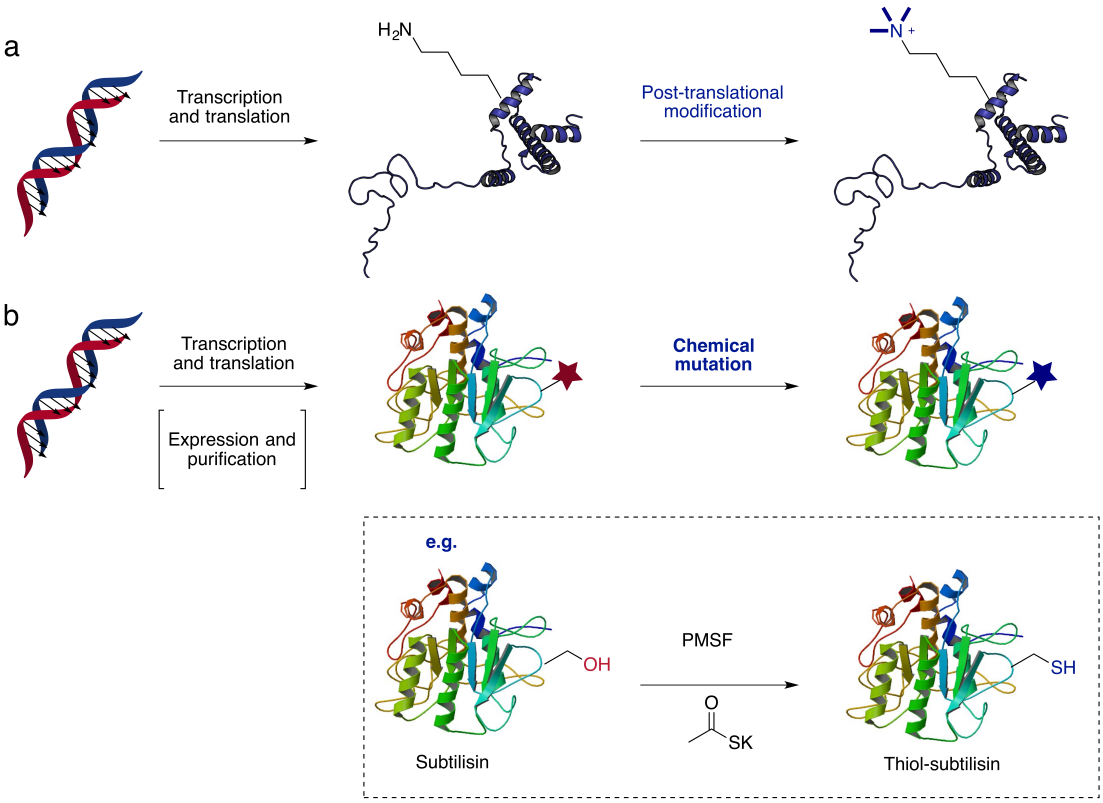


Figure 1: Natural (a.) and synthetic (b.) approaches to post-translational modification. a. In nature, enzymatic post-translational modification expands the diversity of protein side-chains. b. 'Chemical mutagenesis' mimics enzymatic post-translational modification by altering protein side-chains on intact proteins following transcription and translation. Chemical modification takes the place of enzymatic modification, allowing almost unlimited scope for side-chain alteration.

In the late 1960s, working before the biotechnology revolution enabled by recombinant DNA techniques, a group of prescient biochemists proposed a chemical strategy that would enable the divergent construction of modified proteins[13-16]. Termed 'chemical mutagenesis'[15], these scientists proposed a

system for protein chemical modification in which natural residues (and their post-translationally modified counterparts) could be interchanged by purely chemical techniques (**Figure 1b**). However, their work was seemingly forgotten as purely genetic techniques colonized biological research and led to the explosion of the 'biotechnology revolution'. In this thesis, we revise the concept of 'chemical mutagenesis', which we have termed 'post-translational mutagenesis' to avoid confusion with other uses of the original term and to highlight the facile access to post-translationally modified residues enabled by the method.

As alluded to above, a plethora of techniques have been developed for chemical modification of proteins, motivated by a number of research challenges. In the introduction to this thesis, the focus is placed on efforts to obtain proteins bearing natural post-translational modifications, or their functional chemical mimics, in quantities sufficient for biochemical study. Efforts to 'label' proteins with (for example) non-natural fluorescent probes for tracking *in-cellulo* are largely outside the scope of this thesis, and have been extensively reviewed elsewhere[17,18].

Total chemical synthesis of proteins

Chemically speaking, a protein can be viewed as a linear chain of amino acids linked into a polypeptide biopolymer by amide bonds between the N-terminal amino and C-terminal carboxyl functionalities of free amino acids (**Figure 2**). Subsequent folding into secondary and tertiary structure is determined by the biophysical properties of the specific chain of amino acids present. The chemical formation of the peptide bond by reaction of an activated carboxyl group with a free amine group on a second amino acid is the basis of both nature's protein synthesis (via the ribosome) and the contemporary chemical synthesis of peptides and proteins (**Figure 2**).

Since the early work of Emil Fischer[19,20] and Vigneaud[21] on solution phase peptide coupling, culminating in the first synthesis of a peptide hormone, oxytocin[22,23], organic chemists have been challenged by the purely chemical

synthesis of peptides. Major difficulties in the solution phase approach to peptide synthesis include epimerization of the Cα of the C-terminal amino acid upon activation in solution, the challenge of purifying protected peptide intermediates, and, critically, the extremely low solubility of protected peptides in solution as the chain length increases[10].

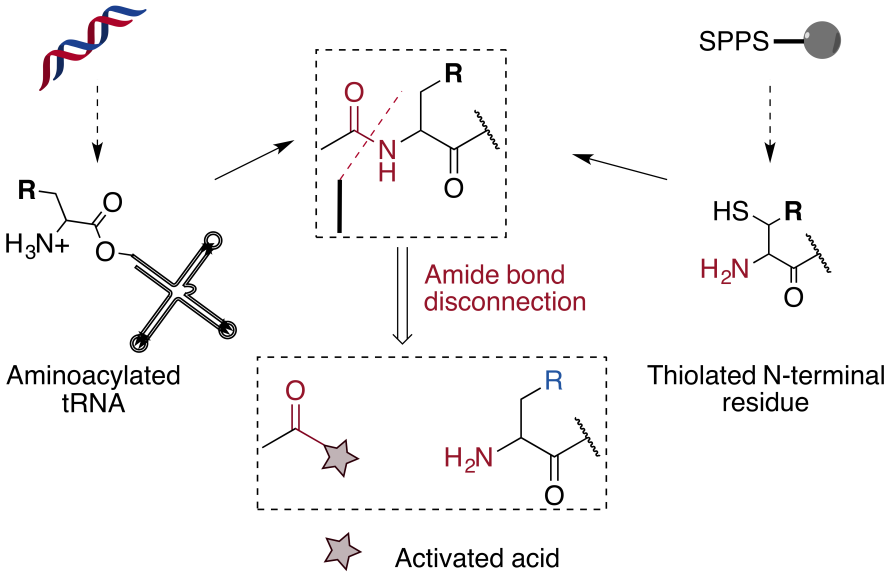


Figure 2: In nature and in the laboratory, amide bond formation is the key step in protein synthesis[14].

A key breakthrough occurred with the work of Bruce Merrifield[24], who demonstrated that immobilization of the growing polypeptide chain on a polystyrene support circumvented many of the problems associated with solution-phase approaches[25]. Attachment of the growing peptide chain to an insoluble resin support creates a copolymer-type species that swells in organic solvent, exerting a strong solvation effect[26]. The use of a resin support thus overcomes the solubility problems associated with the solution phase approach to peptide synthesis. Indeed, this technological development enabled Merrifield to complete the first total synthesis of a natural protein, that of ribonuclease A, in 1969[27]. This monumental feat involved 369 chemical reactions, followed by air oxidation to form disulfide bonds and fold the final enzyme[27]. In addition to being a heroic example of persistence in organic synthesis, key hypotheses as to protein structure and function were confirmed by this work, namely, that the sequence of the

polypeptide encodes all the information necessary for correct protein folding and enzymatic activity (extending earlier work by Anfinsen[28] on isolated bovine RNase). Such 'proof by synthesis' remains a key motivation for experimentation in organic chemistry as applied to biological systems.

Following the introduction of solid-phase synthesis, further advances in protecting group chemistry, particularly the development of the Fmoc[29,30] amino protecting scheme, have enabled the routine synthesis of simple peptides up to about 40-50 residues in length. Continual incremental advances in resins, linkers, coupling agents and coupling conditions[31], in tandem with more fundamental shifts, such as the introduction of flow-based synthetic methods[32], promise to render ever more facile access to short peptides by SPPS.

Native chemical ligation

The difficulty in synthesis of large polypeptides by SPPS mandates the convergent assembly of (relatively) small peptide fragments into larger polypeptides and hence proteins. This is accomplished by chemical 'ligation' reactions that are capable of selective condensation of unprotected peptide segments in solution and, more recently, on solid-phase. In the original incarnation of the chemical ligation concept[33], two unprotected peptides were linked in solution *via* reaction of a C-terminal thioacid on one peptide segment with a N-terminal bromoacetate functionality on the other peptide. This reaction resulted in the union of the two peptides, with the fragments joined by a thioester group in place of the native amide. While this condensation reaction was successfully applied to the synthesis of full-length, active HIV-1 protease[33], the instability of thioesters towards hydrolysis in basic aqueous solution limits its generality. Methods to achieve chemical ligation with retention of a fully natural amide/peptide backbone were thus highly sought after.

In 1994**,** the Kent laboratory reported a remarkable reaction[34], which they termed 'native chemical ligation' (NCL), that built on the previous chemical

ligation concept by enabling the aqueous condensation of unprotected peptides with generation of the native amide bond. In NCL, a C-terminal thioester group on one peptide fragment reacts with an N-terminal cysteine residue on another fragment, ultimately forming the native amide linkage. The power of the approach relies on two chemical recognition events that ensure that only N-terminal cysteine residues form irreversible linkages[35]. A schematic of the reaction is shown in **Figure 3.** In the first step, reversible capture (by trans-thioesterification) of the N-terminal cysteine results in a thioester-linked condensation product. Nucleophilic attack by the N-terminal amine of the cysteine then results in the irreversible formation of the native amide-linked product. This two-step process is responsible for the exquisite chemoselectivity of the NCL reaction – while any free cysteine thiol in either peptide can react in the initial transthioesterification, the reversibility of this initial step coupled with the presence of a free amine only at the N-terminal cysteine, leads to eventual 'capture' of the desired product due to the irreversible formation of the desired amide.

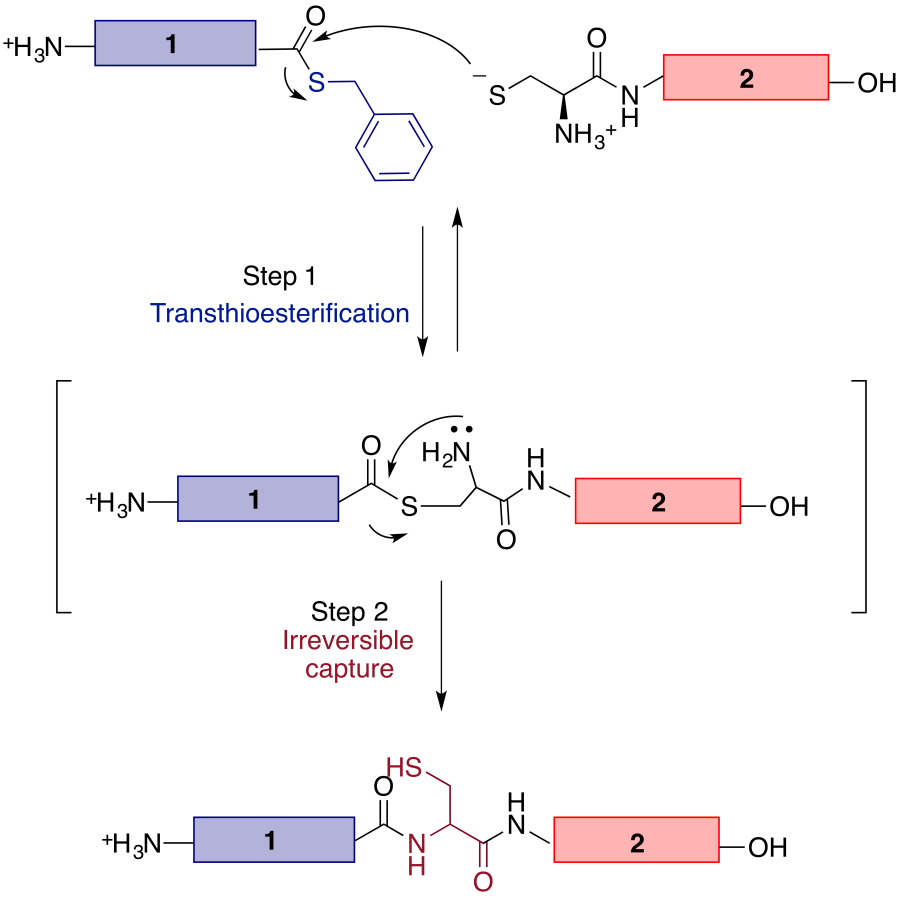


Figure 3: Native chemical ligation

Advances in chemical ligation: extension to non-Cys residues

The requirement for an N-terminal cysteine and C-terminal thioester simultaneously enables and provides the key limitations of the NCL technique. Cysteine is the natural amino acid that is found in lowest abundance in nature, therefore many proteins either lack cysteines or do not possess cysteine residues in convenient locations for disconnection into suitable peptide fragments. The first key advance that demonstrated this limitation could be overcome was the development of techniques for the post-ligation desulfurization of cysteine[36], forming alanine residues at the site of ligation *via* hydrogenation[36] or a radical process[37] (Cys→Ala 'chemical mutation'). Alanine residues are highly abundant in proteins, giving the researcher wide scope in selecting sites for disconnection into peptide fragments for ligation. The ligation- desulfurization approach was initially reported by Yan and Dawson, who employed a metal-mediated (Raney Ni) protocol to effect the conversion of cysteine to alanine[36]. The Danishefsky group then reported[37] a milder, metal-free approach, employing tris(2-carboxyethyl)phosphine (TCEP), the radical initiatior VA-044 and tert-butylthiol (*t*BuSH) to accomplish radical desulfurization. Further optimization of these conditions has enabled the ligation and desulfurization reactions to be employed in a straightforward one-pot procedure**.** Notably, the development of orthogonal cysteine protecting groups has enabled the desulfurization of the 'ligation' cysteine while leaving other cysteine residues in the protein sequence intact[38].

A related strategy to overcome the reliance on cysteine residues for NCL is the use of thiol-containing ligation auxiliaries that can mediate NCL before being chemically removed. These auxiliaries can either be non-native thiol-containing 'appendages'[39-41] or thiolated amino acids[42,43], the latter of which can be subjected to desulfurization to give a native amino acid at the linkage site. Although many thiolated amino acids have been reported as building blocks for solid-phase synthesis, their syntheses tend to be non-

trivial, limiting their use and commercial availability. Methods for the 'direct thiolation' of amino acids and peptides are sorely lacking, with the direct thiolation of tryptophan being a recent exception[44].

An alternative approach to gain the requisite chemoselectivity in radical desulfurization reactions is the use of selenol residues for ligation[45,46]. Selective deselenisation is possible in the presence of cysteine due to the weaker carbon–selenium bond allowing more facile alanyl radical formation[47]. Taking advantage of this fact, the Payne group has recently reported[48] an intriguing alternative to classical NCL-desulfurization that is directly relevant to the 'chemical mutation' concept outlined in this thesis. In this report, an N-terminal selenocysteine (Sec) residue is first used to mediate the ligation of two polypeptides. Following ligation, Sec can be converted ('chemically mutated') to a native serine residue via the intermediacy of the alanyl radical[48]. It was found that addition of oxone to a reaction mixture for radical deselenisation resulted in trapping of the alanyl radical, formed upon C-Se bond fragmentation, forming serine[48]. This reaction thus offers a facile method for programmed peptide ligations at Ser junctions, extending the scope of NCL methods significantly.

The second key limitation of the NCL method is the requirement for a C-terminal peptide thioester. Direct synthesis of peptide thioesters via standard Fmoc chemistry is difficult due to competing base-mediated hydrolysis during deprotection cycles[10]. Therefore, early reports of NCL strictly employed Boc-SPPS protocols for synthesis of C-terminal thioesters[34]. While a robust method for peptide synthesis, Boc-based protocols employ highly toxic and dangerous liquid HF for cleavage of the peptide resin and removal of side-chain protecting groups. Very few laboratories are able to maintain the special facilities required for the safe handling of liquid HF; as a consequence the application of Boc strategies has been limited to a small number of research groups. To overcome this reliance on Boc-SPPS for thioester synthesis, a number of 'crypto-thioester' precursors, that are compatible with Fmoc-SPPS protocols, have been

developed. Notable advances in this area include the use of *C*-terminal hydrazides[49,50] and *N*-acylurea 'Dawson linker'[51] strategies, which rely on in-solution conversion to peptide thioesters following resin cleavage, thus avoiding exposure of the thioester to hydrolytic conditions. These and other methods[52-54] have begun to overcome the 'thioester synthesis' challenge to the widespread employment of NCL methods.

Expressed protein ligation

The above limitations in synthetic access to the requisite *C*-terminal peptide thioesters for NCL have also motivated the development of an alternate strategy, termed expressed protein ligation (EPL)[55,56], that substantially expands the scope of ligation methods. In EPL, the thioester fragment for peptide coupling is generated by recombinant expression in *E. coli* and subsequent purification from cell lysate. The EPL technique derives from study of natural protein-splicing systems[57], in which two flanking domains, termed exteins, are spliced together intra-molecularly via the extrusion of the intervening peptide domain (intein). The mechanism of protein trans-splicing (**Figure 4**) resembles that of NCL, with an initial *N*-*S* acyl shift at the intein N-terminus followed by trans-thioesterification of a cysteine residue at the opposing extein, followed by an *S*-*N* acyl shift mediated by succinimide formation at the intein C-terminus, releasing the spliced extein fragments.

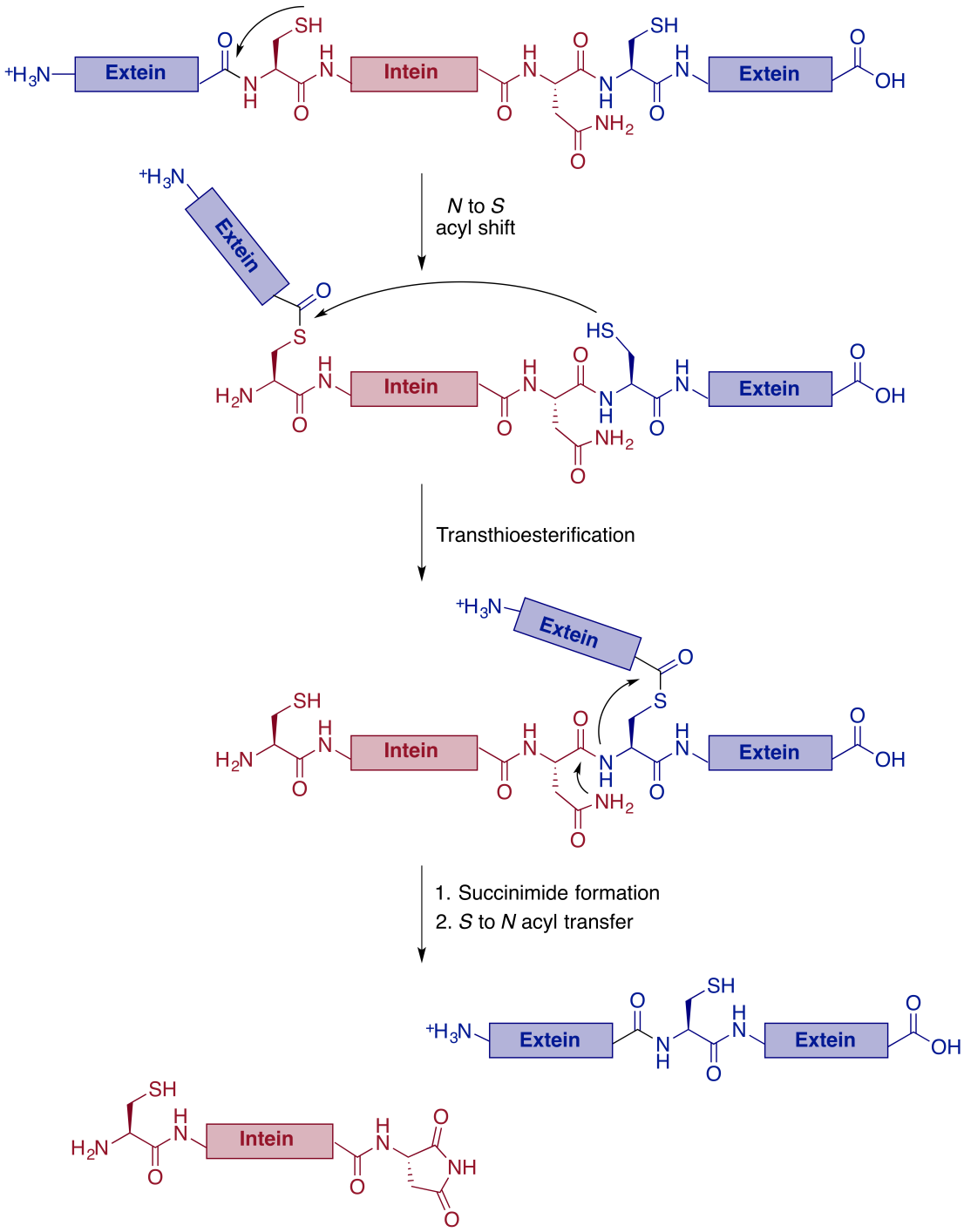


Figure 4: Mechanism of protein *trans*-splicing[57]

To apply this natural process to *in vitro* protein ligation (**Figure 5**), mutation of the intein C-terminal asparagine to alanine prevents succinimide formation and thus the key splicing *S*-*N* acyl shift[56]. This allows the interception of the initially formed thioester linkage via cleavage with free thiols to give a free C-terminal thioester. Recombinant attachment of various affinity tags to the intein C-terminus allows the thioester-bridged

extein-intein fragment to be purified from the cell lysate before thiol cleavage. The free recombinant C-terminal thioester can then be reacted with a N-terminal cysteine containing peptide fragment, as in NCL, to give the final protein product.

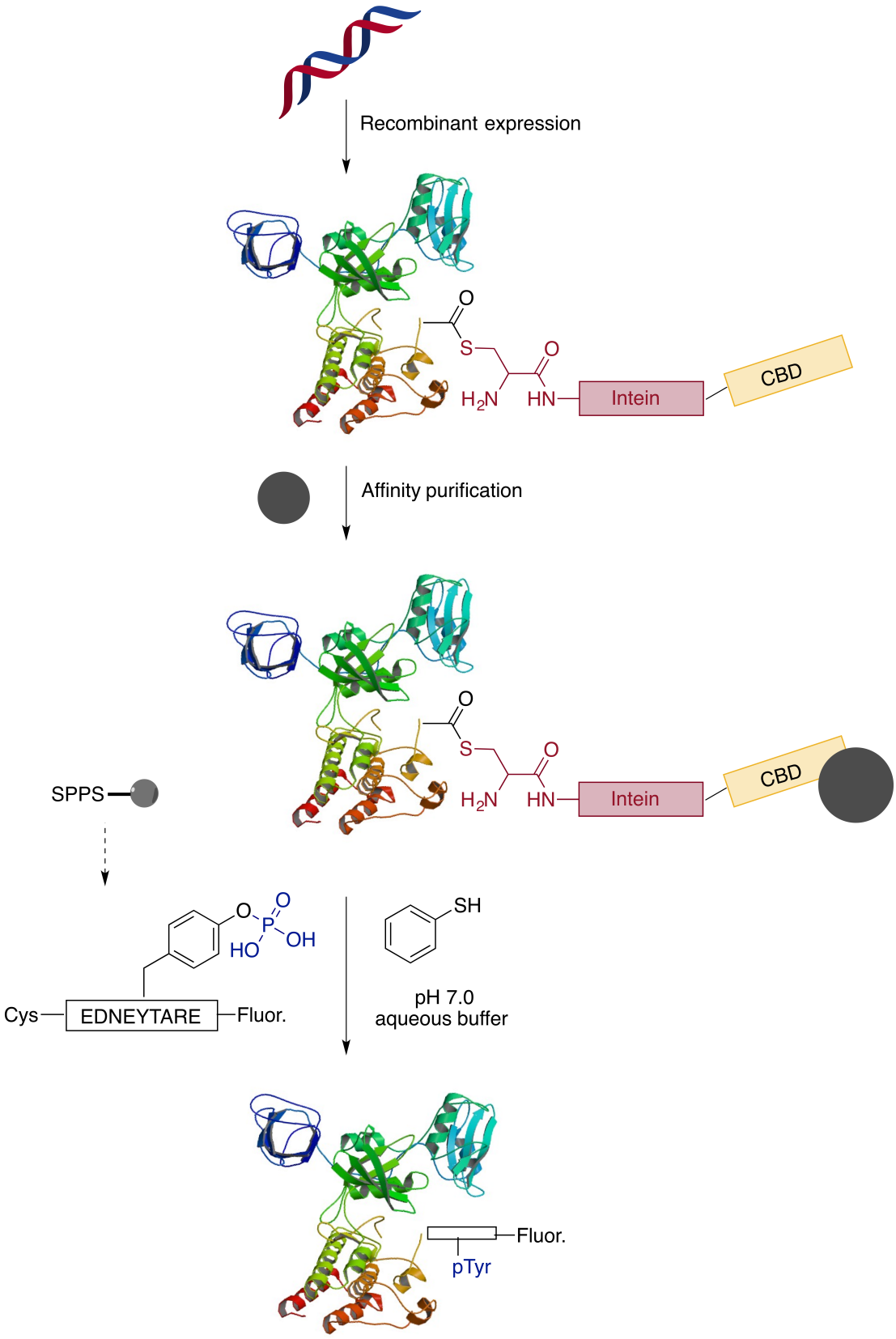


Figure 5: The first report of expressed protein ligation[55]. An unnatural variant of Csk with pTyr and a fluorophore conjugated to the *C*-terminus was assembled *via* the ligation of a recombinant and a synthetic fragment. CBD= chitin binding domain.

The key advantage of EPL relative to purely chemical ligation techniques is that, as the N-terminal fragment is produced recombinantly, it can be of any size, thus circumventing the limitation of SPPS in producing long peptide fragments. This advantage has seen EPL widely applied in protein chemistry, particularly to label proteins *in vitro* and even *in cellulo*[58], and to install defined post-translational modifications for biochemical studies. Notable examples of the use of EPL for the biochemical study of protein PTMs are highlighted in **Chapter 5: Methylation** and **Chapter 6: Phosphorylation.**

Biological methods for the incorporation of post-translationally modified and unnatural amino acids into proteins

The widespread use of recombinant DNA techniques to express heterologous proteins in *E. coli* has motivated efforts to incorporate post-translationally modified residues and unnatural chemical labels directly during biosynthesis. Two such methods have found widespread use in chemical biology research: the use of auxotrophic strains to effect *global* replacement of a given amino acid with a chemically similar analogue, and the reprogramming of the genetic code to allow unnatural amino acid incorporation at the ribosome in response to reassigned stop or, more recently, sense codons. These powerful methods have together allowed the incorporation of well over 100 different amino acids into recombinant proteins[59], that have found varied use in protein chemistry and biology.

Global codon replacement via the use of auxotrophic bacteria

Auxotrophic strains of *E. coli* are unable to synthesise a given amino acid, and thus must be provided with an external source of the amino acid to survive[60]. Under these conditions of stress, the bacteria can be forced to incorporate an analogue of the amino acid of interest by simply adding the analogue amino acid to the growth media during microbial culture. Analogues of methionine[61-64] are the most widely used for auxotrophic incorporation due to the relaxed substrate specificity of the methionyl-aminoacyl-tRNA synthetase. Methionine analogues containing alkene[65], azide[66] and alkyne[65] groups have been extensively used for subsequent bio-orthogonal chemical modification for protein labeling and enrichment. One particularly useful application of the technique is the temporal labeling of proteins by a 'pulse-chase' type experiment, in which a 'pulse' of a chemically reactive methionine surrogate is added to the growth media, followed by a 'chase' with regular methionine[67]. This technique thus only labels newly-synthesised proteins, a fact that has been taken advantage of

in order to study the dynamics of translation and protein turnover, for example in the context of histone incorporation within nucleosomes[68].

Recent developments in this area include the use of mutant aminoacyl tRNA synthetases[69,70] (aaRSs) with increased selectivity for the non-canonical amino acid, which can be coupled to cell-type specific promoters for cell-specific labeling *in vivo*, for example, for proteomic characterization of cell-specific expression. In addition to methionine analogues, phenylalanine analogues can also be incorporated through evolution of an appropriate phenylalanyl- aminoacyl-tRNA synthetase towards relaxed substrate specificity. In this manner, aryl azide[71], aryl halide[72] and aryl ketone[73] functionalities have been incorporated into proteins for subsequent chemical labeling.

Auxotrophic incorporation implies global replacement of a natural residue with the amino acid analogue, and thus relies on efficient depletion of the natural amino acid, which can be tedious in practice and lead to inconsistent yields and incorporation frequencies. Furthermore, the amino acid analogue will be incorporated at all codons encoding the natural residue hence unnatural amino acid incorporation is *residue*- but not *site*-specific[59]. In order to achieve site-specific non-standard amino acid incorporation into proteins via genetic means, a more sophisticated technique is required.

Site-specific amino acid incorporation via stop codon suppression

Stop codon suppression, pioneered by the Schultz laboratory[74-76], allows protein and site-specific incorporation of non-canonical amino acids. In this strategy, a redundant stop codon (usually the amber stop codon), is 'reassigned' by directed evolution of an orthogonal tRNA / aminoacyl-tRNA synthetase pair specific for the non-canonical amino acid of interest. The amino acid can then be incorporated into the protein target by simple site-directed mutagenesis to install an amber codon at the desired position. Co-expression of the protein with the mutant tRNA and synthetase results in

ribosome-programmed incorporation in response to the amber codon, when the amino acid of interest is added to the growth medium (**Figure 6**). Although pioneered in *E. coli*, the technique has since been extended to a range of bacterial and eukaryotic systems, including whole organisms such as *C. elegans*[77] and *A. thaliana*[78].

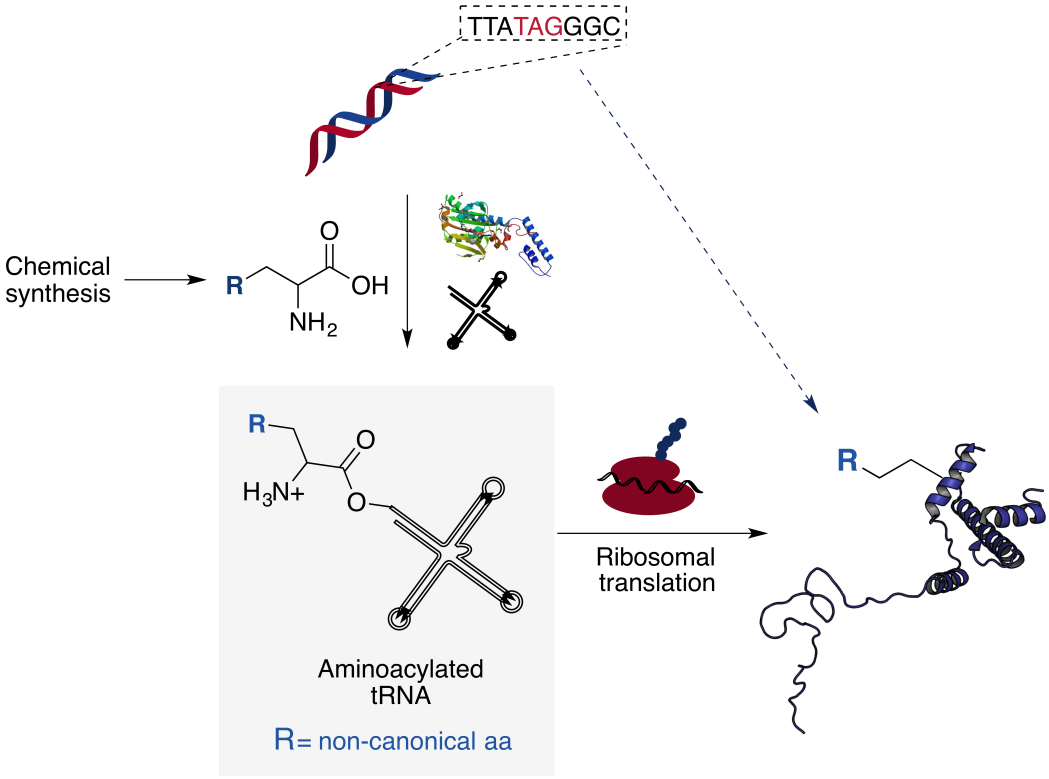


Figure 6: Stop-codon suppression for the genetic incorporation of non-canonical amino acids (ncAAs). ncAAs are coded for by stop-codons, which program incorporation of the ncAA by a mutant aminoacyl-tRNA synthetase.

Earlier efforts at amber codon suppression relied on the chemical aminoacylation of tRNAs and injection of the acylated product into cells[74]. The discovery of the tyrosyl tRNA/ tyrosyl-tRNA synthetase pair from the archaea *Methanococcus jannaschii* provided the first system for direct incorporation *in cellulo*[75]. The unique features of the *M. jannaschii* system include an acceptor loop in the tRNA distinct from that found in *E. coli* and the absence of an anticodon-binding pocket in the synthetase, allowing mutation of the anti-codon loop to recognize a stop codon[11]. These factors together enabled the evolution of tyrosyl tRNA/ tyrosyl-tRNA synthetase

pairs orthogonal to those found in *E. coli*, by selections for synthetases that charge the mutated stop-codon specific tRNA with the non-canonical amino acid of interest, but do not charge the natural tyrosyl-tRNA already present in the cell. Counter-selection against the natural amino acid results in a high-fidelity *orthogonal* tRNA-synthetase pair that allows incorporation of the desired amino acid, but not tyrosine, in response to the amber codon. A number of unnatural aromatic-containing amino acids have been incorporated into recombinant proteins using such evolved tyrosyl tRNA / tRNA synthetase pairs, including useful reactive handles such as azides[79]/alkynes[80], ketones[81] and halides[82].

The discovery of the *Methanosarcina barkeri* pyrrolysyl-tRNA/ pyrrolysyl-tRNA synthetase pair[83-85], which naturally charges an amber codon in this organism with the '22nd amino acid', pyrrolysine, greatly expanded the scope of the amber codon suppression method. Indeed, the unusually plastic substrate-binding pocket of the pyrrolysyl-tRNA synthetase[86-88] enables facile incorporation of many pyrrolysine analogues using only the wild-type protein, termed polyspecificity. A number of unnatural amino acids have been incorporated, with the key recognition feature being the presence of an amide or carbamate linkage at the γ-amino group of lysine[89]. Derivatives with short-chain azide[90] and alkyne[90] functionalities for copper-catalysed 'click' conjugation were amongst the first unnatural amino acids to be incorporated. More structurally divergent analogues of pyrollysine, such as strained alkynes[91] for copper-free click chemistry have since been incorporated, demonstrating the surprising plasticity of the active site of this synthetase. Taking advantage of the recognition of N^{ε}-amide groups by the pyrrolysyl-tRNA synthetase active sites, a number of groups have reported genetic encoding of acyl-lysine derivatives, now recognized as an important class of naturally occurring PTMs. The Chin group first reported the directed evolution of a mutant synthetase for site-specific incorporation of acetyllysine[92] into recombinant histones produced in *E. coli*, which were then used in a number of assays to elucidate the mechanistic basis for the biological function of this modification[93]. In a

recent advance, the same group has demonstrated the stable integration of the related *Methanosarcina mazei* pyrrolysyl-tRNA synthetase into mammalian cell lines[94], enabling homogeneous production of acetyllysine-incorporating proteins in these challenging systems. Further, a number of groups have recently demonstrated genetic incorporation of a range of newly discovered acyl modifications of lysine, including propionylation[95], butyrylation[95], octenoylation[96] and crotonylation[95], using both wild-type and evolved pyrrolysyl-tRNA synthetases (**Figure 7**). This ability to produce proteins with these newly discovered modifications is already yielding dividends in terms of biochemical understanding.

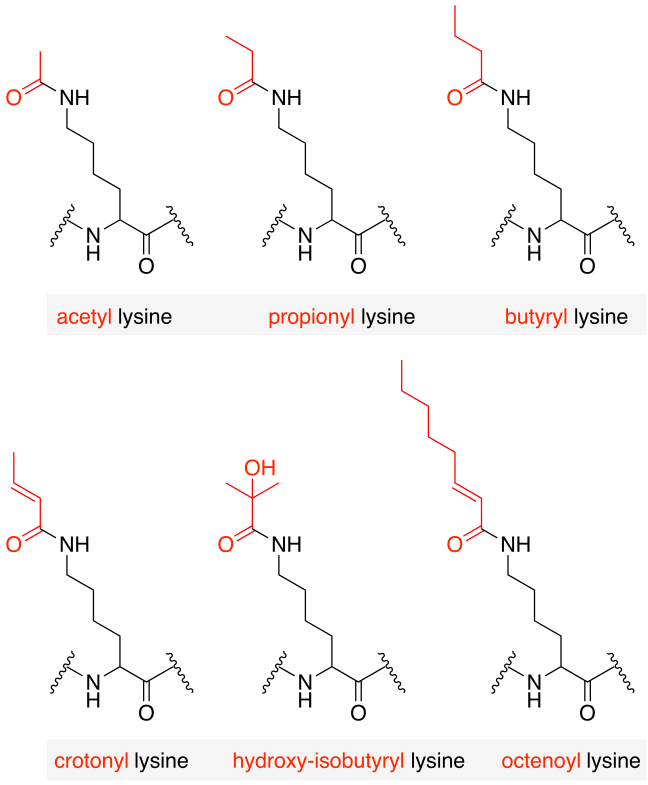


Figure 7: Acylated PTM derivatives of lysine incorporated into proteins using stop-codon suppression.

While the number of non-canonical amino acids incorporated by stop-codon suppression is impressive, and growing, shortcomings of the system still remain. One such shortcoming is the lack of structural diversity beyond the derivatives of lysine or tyrosine described above, reflecting the original propensities of the synthetases that have been selected for use and evolution. Sorely lacking are synthetases for direct incorporation of

methylated lysine derivatives, any derivatives of arginine and glutamine (which are increasingly recognized as key PTMs in biology) or, notably, for any glycosylated amino acids[97-99]. These absences likely reflect the difficulties inherent in engineering a synthetase capable of discriminating amino acids with very small modifications such as methylation, or of charging very large appendages in the case of glycoproteins. The other chief limitation of codon suppression techniques lies in the variability of expression yields, which depend on a large number of interconnected and sometimes conflicting factors[100,101]. Typically, expression yields of proteins containing non-canonical amino acids demonstrate a strong dependence on the particular amino acid incorporated, the synthetase system used for incorporation and the protein sequence itself. It must be noted that the above limitations are not insurmountable, and given the ease of use of the technology and the unlimited applications, further research in the area is likely to further expand access to PTMs and unnatural amino acids with profound implications for protein engineering.

Hybrid approaches – the 'tag and modify' strategy

A third route to modified proteins, the 'tag and modify'[12] approach, attempts to combine the strengths of chemical and biological approaches for the production of proteins with defined non-canonical amino acids. In a two-step strategy, the genetic/biosynthetic introduction of a uniquely reactive 'tag' residue is followed by its *in-vitro* chemical modification (**Figure 8**)[12]. The 'tag' may be a reactive canonical amino acid such as cysteine or an unnatural residue introduced by auxotrophic incorporation or stop-codon engineering, described above. The use of chemoselective chemistry specific for the introduced tag avoids the production of the heterogenous mixtures that can plague attempts at protein modification at canonical residues. One key aspect allowed by this approach is the divergent production of multiple protein modifications from a single 'tag' and hence from a single expression of recombinant protein, allowing ready generation of analogues for functional assays. For convenience, the below discussion of various 'tag and modify' approaches is organised according to the 'tag' residue employed. As previously, the focus is on reactions that generate natural PTMs or their analogues, and not on the introduction of various unnatural affinity probes or fluorescent labels for localisation and tracking studies.

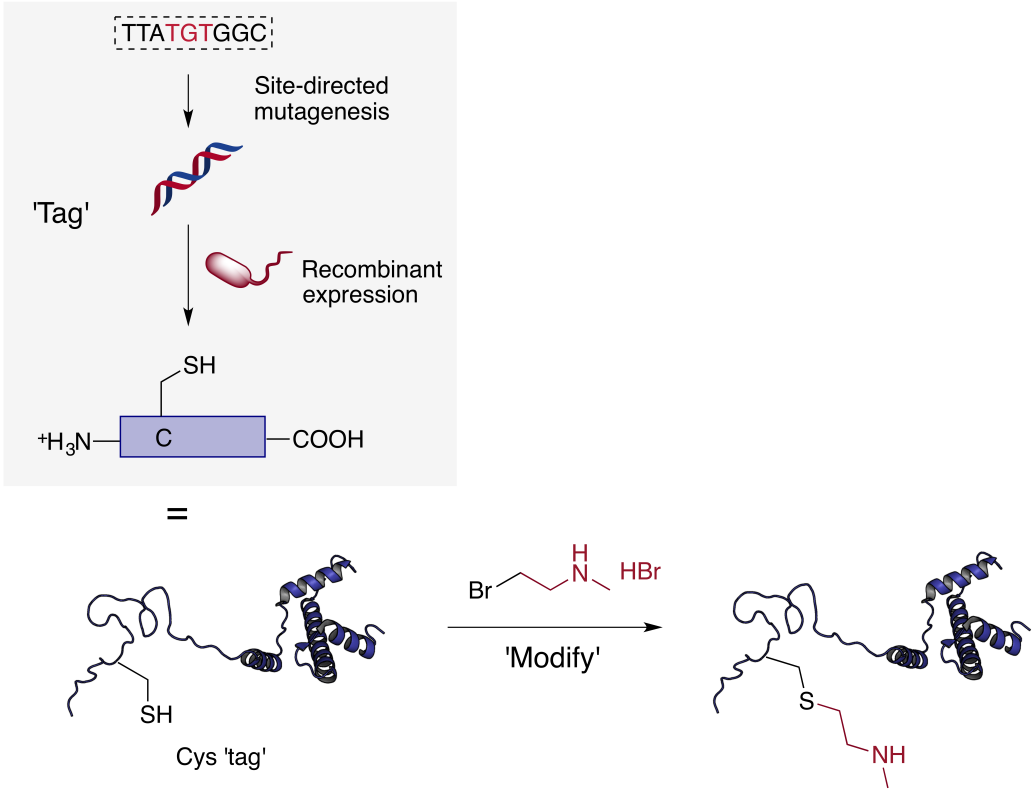


Figure 8: The 'tag-and-modify' approach to selective protein modification, illustrated by 'Shokat'[102] alkylation of cysteine to give methyl-lysine analogues (MLAs)

Modification of azides and alkynes

Azides and alkynes are minimalist 'tags' that have been incorporated into proteins via both auxotrophic incorporation and stop-codon engineering. Due to the ease of incorporation and largely 'bio-orthogonal' nature of these functional groups, a large number of reactions and conditions have been developed for their modification both *in-vitro* and in a cellular context. The Staudinger ligation between azides and phosphines and the [3+2]-cycloaddition ('click' reaction) between azides and alkynes are the most common approaches employed.

The classical Staudinger ligation was first introduced into chemical biology by the Bertozzi laboratory, where it was used to modify metabolically incorporated sugar azides for detection of cellular glycoproteins[103]. The development of techniques for incorporating azide-containing amino acids enabled appplication of this reaction to the modification of protein side chains, at azidohomoalanine (Aha)[66] and at azidophenylalanine. Notably, *in-*

vitro methods for glycosylation[104] and phosphorylation[105] of proteins using the Staudinger ligation have been reported, although an incompatibility of the phosphine reagents required with cysteine residues has been noted[104].

The [3+2]-cycloaddition of azides and alkynes, known to organic chemistry since the time of Huisgen[106], found widespread application in protein modification following the reports of Sharpless[107] and Meldal[108] of a mild copper-catalysed procedure that has become synonymous with the term 'click chemistry' (**Figure 9**). The rapid kinetics, selectivity and 'orthogonality' towards biological functionality has enabled application of the reaction for the stoichiometric labeling of proteins[109], *in-vivo* imaging[110], proteomics[111] and myriad other biological applications. The toxicity of copper in a cellular context[112] has been overcome through the use of ligand engineering[110], encapsulation[113] and replacement by copper-free variants utilizing strained alkynes[114], such as cycloocytnes, which are now favoured for *in-vivo* and cellular applications[115].

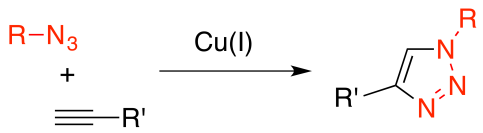


Figure 9: The copper-catalysed azide-alkyne cycloaddition reaction (CuAAC).

The Davis group first reported the use of CuAAC (copper-catalysed, azide-alkyne cycloaddition) for the introduction of defined mimics of glycosylated PTMs in 2007[116]. Using the auxotrophic suppression method, alkyne- (homopropargylglycine, Hpg) and azide- (azidohomoalanine, Aha) containing amino acids were incorporated into proteins as surrogates of methionine. Using ligand-stabilised Cu(I), copper-catalysed Huisgen cycloaddition could be performed on the tagged proteins, employing sugar donors to incorporate glycosylated structures. Notably, the CuAAC reaction could be combined with disulfide-directed modification to enable installation of two different PTMs to a single protein scaffold.

Modification of aryl halides

As mentioned above, the incorporation of iodo- and bromo- analogues of tyrosine has been accomplished by both amber codon suppression and auxotrophic incorporation. These aryl halide 'tags' have been employed in cross-coupling reactions on intact proteins; Sonogoshira, Heck and Suzuki cross-coupling reactions have all been performed. Kodama et al.[117,118] first reported the modification of p-iodophenylalanine (pIPhe) via Heck and Sonogoshira reactions, although with very poor yields. A breakthrough in the application of Pd-mediated cross-coupling chemistry to proteins occurred in 2009[119], with the report by Chalker et al. of quantitative protein modification, including glycosylation, via Suzuki-Miyaura coupling at a cysteine-linked aryl iodide. Key to the success of the Suzuki chemistry was the discovery of an air and moisture-stable ligand for palladium(II). It was later shown that this Pd/ligand combination also enabled reaction at genetically encoded pIPhe residues[120], even in the context of living *E. coli* cells[121].

Modification of ketones and aldehydes

Ketone and aldehyde containing proteins have been used for the selective chemical modification of proteins since the 1950s[122]. Aldehydes and ketones have been introduced into proteins by a number of chemical and biological routes. Indeed, formylglycine is a naturally occuring post-translational modification that arises in nature from enzymatic processing of cysteine or serine at the active site of type I sulfatases[123]. The Bertozzi laboratory has shown that conjugation of a short recognition sequence for the formylglycine-generating enzyme (FGE) can result in efficient enzymatic generation of formylglycine in a target protein[124]. Careful oxidation of N-terminal serine residues using periodate is an alternative, strictly chemical approach to aldehyde incorporation into proteins[125]. Ketones have also been incorporated into proteins by auxotrophic incorporation[73] and stop-codon engineering[81].

Aldehydes and ketones possess electrophilic reactivity orthogonal to that of the canonical protein side-chains and can be functionalised with classical bioconjugation techniques such as the formation of hydrazide and oxime linkages by reaction of tagged proteins with hydrazides[125,126] and hydroxylamines[126,127], respectively. Notably, these linkages are inherently unstable to hydrolysis, limiting the applications of such bioconjugates. The reaction of protein ketones and aldehydes with hydroxylamines has been used to generate glycoprotein mimics[127-129]. The slow kinetics of these reactions can be overcome somewhat through the use of small-molecule aniline catalysts[130-132]. Application of Pictet-Spengler type chemistry by the Bertozzi group has extended the bond-formation strategies available for functionalisation at formylglycine to include C-C bond-forming processes, producing more hydrolytically stable conjugates[133]. However, this work has not yet been directed towards the installation of naturally occuring side-chains.

Modification of natural tags: cysteine

The low natural abundance and relative nucleophilicity of cysteine have made this residue a favoured target for chemical modification[134]. As many proteins contain no or few essential cysteines, it is often possible to genetically encode a single reactive cysteine as a unique 'tag' residue. Alkylation of cysteine with halogenated electrophiles is a classic strategy in protein modification[135], with iodoacetamide 'tagging' still being used routinely in proteolytic peptide mapping to 'cap' cysteine residues and prevent disulfide formation.[136] Aminoethylation[137] of cysteine using aminoethyl halides has been used to install lysine and its methylated derivatives by simple alkylation. This approach was originally developed to install a thioether-mimic of lysine, 'thialysine', in order to generate additional sites of trypsin proteolytic cleavage for peptide mapping purposes[138] (**Figure 10**). Extension to 'chemical rescue' studies of catalytic sites in enzymes[139,140] and, more recently, to the study of lysine methylation in histone proteins[102], demonstrate the versatility of this simple alkylation reaction.

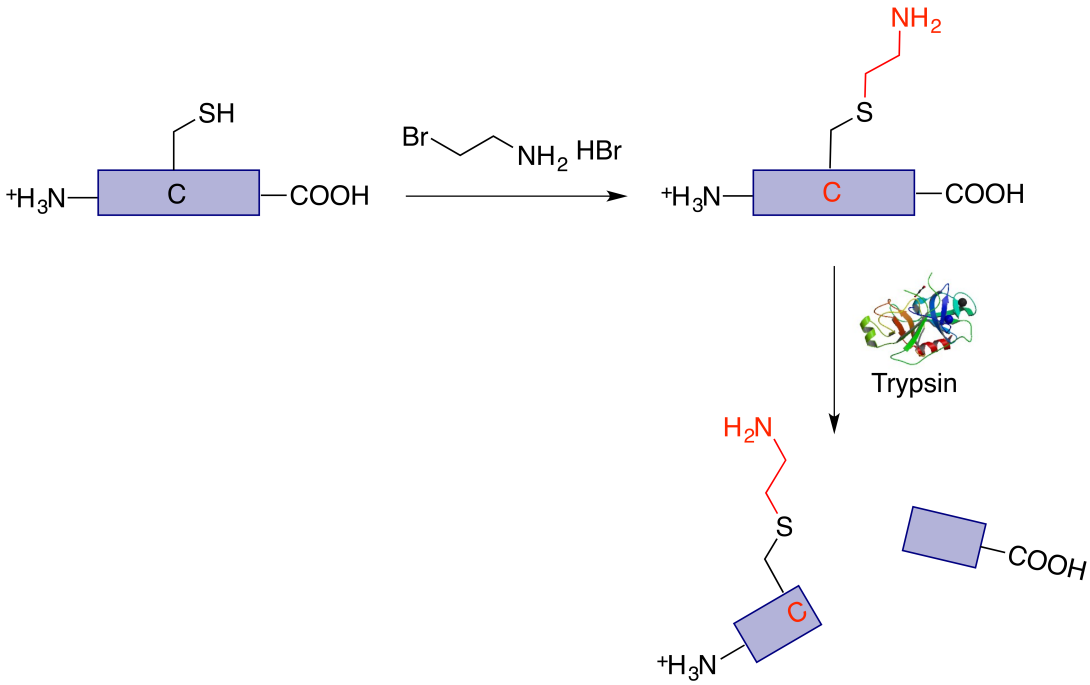


Figure 10: Alkylation of cysteine with aminoethyl halides produces 'thia-lysine' derivatives that can be used as substrates for trypsin-cleavage.

Cysteine can also be modified by conjugate addition of the thiolate to α,β-unsaturated electrophiles in a thia-Michael reaction. Maleimide reagents have been employed for many years to label protein cysteine residues, for example in the generation of antibody-drug conjugates[141-143]. In the context of post-translational modifications, the Waldmann group has demonstrated the use of functionalized maleimide lipopeptide reagents for the incorporation of natural lipid chains, such as polyprenyl, into Ras proteins key to a number of signal transduction events[144,145]. More recently, the use of α,β-unsaturated amidine reagents has been reported for installation of thioether mimics of arginine PTMs[146]. Methylarginine mimetics were installed into histone proteins and could be recognized by the cognate antibodies for the native modifications and also by a natural reader of methylarginine.

One disadvantage inherent to approaches dependent on the nucleophilicity of cysteine is the possibility of side-reaction at other nucleophilic sites of the protein, such as lysine residues[102]. Although selectivity is usually possible through careful control of reaction pH, this may not be feasible in

every case. An alternative approach is to generate a transient electrophilic group at cysteine by generation of thiyl radicals[147]. The Dondoni laboratory first reported the trapping of cysteine-derived thiyl radicals with terminal alkenes, termed the 'thiol-ene' (or thiyl-ene) reaction[148]. Defined glycosylation of bovine serum albumin was possible using an alkenyl-galactoside and a photoinitiator under UV light. Shortly after, the Davis group reported the 'inverse' strategy – reaction of a protein-bound homoallylglycine (Hag) residue, installed as a tag by auxotrophic incorporation, with glyocsyl thiols afforded the desired protein glycoconjugates, again using photoinitiation under UV light[149].

An alternative approach towards cysteine modification is the use of redox chemistry to form mixed disulfides[150] (or selenides), which can then be used directly or as substrates for further reactions. The use of Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid, DTNB)[151-153] to quantify the abundance and accessibility of protein thiols relies on such mixed disulfide formation. The Davis group has explored the formation of mixed disulfides at 'tag' cysteine residues as a means to generate defined glycoprotein constructs[154]. Simple mixture of a external thiol and a cysteine-tagged protein under air will form mixed disulfides by air oxidation. However, long reaction times and formation of symmetrical disulfide side-products (such as unreactive protein dimers) motivated exploration of alternative methods. Glycosyl methanethiosulfonates[155], phenylthiosulfonates[156] and phenylselenylsulfides[157] were developed as activated disulfide exchange reagents with greatly improved reaction rates. The ease of disulfide-linked glycoprotein synthesis using these reagents, particularly a low reagent:protein ratio, allowed the synthesis of complex glycoconjugates, including polyvalent structures[158,159] and the attachment of polysaccharides of varied chain length and branching pattern. Glycosylated proteins were shown to be active in enzymatic extension[157] and, despite the unnatural linkage, were also able to bind their cognate lectins[160]. The use of such activated disulfides is not limited to protein glycosylation, with introduction of disulfide mimics of other PTMs also possible[116].

Activation of a protein cysteine thiol is of course, also possible. Reaction of a free protein cysteine with phenylselenyl bromide, forming a selenylsulfide, generates a cysteine electrophile that can react with external thiols to form mixed disulfides[157]. This 'inverse' approach has been used both to glycosylate, proteins and to incorporate a range of other disulfide mimics of natural post-translational modifications, such as cysteine prenylation[161].

Modification at dehydroalanine

The advantages of selectively incorporating electrophilic functionality into proteins have been alluded to above. As there is no truly electrophilic canonical residue, introduction of a electrophilic modification as a 'tag' allows subsequent chemo- and site- selective chemical modification by exploitation of this polarity difference[162]. The α,β-unsaturated residue dehydroalanine (Dha) is a natural post-translational modification found in antibiotic lanthipeptides[163] as a product of the enzymatic dehydration of serine and in thyroglobulin[164] where it results from elimination of iodinated tyrosine during hormone maturation. Synthetically, dehydroalanine has long been recognised as a valuable handle for protein chemical modification[165].

The earliest method for generation of dehydroalanine on a protein substrate was reported by Koshland in 1963[166] (**Figure 11**). The active site serine of chymotrypsin was converted to dehydroalanine by sulfonylation with tosyl chloride followed by basic elimination. Although used in this first study to study the catalytic mechanism of the protease, Koshland recognised the broader potential of dehydroalanine as a tag for chemical modification. Conjugate additions with thiol reagents were demonstrated and later, Koshland proposed broader use of nucleophillic additions to dehydroalanine as a platform for 'chemical mutation'[15], discussed further below.

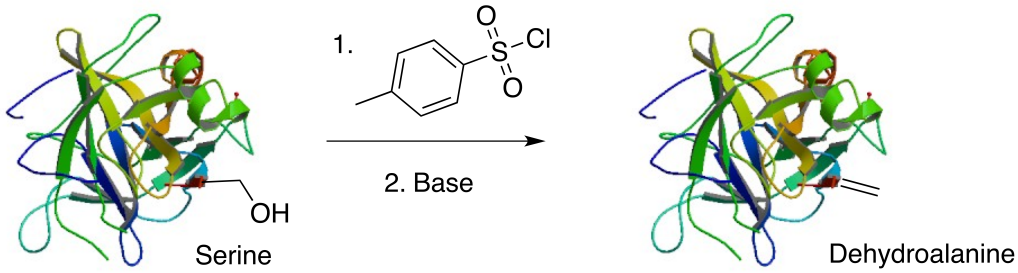


Figure 11: Conversion of the active-site serine to dehydroalanine on chymotrypsin[166].

While Koshland's protocol relied on the enhanced nucleophilicity of the active site serine in serine proteases, more general methods for dehydroalanine formation have since emerged. The nucleophilicity of cysteine allows dehydroalanine formation via a bis-alkylation, elimination route which has been explored and optimised by the Davis[165,167,168] and other laboratories[169] (see the next Chapter). Alternative routes to dehydoalanine formation, employing oxidative elimination of unnatural aryl and alkyl cysteines or phenylselenocysteine[170], have also been reported.

The conjugate addition of thiols to dehydroalanine residues occurs in nature during the formation of lanthionine bridges in lanthipeptide biosynthesis[163]. Taking inspiration from nature, a number of research groups have reported the use of such conjugate additions for introduction of post-translational modifications and their mimics to peptides[171] and proteins. In the laboratory, nucleophilic attack on protein dehydroalanine with small molecule thiols typically results in a mixture of $_{D}/_{L}$ epimers at the α-carbon, although few studies have attempted to provide an accurate assessment of their ratio[168,172].

Reaction with thiophosphate provides access to phosphocysteine[168], a natural PTM that has also been validated as a mimic of phosphorylated threonine in the context of kinase activation[173]. Addition of glycosyl-thiol reagents forms naturally occuring glycosylated cysteines, unusual glycosylated residues that possess increased metabolic stability relative to O-linked glycosylation and remain little studied in biology[174,175]. Recently,

use of a GlcNAc-thiol reagent to glycosylate histone proteins as a mimic of O-GlcNAcylation has been reported[168]. Histones bearing *S*-linked GlcNAc at sites of reported threonine or serine glycosylation, followed by incorporation of the modified histones into nucleosomes, enabled novel biophysical[176] and proteomic insights[177] into the function of the naturally occuring modifications. Notably, the use of native *O*-linked glycoproteins in proteome 'fishing' for binding partners is complicated by the hydrolytic instability of these residues, and their rapid enymatic turnover in cell extracts.

The conjugate addition of alkyl thiols to dehydroalanine is an alternative approach to the incorporation of thioether mimics of lysine PTMs[167,168]. This approach to 'methyl-lysine analogues', MLAs, is discussed further in **Chapter 5.**

'Chemical mutagenesis'

In 1966, the Koshland[15] and Bender[16] groups independently reported the chemical transformation of one specific amino acid side chain to another – the first point mutation of a protein (**Figure 12**). Both groups were motivated by the need for milder methods to probe the function of catalytic residues in enzymes. In both reports, the active site serine of *subtilisin* was chemically converted to a cysteine residue. Firstly, the active site serine was sulfonylated with phenylmethanesulfonyl fluoride, forming an excellent leaving group. Displacement by thioacetate, followed by acetyl hydrolysis through the inherent protease activity of the enzyme, provided thio-subtilisin, bearing a cysteine residue in the active site pocket. The presence of the desired cysteine was confirmed by amino acid analysis and colorimetric titration. To appreciate the importance of this first 'chemical mutation', it should be noted that the first mutation of a protein achieved via DNA-level site-directed mutagenesis was not reported until 1978[178,179].

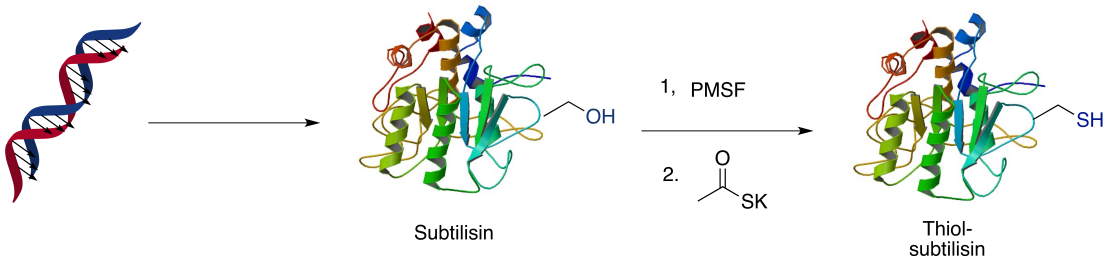


Figure 12: The first point mutation of a protein.

Conceptually, the goal of 'chemical mutation' can be seen as an extension of hybrid biosynthetic/chemical approaches to protein modification, with an emphasis on the inter-conversion of natural amino acids and their modifications[13,14]. The tools of retrosynthesis can be applied to protein modification to understand the unique strategic approach embedded in the concept[14]. Biosynthetic approaches (that ultimately rely on the ribosomal incorporation of non-canonical amino acids into polypeptides) and contemporary chemical ligation techniques are united by the use of amide-bond formation via activated carboxy-derivatives as the key chemical step for incorporation of modified amino acids. In contrast, chemical mutation approaches resemble and extend the 'tag-and-modify' hybrid approach by building side-chain diversity directly at the residue level, on a fully folded, expressed protein. This side-chain focused strategy thus resembles nature's method of extending protein chemical diversity beyond the canonical set of ribosomally-processed amino acids by extensive enzymatic post-translational modification.

One aspect in which the 'tag-and-modify' approach to protein modification extends the early work of Koshland and Bender is the idea of a generic 'tag' residue that, by virtue of unique reactivity, can act as a 'handle' for divergent modification into a number of diverse amino acids[13]. The work of Clark and Lowe on 'chemical mutations' of papain demonstrated for the first time the purely chemical conversion of a single amino acid precursor to a number of different side-chains – a key necessity for a general mutagenesis[180,181]. The reactive cysteine of papain was chemically mutated to either serine or glycine via the intermediate formylglycine. Synthetically,

Clark and Lowe's chemical mutation of papain commenced with selective alkylation of the reactive active site cysteine using phenacyl bromide. Repeated photolysis was necessary to convert the phenacyl-inhibited protein to the desired thioaldehyde product. The thioaldehye slowly hydrolyses, forming formylglycine with the release of H_2S. From formylglycine, reduction with $NaBH_4$ provides serine while incubation at pH 9.0 for prolonged periods leads to glycine *via* dehydroserine.

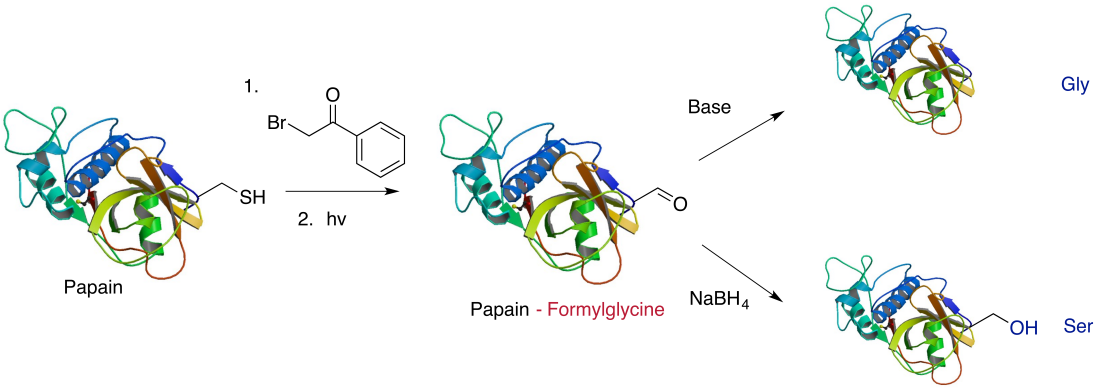


Figure 13: Synthetic divergence in the chemical mutations of papain by Clark and Lowe[180,181]

As in the work of Koshland of Bender, Clark and Lowe relied on the distinct active-site nucleophilicity of a protease active site residue to accomplish their chemical mutations. To achieve a truly general approach to chemical mutation, a 'tag' residue bearing unique reactivity in any protein is necessary. Building on the prescient vision of Koshland, as alluded to previously, we have proposed the use of dehydroalanine as such a general 'tag' residue[13].

Goals of this thesis

The key goal of this thesis is the development of a general method for chemical mutation, which we have termed 'post-translational mutagenesis'[182] to highlight both the strategic similarity to nature's approach of enzymatic protein modification and the goal of providing ready access to proteins bearing defined post-translational modifications. Building on previous work using the reaction of thiol reagents with

protein-bound dehydroalanine to build thioether mimics of PTMs and early synthetic work on modification of dehydroalanine peptides via radical means, we propose that the conjugate addition of *alkyl radicals* to protein dehydroalanine could enable a general approach to 'post-translational mutagenesis'. Addition of alkyl radicals would allow construction of C(sp^3)–C (sp^3) bonds at the β-γ linkage[14], theoretically enabling access to a number of natural amino acids and their PTM derivatives.

References: Introduction

1 Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie International Edition* **48**, 6974-6998, doi:10.1002/anie.200900942 (2009).

2 Sievers, E. L. & Senter, P. D. Antibody-drug conjugates in cancer therapy. *Annual review of medicine* **64**, 15-29 (2013).

3 Jones, L. H. Recent advances in the molecular design of synthetic vaccines. *Nat Chem* **7**, 952-960 (2015).

4 Wenda, S., Illner, S., Mell, A. & Kragl, U. Industrial biotechnology—the future of green chemistry? *Green Chemistry* **13**, 3007-3047 (2011).

5 Yang, H. *et al.* Molecular engineering of industrial enzymes: recent advances and future prospects. *Applied Microbiology and Biotechnology* **98**, 23-29, doi:10.1007/s00253-013-5370-3 (2014).

6 Davis, B. G. Chemical modification of biocatalysts. *Current opinion in Biotechnology* **14**, 379-386 (2003).

7 Walsh, C. T., Garneau‐Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie International Edition* **44**, 7342-7372 (2005).

8 Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* **13**, 343-357 (2012).

9 Shahbazian, M. D. & Grunstein, M. Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* **76**, 75-100 (2007).

10 Kent, S. B. Total chemical synthesis of proteins. *Chemical Society Reviews* **38**, 338-351 (2009).

11 Xie, J. & Schultz, P. G. A chemical toolkit for proteins—an expanded genetic code. *Nature Reviews Molecular Cell Biology* **7**, 775-782 (2006).

12 Chalker, J. M., Bernardes, G. a. J. & Davis, B. G. A "tag-and-modify" approach to site-selective protein modification. *Accounts of chemical research* **44**, 730-741 (2011).

13 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).

14 Wright, T. H., Vallée, M. R. J. & Davis, B. G. From Chemical Mutagenesis to Post‐Expression Mutagenesis: A 50 Year Odyssey. *Angewandte Chemie International Edition* **55**, 5896-5903 (2016).

15 Neet, K. E. & Koshland, D. The conversion of serine at the active site of subtilisin to cysteine: a" chemical mutation". *Proceedings of the National Academy of Sciences* **56**, 1606-1611 (1966).

16 Polgar, L. & Bender, M. L. The reactivity of thiol-subtilisin, an enzyme containing a synthetic functional group. *Biochemistry* **6**, 610-620 (1967).

17 Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nature communications* **5** (2014).

18 Boutureira, O. & Bernardes, G. a. J. Advances in chemical protein modification. (2015).

19 Kunz, H. Emil Fischer—unequalled classicist, master of organic chemistry research, and inspired trailblazer of biological chemistry. *Angewandte Chemie International Edition* **41**, 4439-4451 (2002).

20 Fischer, E. Faraday lecture. Synthetical chemistry in its relation to biology. *Journal of the Chemical Society, Transactions* **91**, 1749-1765 (1907).

21 Vigneaud, V. Experiences in the polypeptide field: insulin to oxytocin. *Annals of the New York Academy of Sciences* **88**, 537-548 (1960).

22 du Vigneaud, V., Ressler, C., Swan, J. M., Roberts, C. W. & Katsoyannis, P. G. The synthesis of oxytocin. *Journal of the American Chemical Society* **76**, 3115-3121 (1954).

23 Vigneaud, V. d. *et al.* The synthesis of an octapeptide amide with the hormonal activity of oxytocin. *Journal of the American Chemical Society* **75**, 4879-4880 (1953).

24 Merrifield, R. B. Solid phase synthesis (Nobel lecture). *Angewandte Chemie International Edition in English* **24**, 799-810 (1985).

25 Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154, doi:10.1021/ja00897a025 (1963).

26 Merrifield, B. The role of the support in solid phase peptide synthesis. *British Polymer Journal* **16**, 173-178, doi:10.1002/pi.4980160404 (1984).

27 Gutte, B. & Merrifield, R. B. Total synthesis of an enzyme with ribonuclease A activity. *Journal of the American Chemical Society* **91**, 501-502 (1969).

28 Anfinsen, C. B., Haber, E., Sela, M. & White, F. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences* **47**, 1309-1314 (1961).

29 Chang, C.-D. & Meienhofer, J. Solid-phase peptide synthesis using mild base cleavage of Nα-fluorenylmethyloxycarbonylamino acids, exemplified by a synthesis of dihydrosomatostatin. *International Journal of Peptide and Protein Research* **11**, 246-249 (1978).

30 Chan, W. C. & White, P. D. *Fmoc solid phase peptide synthesis*. (Oxford University Press, 2000).

31 Pattabiraman, V. R. & Bode, J. W. Rethinking amide bond synthesis. *Nature* **480**, 471-479 (2011).

32 Simon, M. D. *et al.* Rapid Flow‐Based Peptide Synthesis. *ChemBioChem* **15**, 713-720 (2014).

33 Schnolzer, M. & Kent, S. B. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science* **256**, 221 (1992).

34 Dawson, P. E., Miur, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-780 (1994).

35 Johnson, E. C. & Kent, S. B. Insights into the mechanism and catalysis of the native chemical ligation reaction. *Journal of the American Chemical Society* **128**, 6640-6646 (2006).

36 Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *Journal of the American Chemical Society* **123**, 526-533 (2001).

37 Wan, Q. & Danishefsky, S. J. Free‐radical‐based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angewandte Chemie International Edition* **46**, 9248-9252 (2007).

38 Pentelute, B. L. & Kent, S. B. Selective desulfurization of cysteine in the presence of Cys (Acm) in polypeptides obtained by native chemical ligation. *Organic letters* **9**, 687-690 (2007).

39 Brik, A., Yang, Y.-Y., Ficht, S. & Wong, C.-H. Sugar-Assisted Glycopeptide Ligation. *Journal of the American Chemical Society* **128**, 5626-5627, doi:10.1021/ja061165w (2006).

40 Offer, J., Boddy, C. N. C. & Dawson, P. E. Extending Synthetic Access to Proteins with a Removable Acyl Transfer Auxiliary. *Journal of the American Chemical Society* **124**, 4642-4646, doi:10.1021/ja016731w (2002).

41 Wu, B. *et al.* Building Complex Glycopeptides: Development of a Cysteine-Free Native Chemical Ligation Protocol. *Angewandte Chemie International Edition* **45**, 4116-4125, doi:10.1002/anie.200600538 (2006).

42 Wong, C. T., Tung, C. L. & Li, X. Synthetic cysteine surrogates used in native chemical ligation. *Molecular BioSystems* **9**, 826-833 (2013).

43 Malins, L. R. & Payne, R. J. Synthetic amino acids for applications in peptide ligation–desulfurization chemistry. *Australian Journal of Chemistry* **68**, 521-537 (2015).

44 Malins, L. R., Cergol, K. M. & Payne, R. J. Chemoselective sulfenylation and peptide ligation at tryptophan. *Chemical Science* **5**, 260-266 (2014).

45 Malins, L. R., Mitchell, N. J. & Payne, R. J. Peptide ligation chemistry at selenol amino acids. *Journal of Peptide Science* **20**, 64-77 (2014).

46 Hondal, R. J., Nilsson, B. L. & Raines, R. T. Selenocysteine in native chemical ligation and expressed protein ligation. *Journal of the American Chemical Society* **123**, 5140-5141 (2001).

47 Metanis, N., Keinan, E. & Dawson, P. E. Traceless Ligation of Cysteine Peptides Using Selective Deselenization. *Angewandte Chemie International Edition* **49**, 7049-7053, doi:10.1002/anie.201001900 (2010).

48 Malins, L. R., Mitchell, N. J., McGowan, S. & Payne, R. J. Oxidative deselenization of selenocysteine: applications for programmed ligation at serine. *Angewandte Chemie* **127**, 12907-12912 (2015).

49 Fang, G. M., Wang, J. X. & Liu, L. Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angewandte Chemie International Edition* **51**, 10347-10350 (2012).

50 Fang, G. M. *et al.* Protein chemical synthesis by ligation of peptide hydrazides. *Angewandte Chemie International Edition* **50**, 7645-7649 (2011).

51 Blanco‐Canosa, J. B. & Dawson, P. E. An efficient Fmoc‐SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angewandte Chemie International Edition* **47**, 6851-6855 (2008).

52 Tofteng, A. P., Sørensen, K. K., Conde‐Frieboes, K. W., Hoeg‐Jensen, T. & Jensen, K. J. Fmoc Solid‐Phase Synthesis of C‐Terminal Peptide Thioesters by Formation of a Backbone Pyroglutamyl Imide Moiety. *Angewandte Chemie International Edition* **48**, 7411-7414 (2009).

53 Zheng, J.-S., Chang, H.-N., Wang, F.-L. & Liu, L. Fmoc synthesis of peptide thioesters without post-chain-assembly manipulation. *Journal of the American Chemical Society* **133**, 11080-11083 (2011).

54 Hou, W., Zhang, X., Li, F. & Liu, C.-F. Peptidyl N, N-bis (2-mercaptoethyl)-amides as thioester precursors for native chemical ligation. *Organic letters* **13**, 386-389 (2010).

55 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**, 6705-6710 (1998).

56 Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annual review of biochemistry* **72**, 249-289 (2003).

57 Noren, C. J., Wang, J. & Perler, F. B. Dissecting the chemistry of protein splicing and its applications. *Angewandte Chemie International Edition* **39**, 450-466 (2000).

58 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).

59 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).

60 Davis, B. D. & Mingioli, E. S. Mutants of Escherichia coli requiring methionine or vitamin B12. *Journal of bacteriology* **60**, 17 (1950).

61 Duewel, H., Daub, E., Robinson, V. & Honek, J. F. Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein 19F NMR. *Biochemistry* **36**, 3404-3416 (1997).

62 Nediljko Budisa, B. S. & Pascal Demange, C. E. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur. J. Biochem* **230**, 788-796 (1995).

63 Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO journal* **9**, 1665 (1990).

64 Kiick, K. L., van Hest, J. C. M. & Tirrell, D. A. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. *Angewandte Chemie International Edition* **39**, 2148-2152 (2000).

65 van Hest, J. C., Kiick, K. L. & Tirrell, D. A. Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *Journal of the American Chemical Society* **122**, 1282-1288 (2000).

66 Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National academy of Sciences* **99**, 19-24 (2002).

67 Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences* **103**, 9482-9487 (2006).

68 Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161-1164 (2010).

69 Ngo, J. T. *et al.* Cell-selective metabolic labeling of proteins. *Nat Chem Biol* **5**, 715-717 (2009).

70 Ngo, J. T., Schuman, E. M. & Tirrell, D. A. Mutant methionyl-tRNA synthetase from bacteria enables site-selective N-terminal labeling of proteins expressed in mammalian cells. *Proceedings of the National Academy of Sciences* **110**, 4992-4997, doi:10.1073/pnas.1216375110 (2013).

71 Kirshenbaum, K., Carrico, I. S. & Tirrell, D. A. Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues. *ChemBioChem* **3**, 235-237 (2002).

72 Ibba, M. & Hennecke, H. Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows in vitro and in vivo synthesis of proteins containing unnatural amino acids. *FEBS letters* **364**, 272-275 (1995).

73 Datta, D., Wang, P., Carrico, I. S., Mayo, S. L. & Tirrell, D. A. A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *Journal of the American Chemical Society* **124**, 5652-5653 (2002).

74 Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **244**, 182-188 (1989).

75 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **292**, 498-500, doi:10.1126/science.1060077 (2001).

76 Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79**, 413-444 (2010).

77 Greiss, S. & Chin, J. W. Expanding the genetic code of an animal. *Journal of the American Chemical Society* **133**, 14196-14199 (2011).

78 Li, F. *et al.* Expanding the genetic code for photoclick chemistry in E. coli, mammalian cells, and A. thaliana. *Angewandte Chemie International Edition* **52**, 9700-9704 (2013).

79 Chin, J. W. *et al.* Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia c oli. *Journal of the American Chemical Society* **124**, 9026-9027 (2002).

80 Deiters, A. & Schultz, P. G. In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorganic & medicinal chemistry letters* **15**, 1521-1524 (2005).

81 Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences* **100**, 56-61, doi:10.1073/pnas.0234824100 (2003).

82 Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotech* **22**, 1297-1301 (2004).

83 Blight, S. K. *et al.* Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. *Nature* **431**, 333-335 (2004).

84 Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459-1462 (2002).

85 Polycarpo, C. *et al.* An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12450-12454, doi:10.1073/pnas.0405362101 (2004).

86 Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the National Academy of Sciences* **104**, 11268-11273 (2007).

87 Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase–tRNAPyl structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163-1167 (2009).

88 Yanagisawa, T. *et al.* Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *Journal of molecular biology* **378**, 634-652 (2008).

89 Li, W.-T. *et al.* Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *Journal of molecular biology* **385**, 1156-1164 (2009).

90 Nguyen, D. P. *et al.* Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry. *Journal of the American Chemical Society* **131**, 8720-8721 (2009).

91 Kaya, E. *et al.* A Genetically Encoded Norbornene Amino Acid for the Mild and Selective Modification of Proteins in a Copper-Free Click Reaction. *Angewandte Chemie International Edition* **51**, 4466-4469, doi:10.1002/anie.201109252 (2012).

92 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nature chemical biology* **4**, 232-234 (2008).

93 Neumann, H. *et al.* A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Molecular cell* **36**, 153-163 (2009).

94 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* (2016).

95 Wilkins, B. J. *et al.* Genetically encoding lysine modifications on histone H4. *ACS chemical biology* **10**, 939-944 (2015).

96 Wang, W. W., Zeng, Y., Wu, B., Deiters, A. & Liu, W. R. A chemical biology approach to reveal Sirt6-targeted histone H3 sites in nucleosomes. *ACS chemical biology* **11**, 1973-1981 (2016).

97 Zhang, Z. *et al.* A new strategy for the synthesis of glycoproteins. *Science* **303**, 371-373 (2004).

98 Xu, R. *et al.* Site-Specific Incorporation of the Mucin-Type N-Acetylgalactosamine-α-O-threonine into Protein in Escherichia coli. *Journal of the American Chemical Society* **126**, 15654-15655, doi:10.1021/ja044711z (2004).

99 Wang, L.-X. & Davis, B. G. Realizing the promise of chemical glycobiology. *Chemical Science* **4**, 3381-3394 (2013).

100 Short, G. F., Golovine, S. Y. & Hecht, S. M. Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry* **38**, 8808-8819 (1999).

101 Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in E. coli. *Journal of molecular biology* **395**, 361-374 (2010).

102 Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003-1012 (2007).

103 Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007-2010 (2000).

104 Bernardes, G. J., Linderoth, L., Doores, K. J., Boutureira, O. & Davis, B. G. Site‐Selective Traceless Staudinger Ligation for Glycoprotein Synthesis Reveals Scope and Limitations. *ChemBioChem* **12**, 1383-1386 (2011).

105 Serwa, R. *et al.* Chemoselective Staudinger‐Phosphite Reaction of Azides for the Phosphorylation of Proteins. *Angewandte Chemie International Edition* **48**, 8234-8239 (2009).

106 Huisgen, R. 1, 3 - dipolar cycloadditions. Past and future. *Angewandte Chemie International Edition in English* **2**, 565-598 (1963).

107 Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper (I) - Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition* **41**, 2596-2599 (2002).

108 Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase:[1, 2, 3]-triazoles by regiospecific copper (I)-catalyzed 1, 3-dipolar cycloadditions of terminal alkynes to azides. *The Journal of organic chemistry* **67**, 3057-3064 (2002).

109 Wang, Q. *et al.* Bioconjugation by copper (I)-catalyzed azide-alkyne [3+ 2] cycloaddition. *Journal of the American Chemical Society* **125**, 3192-3193 (2003).

110 Del Amo, D. S. *et al.* Biocompatible copper (I) catalysts for in vivo imaging of glycans. *Journal of the American Chemical Society* **132**, 16893 (2010).

111 Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper (i)-catalyzed azide-alkyne [3+ 2] cycloaddition. *Journal of the American Chemical Society* **125**, 4686-4687 (2003).

112 Kennedy, D. C. *et al.* Cellular consequences of copper complexes used to catalyze bioorthogonal click reactions. *Journal of the American Chemical Society* **133**, 17993-18001 (2011).

113 Yamada, Y. M. A., Sarkar, S. M. & Uozumi, Y. Amphiphilic Self-Assembled Polymeric Copper Catalyst to Parts per Million Levels: Click Chemistry. *Journal of the American Chemical Society* **134**, 9285-9290, doi:10.1021/ja3036543 (2012).

114 Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3+ 2] azide– alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society* **126**, 15046-15047 (2004).

115 Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences* **104**, 16793-16797 (2007).

116 van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446**, 1105-1109 (2007).

117 Kodama, K. *et al.* Regioselective carbon–carbon bond formation in proteins with palladium catalysis; new protein chemistry by organometallic chemistry. *ChemBioChem* **7**, 134-139 (2006).

118 Kodama, K. *et al.* Site‐Specific Functionalization of Proteins by Organopalladium Reactions. *ChemBioChem* **8**, 232-238 (2007).

119 Chalker, J. M., Wood, C. S. & Davis, B. G. A convenient catalyst for aqueous and protein suzuki– miyaura cross-coupling. *Journal of the American Chemical Society* **131**, 16346-16347 (2009).

120 Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki–Miyaura protein modification at genetically encoded aryl halides. *Chemical Communications* **47**, 1698-1700 (2011).

121 Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-mediated cell-surface labeling. *Journal of the American Chemical Society* **134**, 800-803 (2011).

122 Barrnett, R. J. & Seligman, A. M. Histochemical demonstration of protein-bound alpha-acylamido carboxyl groups. *The Journal of biophysical and biochemical cytology* **4**, 169-176 (1958).

123 Appel, M. J. & Bertozzi, C. R. Formylglycine, a post-translationally generated residue with unique catalytic capabilities and biotechnology applications. *ACS chemical biology* **10**, 72 (2015).

124 Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat Chem Biol* **3**, 321-322 (2007).

125 Geoghegan, K. F. & Stroh, J. G. Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjugate chemistry* **3**, 138-146 (1992).

126 Vilaseca, L. A. *et al.* Protein conjugates of defined structure: synthesis and use of a new carrier molecule. *Bioconjugate chemistry* **4**, 515-520 (1993).

127 Cervigni, S. E., Dumy, P. & Mutter, M. Synthesis of Glycopeptides and Lipopeptides by Chemoselective Ligation. *Angewandte Chemie International Edition in English* **35**, 1230-1232, doi:10.1002/anie.199612301 (1996).

128 Lees, A., Sen, G. & LopezAcosta, A. Versatile and efficient synthesis of protein–polysaccharide conjugate vaccines using aminooxy reagents and oxime chemistry. *Vaccine* **24**, 716-729 (2006).

129 Liu, H., Wang, L., Brock, A., Wong, C.-H. & Schultz, P. G. A Method for the Generation of Glycoprotein Mimetics. *Journal of the American Chemical Society* **125**, 1702-1703, doi:10.1021/ja029433n (2003).

130 Lempens, E. H., Helms, B. A., Merkx, M. & Meijer, E. Efficient and chemoselective surface immobilization of proteins by using aniline‐catalyzed oxime chemistry. *ChemBioChem* **10**, 658-662 (2009).

131 Wendeler, M., Grinberg, L., Wang, X., Dawson, P. E. & Baca, M. Enhanced catalysis of oxime-based bioconjugations by substituted anilines. *Bioconjugate chemistry* **25**, 93-101 (2013).

132 Dirksen, A., Hackeng, T. M. & Dawson, P. E. Nucleophilic catalysis of oxime ligation. *Angewandte Chemie International Edition* **45**, 7581-7584 (2006).

133 Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D. & Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proceedings of the National Academy of Sciences* **110**, 46-51 (2013).

134 Fraenkel-Conrat, H. [11] Methods for investigating the essential groups for enzyme activity. *Methods in enzymology* **4**, 247-269 (1957).

135 Sela, M., White, F. H. & Anfinsen, C. The reductive cleavage of disulfide bonds and its application to problems of protein structure. *Biochimica et biophysica acta* **31**, 417-426 (1959).

136 Sechi, S. & Chait, B. T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Analytical Chemistry* **70**, 5150-5158 (1998).

137 Raftery, M. & Cole, R. D. On the aminoethylation of proteins. *Journal of Biological Chemistry* **241**, 3457-3459 (1966).

138 Cole, R. D. [33] S-Aminoethylation. *Methods in Enzymology* **11**, 315-317 (1967).

139 Lorimer, G. H., Chen, Y. R. & Hartman, F. C. A role for the epsilon-amino group of lysine-334 of ribulose-1, 5-bisphosphate carboxylase in the addition of carbon dioxide to the 2, 3-enediol (ate) of ribulose 1, 5-bisphosphate. *Biochemistry* **32**, 9018-9024 (1993).

140 Highbarger, L. A., Gerlt, J. A. & Kenyon, G. L. Mechanism of the Reaction Catalyzed by Acetoacetate Decarboxylase. Importance of Lysine 116 in Determining the pKa of Active-Site Lysine 115. *Biochemistry* **35**, 41-46, doi:10.1021/bi9518306 (1996).

141 Yoshitake, S. *et al.* Mild and Efficient Conjugation of Rabbit Fab' and Horseradish Peroxidase Using a Maleimide Compound and Its Use for Enzyme Immunoassay. *The Journal of Biochemistry* **92**, 1413-1424 (1982).

142 Shen, B.-Q. *et al.* Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nature biotechnology* **30**, 184-189 (2012).

143 Lyons, A. *et al.* Site-specific attachment to recombinant antibodies via introduced surface cysteine residues. *Protein engineering* **3**, 703-708 (1990).

144 Kuhn, K. *et al.* Synthesis of functional Ras lipoproteins and fluorescent derivatives. *Journal of the American Chemical Society* **123**, 1023-1035 (2001).

145 Bader, B. *et al.* Bioorganic synthesis of lipid-modified proteins for the study of signal transduction. *Nature* **403**, 223-226 (2000).

146 Le, D. D., Cortesi, A. T., Myers, S. A., Burlingame, A. L. & Fujimori, D. G. Site-specific and regiospecific installation of methylarginine analogues into recombinant histones and insights into effector protein binding. *Journal of the American Chemical Society* **135**, 2879 (2013).

147 Schöneich, C. Mechanisms of Protein Damage Induced by Cysteine Thiyl Radical Formation. *Chemical Research in Toxicology* **21**, 1175-1179, doi:10.1021/tx800005u (2008).

148 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).

149 Floyd, N., Vijayakrishnan, B., Koeppe, J. R. & Davis, B. G. Thiyl Glycosylation of Olefinic Proteins: S‐Linked Glycoconjugate Synthesis. *Angewandte Chemie International Edition* **48**, 7798-7802 (2009).

150 Wynn, R. & Richards, F. M. [33] Chemical modification of protein thiols: Formation of mixed disulfides. *Methods in enzymology* **251**, 351-356 (1995).

151 Ellman, G. L. Tissue sulfhydryl groups. *Archives of biochemistry and biophysics* **82**, 70-77 (1959).

152 Boyne, A. F. & Ellman, G. L. A methodology for analysis of tissue sulfhydryl components. *Analytical biochemistry* **46**, 639-653 (1972).

153 Riddles, P. W., Blakeley, R. L. & Zerner, B. [8] Reassessment of Ellman's reagent. *Methods in enzymology* **91**, 49-60 (1983).

154 Davis, B. G. Synthesis of glycoproteins. *Chemical reviews* **102**, 579-602 (2002).

155 Davis, B. G., Maughan, M. A., Green, M. P., Ullman, A. & Jones, J. B. Glycomethanethiosulfonates: powerful reagents for protein glycosylation. *Tetrahedron: Asymmetry* **11**, 245-262 (2000).

156 Gamblin, D. P. *et al.* Glycosyl phenylthiosulfonates (Glyco-PTS): novel reagents for glycoprotein synthesis. *Organic & biomolecular chemistry* **1**, 3642-3644 (2003).

157 Gamblin, D. P. *et al.* Glyco‐SeS: Selenenylsulfide‐Mediated Protein Glycoconjugation—A New Strategy in Post‐Translational Modification. *Angewandte Chemie International Edition* **43**, 828-833 (2004).

158 Davis, B. G. The controlled glycosylation of a protein with a bivalent glycan: towards a new class of glycoconjugates, glycodendriproteins. *Chemical Communications*, 351-352, doi:10.1039/B009184G (2001).

159 Rendle, P. M. *et al.* Glycodendriproteins: A Synthetic Glycoprotein Mimic Enzyme with Branched Sugar-Display Potently Inhibits Bacterial

Aggregation. *Journal of the American Chemical Society* **126**, 4750-4751, doi:10.1021/ja031698u (2004).

160 Davis, B. G. *et al.* Selective Protein Degradation by Ligand‐Targeted Enzymes: Towards the Creation of Catalytic Antagonists. *ChemBioChem* **4**, 533-537 (2003).

161 Gamblin, D. P. *et al.* Chemical site-selective prenylation of proteins. *Molecular BioSystems* **4**, 558-561 (2008).

162 Chalker, J. M., Bernardes, G. J., Lin, Y. A. & Davis, B. G. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chemistry–An Asian Journal* **4**, 630-640 (2009).

163 Willey, J. M. & Van Der Donk, W. A. Lantibiotics: peptides of diverse structure and function. *Annu. Rev. Microbiol.* **61**, 477-501 (2007).

164 Gavaret, J.-M., Nunez, J. & Cahnmann, H. Formation of dehydroalanine residues during thyroid hormone synthesis in thyroglobulin. *Journal of Biological Chemistry* **255**, 5281-5285 (1980).

165 Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science* **2**, 1666-1676 (2011).

166 Strumeyer, D., White, W. N. & Koshland, D. Role of serine in chymotrypsin action. Conversion of the active serine to dehydroalanine. *Proceedings of the National Academy of Sciences* **50**, 931-935 (1963).

167 Bernardes, G. J., 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).

168 Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post‐Translational Modifications. *Angewandte Chemie International Edition* **51**, 1835-1839 (2012).

169 Morrison, P. M., Foley, P. J., Warriner, S. L. & Webb, M. E. Chemical generation and modification of peptides containing multiple dehydroalanines. *Chemical Communications* **51**, 13470-13473 (2015).

170 Wang, J., Schiller, S. M. & Schultz, P. G. A Biosynthetic Route to Dehydroalanine‐Containing Proteins. *Angewandte Chemie* **119**, 6973-6975 (2007).

171 Zhu, Y. & van der Donk, W. A. Convergent synthesis of peptide conjugates using dehydroalanines for chemoselective ligations. *Organic letters* **3**, 1189-1192 (2001).

172 Meledin, R., Mali, S. M., Singh, S. K. & Brik, A. Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1, 4-addition step. *Organic & biomolecular chemistry* **14**, 4817-4823 (2016).

173 Chooi, K. P. *et al.* Synthetic phosphorylation of p38α recapitulates protein kinase activity. *Journal of the American Chemical Society* **136**, 1698 (2014).

174 Brimble, M. A. *et al.* Synthesis of the Antimicrobial S‐Linked Glycopeptide, Glycocin F. *Chemistry‐A European Journal* **21**, 3556-3561 (2015).

175 Norris, G. E. & Patchett, M. L. The glycocins: in a class of their own. *Current Opinion in Structural Biology* **40**, 112-119 (2016).

176 Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nature communications* **6** (2015).

177 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).

178 Hutchison, C. A. *et al.* Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry* **253**, 6551-6560 (1978).

179 Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature* **299**, 756-758 (1982).

180 Clark, P. I. & Lowe, G. Chemical mutations of papain. The preparation of ser 25-and gly 25-papain. *Journal of the Chemical Society, Chemical Communications*, 923-924 (1977).

181 CLARK, P. I. & LOWE, G. Conversion of the Active‐Site Cysteine Residue of Papain into a Dehydro‐serine, a Serine and a Glycine Residue. *European Journal of Biochemistry* **84**, 293-299 (1978).

182 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

Chapter 2: General experimental procedures

Chemicals and solvents were purchased from Sigma-Aldrich UK, Acros UK, Alfa Aesar UK, Carbosynth or Fischer UK and were used as delivered unless stated otherwise. Thin layer chromatography (TLC) was carried out using plastic 0.20 mm Polygram® Sil g/UV254 plates that were dried using a heat gun and visualised under UV (λ_{max} 254 nm or 366 nm) or by use of anisaldhyde, potassium permanganate, sulfuric acid or vanillin dip. Melting points are uncorrected and are recorded using Leica hotstage microscope. Flash column chromatography was performed using Geduran® Si 60.8.2 or a Biotage Flash Purification System with a Kinesis Telos or Biotage Snap columns.

NMR Spectroscopy General

Deuterated solvents were used as the lock, and the residual protonated solvent as the internal reference peak. Spectra were analysed using Topspin or MestReNova. Correlation spectroscopy (COSY), heteronuclear multiple bond correlation (HMBC) and heteronuclear single quantum correlation (HSQC) were employed when required for signal assignment. Samples were prepared by dissolving ~5 mg of sample (^{1}H) or ~20 mg (^{13}C) in 500 µL of solvent, filtered if required.

^{1}H NMR Spectroscopy

Chemical shift information (δH) is given in parts per million (ppm) relative to the centre of the solvent signal ($CDCl_3$ 7.26 ppm, DMSO-d6 2.50 ppm, MeOH-d4 3.31 ppm, Acetone-d6 2.05 ppm or D_2O 4.79 ppm). Spectra were recorded on a Bruker AVIIIHD 400 nanobays (400.1 MHz) or AVII 500 (500.1 MHz). The number of protons for a resonance signal is given by nH based on integral values. Multiplicity is denoted by the bracketed symbol. Coupling constants (J) are reported to the nearest 0.1 Hz and are determined using MestReNova or Topspin.

^{13}C NMR Spectroscopy

Chemical shift information (δC) is given in parts per million (ppm) relative to the centre of the solvent signal ($CDCl_3$ 77.16 ppm, DMSO-d6 39.52 ppm, MeOH-d4 49.00 or Acetone-d6 29.84 ppm). Spectra were recorded on a Bruker AVIIIHD 400 nanobays (100.6 MHz) or AVII 500 (125.7 MHz) Signals are assumed to be singlets unless otherwise stated. Distortionless enhancements by polarisation transfer with retention of quaternaries (DEPTQ) pulse sequence with broadband proton decoupling were employed for spectrum recording.

^{19}F NMR Spectroscopy (small molecule)

Chemical shift information (δF) is given in parts per million (ppm) relative to the centre of the external standard $CFCl_3$ Spectra were recorded on a Bruker AVIIIHD 400 nanobays (376.4 MHz). Signals are assumed to be singlets unless otherwise stated. Pulse sequences with broadband proton decoupling were employed for spectrum recording, unless stated.

^{31}P NMR Spectroscopy

Chemical shift information (δP) is given in parts per million (ppm) relative to the centre of the external standard 85 % phosphoric acid. Spectra were recorded on a Bruker AVIIIHD 400 nanobays (161.9 MHz). Signals are assumed to be singlets unless otherwise stated. Pulse sequences with broadband proton decoupling were employed for spectrum recording, unless stated.

IR Spectroscopy

Infrared (IR) spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer using a KBr disk or attenuated total reflectance (ATR) attachment – oil or crystals of the compound were placed directly against the stage. A background measurement was subtracted from each spectrum and a new background measurement was recorded before each spectrum was recorded.

Small Molecule Mass Spectrometry

Low-resolution mass spectra were recorded on an Aglilent 6120 Quadrupole spectrometer using chemical ionisation (CI), positive and negative mode are described where appropriate. High-resolution mass spectra were recorded on a Bruker MicroTOF ESI mass spectrometer

UV-Vis Spectroscopy

UV-Vis spectroscopy was performed in a BMG Labtech Spectrostar Nano. Data was analysed in BMG Labtech Spectrostar Nano MARS. Grenier flat bottom 96 well plates were used for sample containment. Path length corrections were used. Spectra were recorded at 37 °C. *para*-Nitrophenol (*p*NP) absorbance values were recorded at 405 nm.

Glovebox

Anaerobic atmosphere was achieved using a Belle Technology glovebox equipped with the BASF R3-11G catalyst. The oxygen level was measured below 6 ppm during all reactions.

Protein Mass Spectrometry

Protein mass spectrometry: Liquid-chromatography-mass spectrometry (LC-MS) was performed using:

LC/MS method A Waters LCT Classic coupled to a Shimadzu 20 Series HPLC using a Merck Chromolith C18 column (50 x 2 mm). Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, were used as mobile phase at a flow rate of 0.4 ml/min. The gradient program was: 95% A (2 min isocratic) to 95% B after 8 min then isocratic for 3 min. The electrospray source parameters were: capillary voltage: 3000 V, cone voltage: 250 V. Spectra were calibrated to at least 17 matched peaks of the multiply charged ion series of equine myoglobin, run under the same conditions.

LC/MS method B Waters LCT Classic coupled to a Shimadzu 20 Series HPLC using a Thermo Proswift (250 x 4.6 mm x 5 µm) column. The gradient

was programmed as follows: 95% A (5 min isocratic) to 100% B after 15 min then isocratic for 3 min.

LC/MS method C Waters LCT Premier XE coupled to a Waters 1525 Micro HPLC using a Merck Chromolith C18 guard cartridge. Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, were used as mobile phase at a variable flow rate. The gradient program was: 95% A (1 min isocratic, 0.4 ml/min) to 100% B after 5 min (0.4 ml/min) then isocratic for 3 min (flow rate increasing to 1.0 ml/min). The spectra were calibrated automatically. The electrospray source parameters were: capillary voltage: 3000 V, cone voltage: 160 V. Nitrogen was used as the desolvation gas and nebuliser at a total flow of 600 l/hr in both spectrometers. .

MS Data Analysis

Spectra were deconvoluted using MassLynx (Waters) using the Maximum Entropy deconvolution algorithm. Percentage conversions were derived from the intensity of the desired product divided by the sum of intensities from the starting material and products. For the 'chemical mutation' section, all conversions are >85% unless explicitly stated. Where applicable, the deconvoluted mass spectra data was then processed using GraphPad Prism (GraphPad Software).

LC/MS Analysis of protein modification reactions and purified products:

To analyse reactions and purified protein products, mass spectra for the total protein content (pink box in example below) of the total ion count (TIC) chromatogram were combined using MassLynx (Waters) or the MassHunter BioConfirm software tool (Agilent Technologies) to obtain the ion series of the protein.

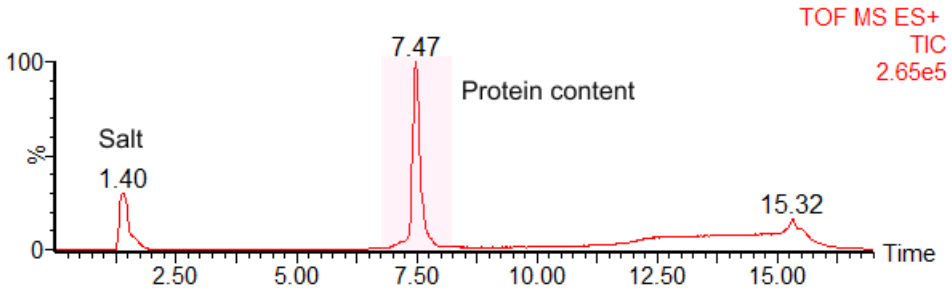
Mass spectra were then calibrated (Masslynx) using a calibration file constructed from a minimum of 17 matched peaks from the multiply charged ion series of equine myoglobin. The calibrated ion series was deconvoluted using the maximum entropy algorithm in MassLynx.

Percentage conversions were derived from the peak intensity of the desired product divided by the sum of intensities from the starting material and products. For the 'chemical mutation' sections, all conversions are >85% unless explicitly stated.

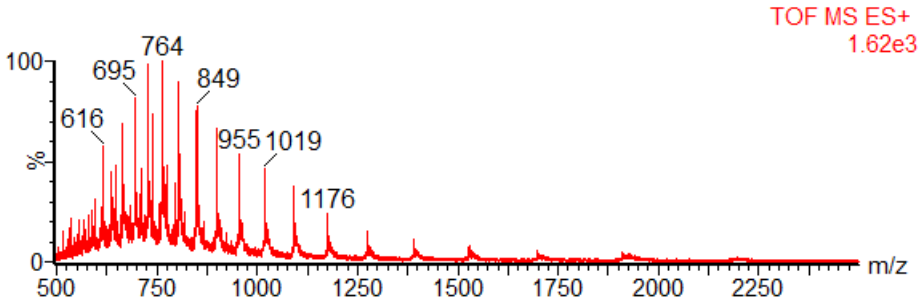
The correlation of peak height of mass spectra to relative amount of protein was validated by correlation with gel and nanodrop A280 analysis.

An example of the LC/MS analysis employed in this thesis:

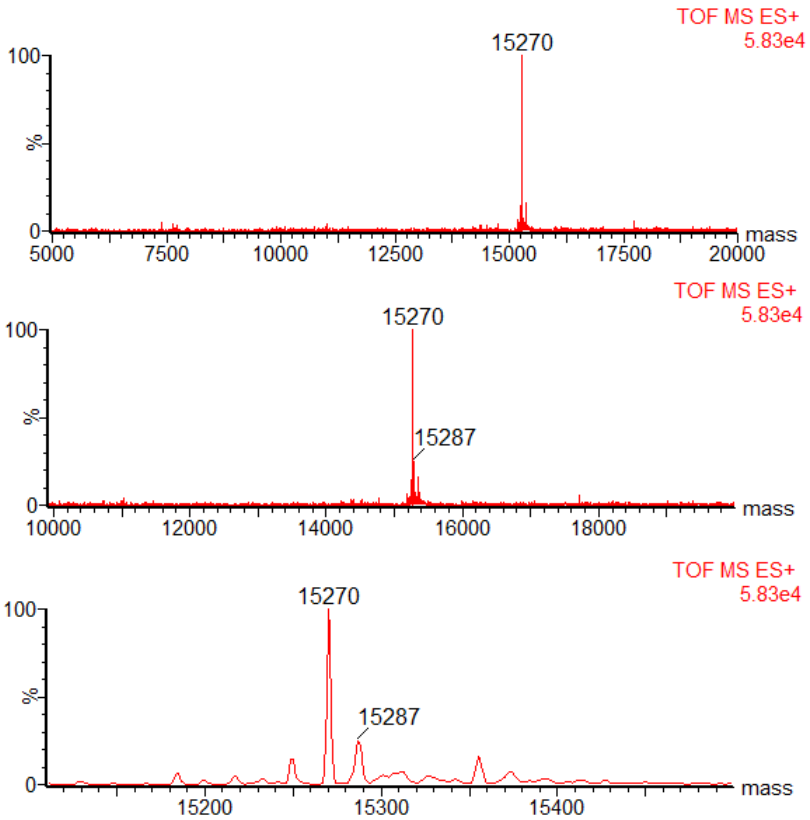
TIC chromatogram of reaction mixture, desalted mixture or purified protein:



Protein ion series:



Deconvoluted specta:



Protein digestion and analysis by mass spectrometry

A 50 µL aliquot of a solution containing modified protein in supplied buffer was further processed for in-solution trypsin digestion as described previously[1]. Histone H3 was digested with trypsin; SBL, Acra, were digested with elastase. In brief, 125 µl of H_2O and 5 mM DTT (final concentration) were added and the sample incubated for 30 minutes at room temperature. A second incubation of 30 minutes at room temperature was performed after adding 20 mM iodoacetamide (final concentration). The sample was then concentrated and desalted by chloroform / methanol precipitation as described previously[2]. 50 µl of 6 M urea, 0.4 M Tris pH 7.8 solution was used to solubilise the sample pellet and 3 µl of trypsin (20 ng/µl, Promega) or 1 µl of elastase (1ug/ul) was added to the sample diluted 6 times with H_2O for an overnight incubation at 37 °C. The sample was then desalted using a C18 SEP-PAK cartridge according to the manufacturer's instructions (Waters), dried via speedvac centrifugation and resuspended in a solution of 20 µl 2% CH_3CN / 0.1% TFA in H_2O. Peptides were analysed by nano–liquid chromatography tandem mass spectrometry (nano–LC-MS/MS) using a Nano-Acquity-UPLC (C18 column 75 µm × 250 mm, 1.7 µm particle size; Waters) coupled to an Orbitrap Elite tandem mass spectrometer (Thermo Scientific, Bremen, Germany) with resolution of 120,000 full-width half maximum at mass/charge 400. Top 20 precursor ions were selected and fragmentation was performed in collision-induced dissociation (CID) mode. The samples were loaded in 99.5 % buffer A (0.1% FA in H_2O). The gradient used to elute the peptides commenced with a 3 minute isocratic gradient composed of 92% buffer A and 8% buffer B (0.1% FA in CH_3CN), followed by a linear gradient from 8 to 35 % of buffer B over 60 minutes and two 8-97% step gradients of buffer B for 3 minutes each for washing. The flow rate was set to 250 nl/minutes and total length of the analysis was 100 minutes to allow column re-equilibration.

Raw MS data were used for MS data analysis and the characterisation of chemical modifications on proteins. Searches were performed using PEAKS

software v7.0 including the following parameters: Data refinement was applied by correcting the precursor's mass. A precursor mass tolerance of 10 ppm and a fragment ion tolerance of 0.5 Da were applied for the De Novo and the PEAKS searches. The enzyme used was trypsin or elastase. No unspecific cleavages were allowed in the case of trypsin. A maximum of 2 or 3 missed cleavages were allowed when trypsin was used. The PTM search settings were as follows: fixed modification for cysteine carbamidomethyl and variable modification for asparagine and glutamine deamidation, methionine oxidation and Lysine methylation (+14), dimethyl (+28) or trimethylation (+42). Expected mass shifts for novel chemical modifications were entered manually. Databases were downloaded in April 2015 from UniProtKB using reviewed sequences (swissprot) of Escherichia coli (strain K12) (4433 sequences). All the proteins searched were added into the Ecoli database. Data was searched against the specific database containing the protein of interest (sp|P84233|H32_XENLA Histone H3.2, AcrA, SBL, Npβ, cAbLys, annexinV).

The false discovery rate (FDR) was calculated using the decoy fusion method[3] which uses a concatenation of the decoy and target sequences of the same protein together as a "fused" sequence enabling a better assignment of the decoy.

For the analysis of glycosylated protein material, digested glycopeptides were analyzed by ultra-high performance liquid chromatography tandem mass spectrometry (UHPLC-MS/MS) using a Dionex Ultimate 3000 UHPLC (C18 column, 75 µm × 250 mm, 1.7 µm particle size; Thermo) coupled to an Orbitrap Fusion Lumos tandem mass spectrometer (Thermo Scientific, Bremen, Germany). The gradient used to elute the peptides consisted of a linear gradient ranging from 2 to 5% of buffer B (0.1% FA in CH_3CN; buffer A: 0.1%FA in H_2O) over 3 minutes, followed by 5-63% buffer B over 63 minutes at a flow rate of 250 nl/minute. Two step gradients of 3-99% buffer B for 5 and 6 minutes, respectively, were used to clean the column. The total length of the analysis was 100 min to allow column re-

equilibration. The method used for the samples analysis, Higher energy collisional dissociation product dependent electron transfer higher energy collisional dissociation (HCDpdEThcD), was performed using the following steps: All MS/MS spectra were acquired in high collision energy dissociation (HCD) in data-dependent analysis (DDA) mode, and all the fragmentations of glycopeptides leading to the formation of glycan oxonium ions (HexNac: 204.0867, HexNac fragment: 138.0545 m/z and HexNacHex: 366.1396 m/z) triggered electron transfer dissociation (ETD) spectra on the glycopeptide precursors allowing specific fragmentation of the glycopeptides in ETD mode and therefore identification and localization of glycan in the peptide (see supplementary for method information) [4,5].

Histone protein expression and purification[6]

pET3 bacterial plasmids encoding all canonical Xenopus laevis histones were kindly donated by Prof. Rob Klose, University of Oxford, Oxford. Site-directed mutagenesis to introduce Cys mutations was performed using QuikChange II (Agilent), according to manufacturer's instructions.

The appropriate histone encoding plasmid was transformed into BL21 (DE3)pLysS cells, with chloramphenicol and ampicillin added. A single colony was selected after overnight growth and used to inoculate 20 mL LB medium with the same antibiotics. This culture was grown at 37 °C overnight. 10 mL of starter culture was then added to 500 mL 2xTY medium containing the same antibiotics and grown at 37 °C until OD600 = 0.4-0.6. IPTG was added, to a final concentration of 0.5 mM and the flask was shaken at 37 °C for 2 h (3 h for Histone H4). Cells expressing histones are prone to lysis and longer expression time is detrimental on the yield. Cells were harvested by centrifugation and the pellet was resuspended in wash buffer (50 mM Tris pH 7.5, 100 mm NaCl) with protease inhibitor (one tablet per 50 mL) and frozen and stored at -80 °C.

Frozen cells were thawed in a water bath, when the solution was viscous DNAase was added, sonicated 5 × at 40 % amplitude in 30 s bursts and then centrifuged at 20k rpm for 20 min. The supernatant was discarded and the pellet was resuspended in 40 mL TW buffer (above lysis buffer +

1% Triton-X). Sonication was repeated at 40 % amplitude for 30 s, and the suspension centrifuged at 20k rpm for 10 min. The pellet was washed two more time with TW buffer and once with wash buffer. After addition of 1 mL DMSO the pellet was minced with a spatula and left at rt for 10 min. 7-15 mL unfolding buffer (7M guanidine HCl, 20 mM Tris pH 7.5, 10 mM DTT) was then added, and shaken for 1h at rt, after which it was centrifuged for 10 min at 20k rpm at rt. The supernatant was decanted and loaded onto an S200 size-exclusion column pre-equilibrated with 1 CV SAU-1000 (7 M urea, 20 mM NaOAc pH 5.2, 1 M NaCl, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine). The histone was eluted with SAU-1000 buffer, and analysed by SDS-PAGE. Fractions containing purified histones were pooled and dialyzed against SAU100 buffer (7 M urea, 20 mM NaOAc pH 5.2, 100 mM NaCl, 1 mM EDTA, 10 mM DTT, 1 mM benzamidine).

The histones were further purified by ion exchange chromatography using a 1 or 5 mL HiTrap SP HP (GE healthcare; 1mL: 17-1151-01; 5mL: 17-1152-01). For this, the lyophilised histone powder was dissolved in 1 – 4 mL (10-40 mg) SAU-100 buffer by rotation at rt for 20 min. Then particulates were removed by centrifugation (15k rpm) and the sample was loaded onto the HiTrap SP HP equilibrated in SAU-100 buffer. The protein was eluted using a linear gradient to 70% SAU-1000 buffer over 20 CV. Pure fractions were pooled, dialyzed against water containing 2 mM BME, and lyophilized.

Dehydroalanine formation on histone substrates (General procedure 1)[7]

The relevant histone cysteine mutant (10 mg of lyophilized powder) was dissolved in 500 μL of reaction buffer (3M guanidinium hydrochloride, 100 mM sodium phosphate, pH 8.0) and 30 mg DTT was added. The resultant protein solution was shaken at room temperature for 30 minutes. Excess DTT was removed by passing the solution through a PD MiniTrap G25 (equilibrated with the same buffer). Eluting with 1 mL of the reaction buffer gave the protein at a final concentration of 10 mg/mL.

To the reduced H3 mutant was added 50 μL of a 2,5-dibromohexane diacetamide (DBHDA) stock solution (0.5 mM, 100 mg in 610 μL of DMF). The reaction was gently shaken for 45 minutes at room temperature and then 90 minutes at 37 °C. The reaction mixture was concentrated to 500 μL using a Vivaspin 500 concentrator (MWCO 10 kDa) and excess reagent was removed by passing the solution through a PD MiniTrap G25 (equilibrated with chemical mutagenesis reaction buffer). Eluting with 1 mL of the same buffer provided the dehydroalanine-containing protein.

References: Chapter 2

1 Fischer, R. & Kessler, B. M. Gel‐aided sample preparation (GASP)—A simplified method for gel‐assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* **15**, 1224-1229 (2015).

2 Wessel, D. & Flügge, U.-I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* **138**, 141-143 (1984).

3 Zhang, J. *et al.* PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics* **11**, M111. 010587 (2012).

4 Wu, S.-W., Pu, T.-H., Viner, R. & Khoo, K.-H. Novel LC-MS2 product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Analytical chemistry* **86**, 5478-5486 (2014).

5 Saba, J., Dutta, S., Hemenway, E. & Viner, R. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *International journal of proteomics* **2012** (2012).

6 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Chromatin Protocols*, 1-16 (1999).

7 Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post‐Translational Modifications. *Angewandte Chemie International Edition* **51**, 1835-1839 (2012).

Chapter 3: Development of a low-valent metal approach to post-translational mutagenesis

Introduction

When Koshland first proposed the use of dehydroalanine as a 'tag' or precursor handle for 'chemical mutation' to divergent side chains[1], he had in mind the nucleophilic addition of thiols to dehydroalanine. However, the reactivity of dehydroalanine is not limited to thia-Michael additions[2]. Indeed, C-N[3] and C-O[4] bond formation is also possible on dehydroalanine model systems using appropriate nucleophiles. In the case of C-N bond formation, this has also been extended to proteins using imidazole as a model nucleophile[5], providing a histidine mimetic that may find use in enzymatic catalysis. The α-β unsaturated nature of dehydroalanine, aside from rendering the residue a good acceptor in Michael-type conjugate additions, should also accord an affinity for alkyl free radicals, enabling C-C bond formation[6-8]. It was speculated that alkyl radicals should offer a unique platform for protein modification as alkyl radicals tend to be both tolerant of the aqueous conditions necessary for biomolecule modification and unreactive with the majority of existing functional groups present in proteins[9].

The addition of alkyl radicals to dehydroalanine is not without precedent. In a series of papers beginning in 1988[10], Crich et al. reported the preparation of both monomeric amino acids and di- and tripeptides[6] *via* radical additions to dehydroalanine containing peptides (**Figure 3.1**). The additions were accomplished by the treatment of alkylmercury halide precursors with sodium borohydride (first reported by Giese[11]), generating free alkyl radicals that reacted with the dehydroalanine peptides to afford the desired aliphatic derivatives. In all cases, the additions proceeded with negligible diastereoselectivity, as expected from a radical process. Notably, reaction with isopropylmercury chloride provided access to racemic leucine, an example of 'chemical mutation' at the peptide level. However, the additions were all performed in organic (dichloromethane or ethanol)

solvents, therefore the tolerance of the reaction for aqueous buffer and hence protein modification remained an open question.

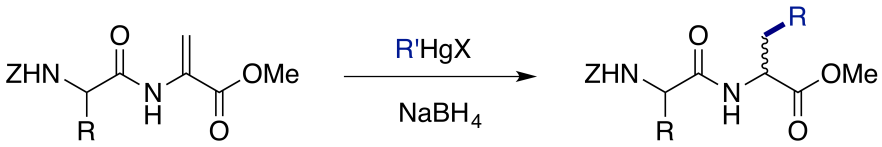


Figure 3.1: Initial report of dehydroalanine alkylation via alkyl radicals by Crich et al.[6,10]

Building on this early work, the formation of *C*-glycopeptides by tin-hydride mediated radical addition of glycosyl bromides to dehydroalanine derivatives was reported[8]. As in Crich's initial report, peptides up to tri-peptides could be successfully used as substrates although, again, the observed diastereoselectivites were poor. Initial work directed towards enabling stereocontrol in these small-molecule systems has also been reported[7], with the use of cyclically constrained dehydroalanine derivatives resulting in modest diastereoselectivity.

In searching for a mild alternative to the use of alkylmercury or tin compounds for the radical addition to dehydroalanine, Chalker and Davis proposed the use of low-valent metals, such as Zn^0, as a means to generate free radicals from organohalides by single electron transfer (SET)[9] (**Figure 3.2**). This proposal was based on the work of Luche[12], who had previously demonstrated metallic zinc – mediated radical additions of organic bromides and iodides to α-β unsaturated carbonyl compounds, such as enones (henceforth termed the 'Luche reaction'). Notably, this transformation could be performed in aqueous solution, a fact of key importance for potential application to protein modification. The 1,4 conjugate addition was found to be promoted by the use of ammonium chloride as an additive[12], or the use of activated Zn(Cu) couple under sonication conditions[13].

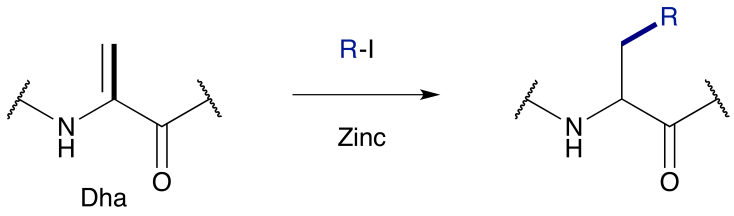


Figure 3.2: Proposed modification of protein dehydroalanine employing alkyl radicals generated from alkyl iodides *via* single electron transfer from Zn[09]

Mechanistic studies performed by Luche and co-workers were inconclusive, although strongly suggestive of a free radical pathway[14]. An organozinc intermediate is unlikely, as aqueous conditions are known to rapidly quench such organometallic species. Further, the reactivity sequence of the organohalides employed (3° > 2° > 1°, iodides > bromides >> chlorides) reflects well the order of radical stability and C-X bond strengths, respectively. The initial addition of an alkyl radical to an enone generates an α-keto radical, which could theoretically be quenched by either a hydrogen atom transfer from solvent or via a reduction-protonation sequence. Luche and co-workers employed isotope- labeling studies[14] to demonstrate that the latter pathway was operative, implying that a second SET event from the zinc metal is necessary to quench the α-keto radical resultant from the initial radical addition. This quenching pathway offers little scope for stereocontrol, and hence it was not unexpected that extensive racemization occurred at the α-position. Together, these inquiries led Luche to propose the mechanism outlined in **Figure 3.3[14]**.

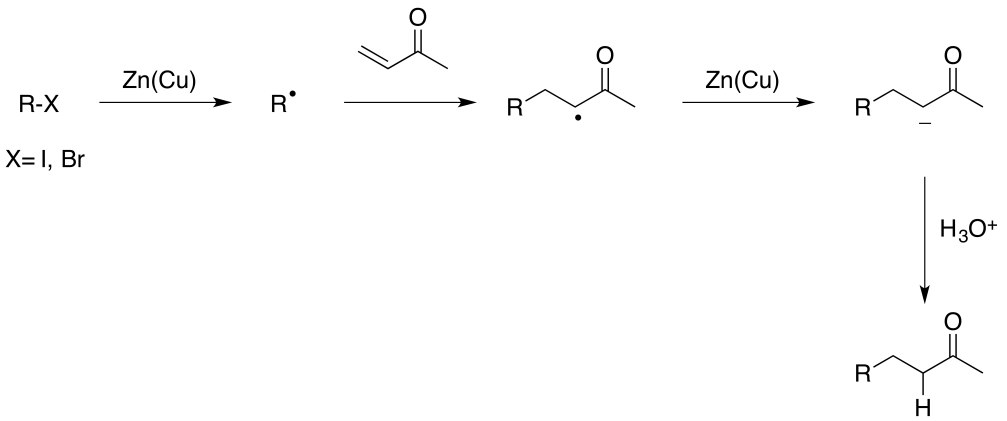


Figure 3.3: Mechanism for the Zn(Cu)-couple mediated radical conjugate addition to enones, proposed by Luche[14].

Chalker and Davis recognized that the aqueous-compatible nature of the zinc-mediated Luche reaction of enones promised potential as a feasible route to modification of dehydroalanine-containing peptides and proteins[9]. Initial studies were focused on demonstrating reaction feasibility on small molecule dehydroalanine model compounds[5]. Leucine was chosen as an initial target side-chain, as it was hypothesized that it could be synthesized from dehydroalanine (Dha) by reaction with readily available isopropyl iodide as a precursor to the isopropyl radical. Pleasingly, reaction of a model Dha derivative with isopropyl iodide using Zn^0 –mediated conditions in aqueous buffer (pH 6, NH_4Cl) afforded the desired leucine product (**Figure 4**). Building on this initial result, Chalker[5] then extended the methodology by varying the alkyl-halide precursor to afford a number of natural and unnatural amino acids. Notably, polar protic functionality, including unprotected hydroxyl and amine groups, was well tolerated, allowing synthesis of the natural residue lysine and homologated serine analogues.

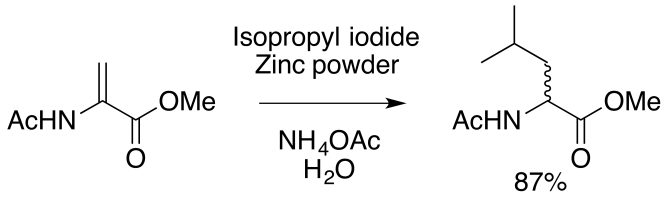


Figure 3.4: Dha→Leu modification in a model system. Experiment performed by Justin Chalker[5].

Following this initial success with small molecule models, Chalker and co-workers then turned to reactions using the model protein subtilisin (from *Bacillus lentus*, hence 'SBL')[5]. A single cysteine derivative of SBL was prepared and converted to the Dha-containing protein by reported methods[15]. Pleasingly, attempted 'chemical mutation' to leucine by reaction of SBL-Dha with isopropyl iodide and Zn^0 in aqueous ammonium acetate buffer (pH 6) resulted in clear formation of SBL-Leu, as ascertained by LC-MS analysis (**Figure 3.5**)[5]. Unfortunately, additional side-products were detected in varying amounts. The masses of these side-products were consistent with a cleavage event at the introduced dehydroalanine residue, although no attempt was made to isolate or further characterize these

products. It was thus proposed that further optimization of the 'Luche reaction' for chemical mutation would necessitate increased understanding of this side-reaction and its elimination by the rational alteration of reaction conditions. Attempts in these directions are outlined in this chapter.

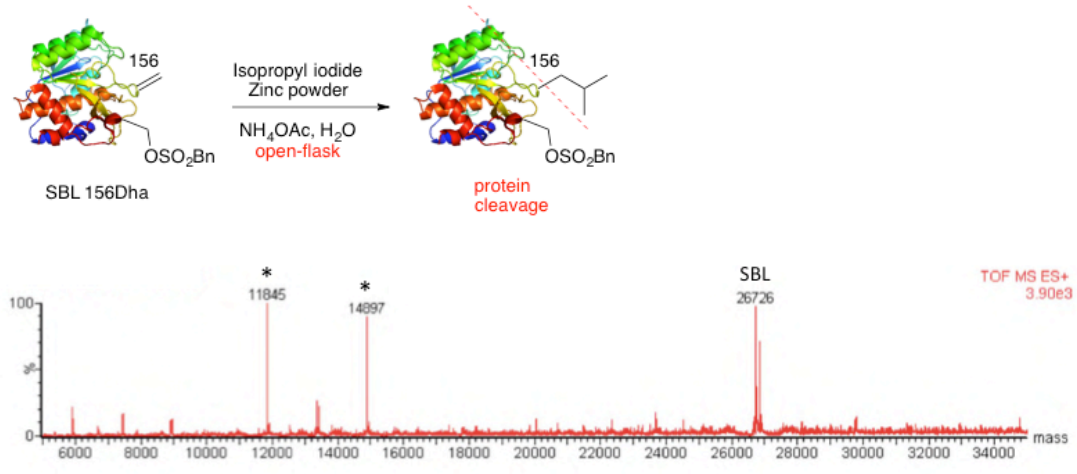


Figure 3.5: Initial attempt at protein chemical mutation by Justin Chalker[b]. * indicates protein fragments generated during the reaction.

Results

In order to examine the generality of the cleavage side-reaction, an additional protein model, histone H3, was selected to complement the single cysteine mutant of SBL (SBL-156C). Histone proteins are tractable to large-scale expression[16] in *E. coli* (SBL must be expressed in *Bacillus lentus*), enabling ready generation of different cysteine mutants *via* standard site-directed mutagenesis. Furthermore, in nature, histones are extensively post-translationally modified, with important roles in the regulation of gene expression and DNA replication and repair[17]. Thus, histone proteins are an important test-bed for new means of incorporation of post-translationally modified proteins, as already alluded to in the introduction to this thesis[18].

Three histone H3 mutants (K9C, K27C and K64C) were expressed and purified using standard procedures[16] (see **Experimental Section**). Along with the use of the differentially structured SBL, which enabled comparison of protein substrate, these three mutants allowed testing of the influence (if any) of varying the reaction site within a single protein scaffold. Dehydroalanine was site-specifically installed into the three histone proteins following the bis-alkylation elimination procedure previously reported by the Davis group[18]. Treatment of the relevant reduced cysteine mutants with 2,5-dibromohexane diacetamide (DBHDA) at slightly elevated pH (pH 8, phosphate buffer) affords a mixture of mono- and di-alkylated (sulfonium) cysteine that is eliminated, forming Dha, upon heating to 37°C (see **Figure 3.6** for example). A simple desalting step utilizing a commercial size-exclusion desalting column afforded the desired H3-Dha protein.

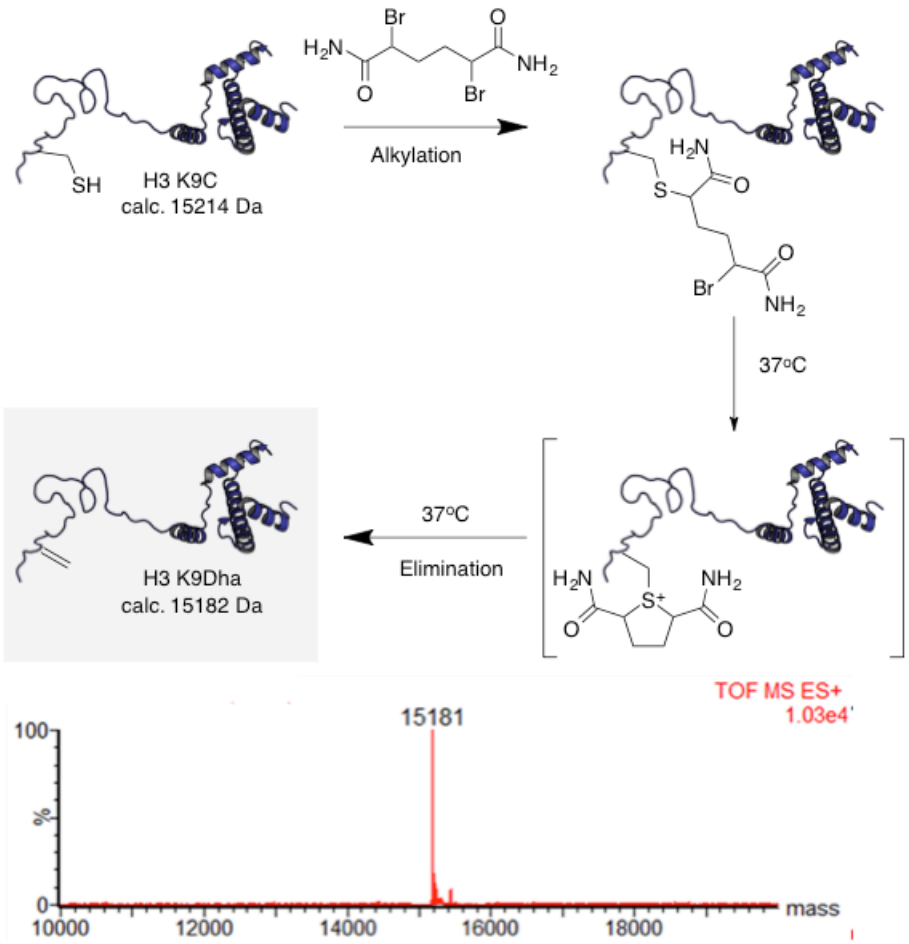


Figure 3.6: Conversion of cysteine to dehydroalanine on H3-K9C using DBHDA. LC/MS spectrum of final Dha-containing product. See **Experimental Section** for details.

With the dehydroalanine-containing proteins in hand, attention turned towards repeating the observations previously reported within the Davis group. Indeed, treatment of both SBL-156Dha and the three H3-Dha proteins with Zn^0 and isopropyl iodide, under nearly identical conditions to those used by Chalker[5], afforded a similar product profile, with successful mutation to leucine (H3-K→Leu calc. 15224 Da., SBL-156L(221PMS) calc. 26879 Da.) and lower molecular weight cleavage products both observed (**Figure 3.7**). This result demonstrated that the cleavage side reaction was a general feature of such a reaction at dehydroalanine, and was not dependent on either the protein substrate or the specific location of the Dha residue within the polypeptide sequence. Further experimentation with SBL156Dha, Zn^0 and a range of alternate alkyl halides also confirmed

that the cleavage side-reaction did not depend on the use of isopropyl iodide as 'side-chain donor' (see **Experimental Section**).

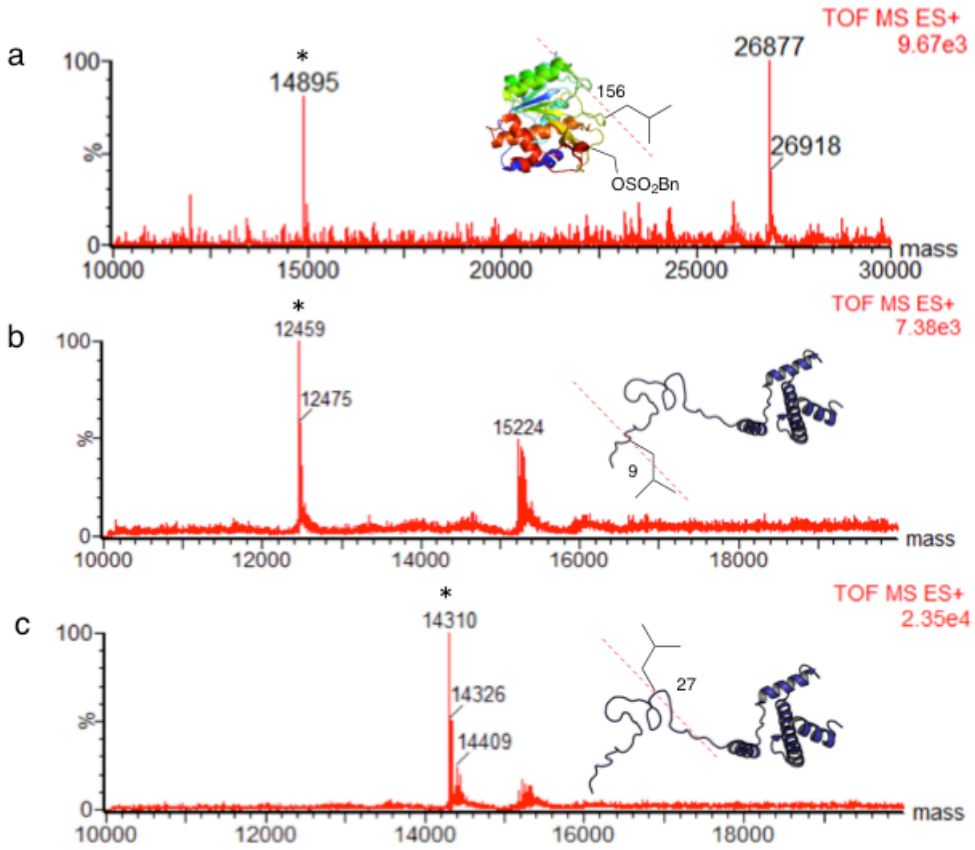


Figure 3.7: Reaction of both SBL-Dha and two mutants of H3-Dha results in protein cleavage under aerobic conditions. * indicates identified protein fragment resultant from cleavage at the introduced dehydroalanine.
See **Experimental Section** for reaction conditions.

Identification and labeling of cleavage products

Protein cleavage at dehydroalanine is a known reaction in biochemistry that has been shown to produce distinctive fragmentation products (**Figure 3.8**)[19,20]. In order to confirm the identity of the cleavage products, both careful analytical LC-MS and a chemical labeling approach were employed. Under certain chromatographic conditions, the two fragments could be separated for both SBL-156Dha and H3-64Dha (**Figure 3.8**) reaction mixtures. These results gave intact mass assignments consistent with cleavage at dehydroalanine following an initial attack by an isopropyl unit, however the exact atomic nature of the products cannot be ascertained with confidence by this method.

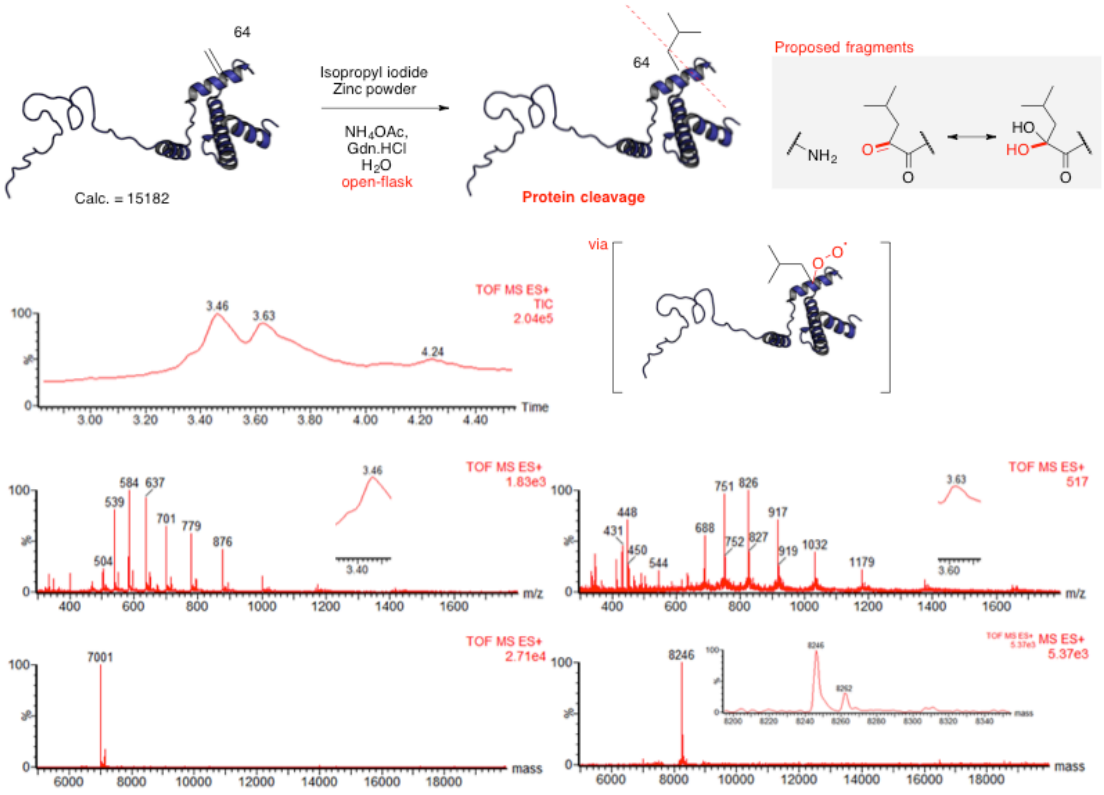


Figure 3.8: Isolation of side product fragments from Leu 'mutation' in H3-K64Dha model system. N-terminal fragment calc. 6999 Da., C-terminal fragment calc. 8243 Da.

The known mechanisms for backbone fragmentation at dehydroalanine all produce a distinctive *N*-terminal pyruvoyl moiety on the *C*-terminal fragment (**Figure 3.8**)[19,20]. Protein-bound diketones are known to react with hydrazides[21-23] and oximes[24] to generate hydrazone and hydroxylamine products, respectively. Indeed, incubation of a cleaved protein solution, derived from the reaction of H3-K9Dha with isopropyl iodide and zinc in open air, with biotin hydrazide, resulted in a distinctive mass shift corresponding to the addition of the hydrazide to the *C*-terminal fragment (**Figure 3.9**). Notably, as in previous reports, nucleophilic catalysis[25] with *m*-phenylenediamine[26] was necessary for hydrazide labeling. This result confirmed the presence of a pyruvoyl moiety at the *N*-terminus of the *C*-terminal protein fragment, consistent with the proposed cleavage occurring site-specifically at the introduced Dha tag. However, the exact mechanism of protein cleavage remained unknown.

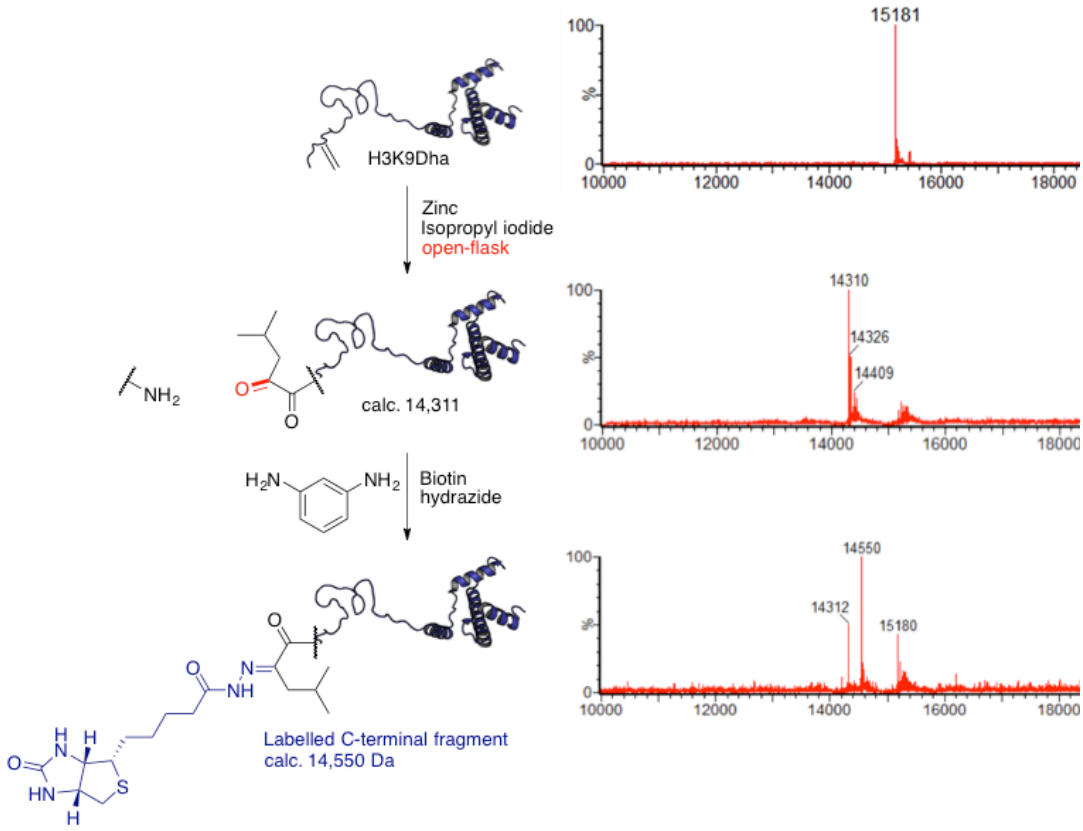


Figure 3.9: Chemical labelling of pyruvoyl-containing N-terminal fragment derived from Leu 'mutation' of H3-K9Dha under aerobic conditions. See **Experimental Section** for details.

Towards mechanistic understanding of the cleavage reaction

Polypeptide cleavage at dehydroalanine can potentially occur through a number of mechanisms[19,20]. Acid-catalysed equilibration to the iminium form of the Dha enamide followed by addition of an equivalent of water could result in an unstable tetrahedral intermediate that readily cleaves to form peptide amide and pyruvoyl peptide fragments. Indeed, such a mechanism has been proposed for the degradation of the commercially available peptide antibiotic, nisin[27,28]. Treatment of Dha-containing peptides and proteins with strong acid has been used synthetically for protein cleavage and subsequent peptide mapping[28-30]. A base-catalysed hydrolysis mechanism is also possible, as evidenced by observation of an increased rate of cleavage upon storage of commercial IgG1 antibodies in alkaline solution[31].

The possibility of an acid-catalysed cleavage reaction, perhaps employing Zn^{2+} ions as Lewis acid catalysts, was discounted as simple incubation of the Dha-protein in the presence of zinc, with no added halide, did not lead to protein cleavage, even under forcing conditions. Further, during Luche reactions leading to protein cleavage, no change in the pH from ~6.0 could be observed. Literature reports of hydrolytic cleavage at dehydroalanine residues indicate a requirement for pH extremes of ~2.0 (for acid-catalysed hydrolysis)[28,31,32] or ~10.0 (for base-catalysed hydrolysis)[33], rendering this mechanism unlikely in this case.

Cleavage of dehydroalanine residues in peptides and proteins has also been reported to occur *via* an oxidative mechanism, particularly under conditions in which radicals are present[19]. In this scenario, radical attack at the β position generates a captodative (hence stabilized) radical at the Cα-position (relative to the amide carbonyl) of the polypeptide backbone (**Figure 3.10**)[19]. In an open-air solution, reaction of carbon-centred radicals with molecular oxygen (in its radical-reactive triplet state, $^3O_2$) is known to take place at diffusion-limited rates. Reaction with oxygen may afford an intermediate protein-bound peroxyl intermediate that can decompose by rapid elimination of $HOO^{\cdot}/O^{\cdot-}{}_2$ (**Figure 3.10**), generating an iminium-type species that leads to backbone cleavage *via* attack of water, as described above for the acid-catalysed reaction. Alternatively, hydrogen atom abstraction by the peroxyl radical results in hydroperoxides, which themselves decompose to alkoxyl radicals that fragment, resulting in backbone cleavage (**Figure 3.10**).

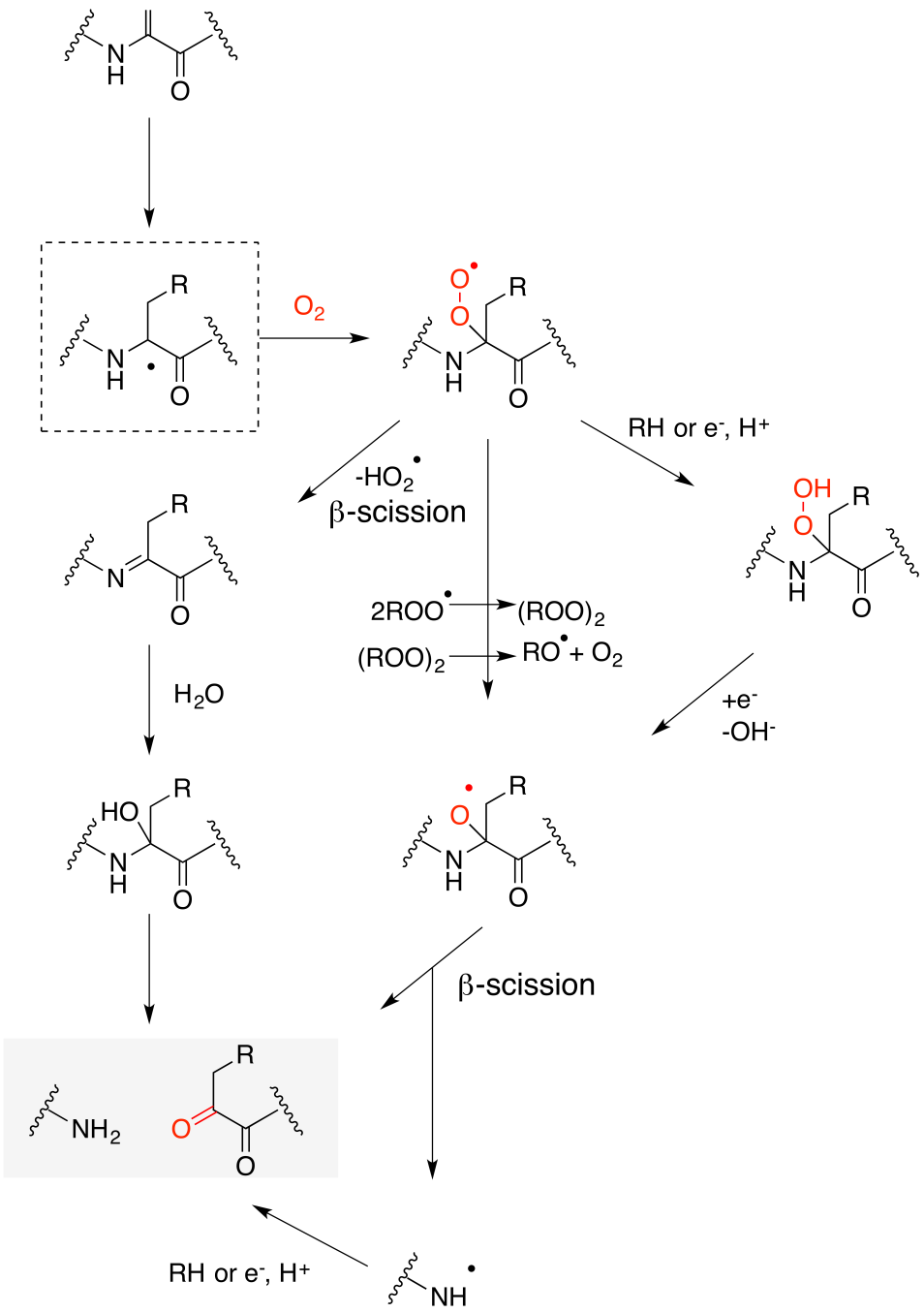


Figure 3.10: Proposed mechanisms for oxidative cleavage of proteins following Cα radical generation. Adapted from[19].

If the observed protein fragmentation resulted from a radical-mediated oxidative cleavage, rigorous exclusion of oxygen from the reaction solution should prevent protein cleavage. Performing the 'Luche' reaction within an anaerobic glovebox environment (<6 ppm O_2), on both SBL-Dha and histone H3-Dha proteins, led to complete elimination of the cleavage side-reaction. This result confirms the hypothesis of backbone cleavage *via*

trapping of an intermediate Cα-radical by molecular oxygen. Notably, a significant equilibration period of the protein solution in the anaerobic environment was necessary to eliminate protein cleavage upon addition of reagents (see **Experimental Section**), highlighting the exquisite sensitivity of the reaction to oxgyen.

Identification of a di-alkylation side-reaction

Employing anaerobic conditions, reaction of SBL156Dha with isopropyl iodide and $Zn^{(0)}$ resulted in complete conversion to the desired SBL156Leu, as evidenced by both intact mass spectra and peptide mapping (**Figure 3.11**). However, in the case of the three histone H3-Dha proteins, an additional side-product, consistent with addition of a second isopropyl unit, was also observed (**Figure 3.12a**, calculated 15267 Da, observed 15266 Da). Control reactions run with isopropyl iodide but in the absence of zinc powder established that this side-product was not due to non-selective alkylation at nucleophilic residues, as in the absence of zinc no addition of isopropyl was observed. It was thus hypothesised that trapping of the presumed Cα radical intermediate[19] with a further isopropyl radical was responsible for the observed 'di-alkylation'. Indeed, this hypothesis was confirmed by peptide-mapping by tryptic digest and LC/MS-MS analysis of a reaction mixture derived from incubation of H3-K64Dha with zinc powder and isopropyl iodide, under anaerobic conditions (**Figure 3.12b**). In addition to the expected H3-64Leu peptide, an additional peptide product bearing an additional ~ +42 mass units at the leucine residue (e.g. +42.08 for the b8 ion in the spectrum shown, +42.05 is the calculated mass for an additional isopropyl unit at the Cα position) was detected. Further evidence for this assignment came from previous work in the Davis group[4] in which Dha-containing dipeptide (Boc-Thr-Dha-Me) was reacted with zinc powder and isopropyl iodide. In addition to the expected leucine-containing product, Boc-Thr-Leu-Me (42% isolated yield), an additional dialkylated product, Boc-Thr-Leu(*i*Pr)-Me was isolated in 32% yield. This side-product was fully characterised by NMR which revealed the

dialkylation to occur at the Cα position, in full accordance with the results obtained on whole proteins.

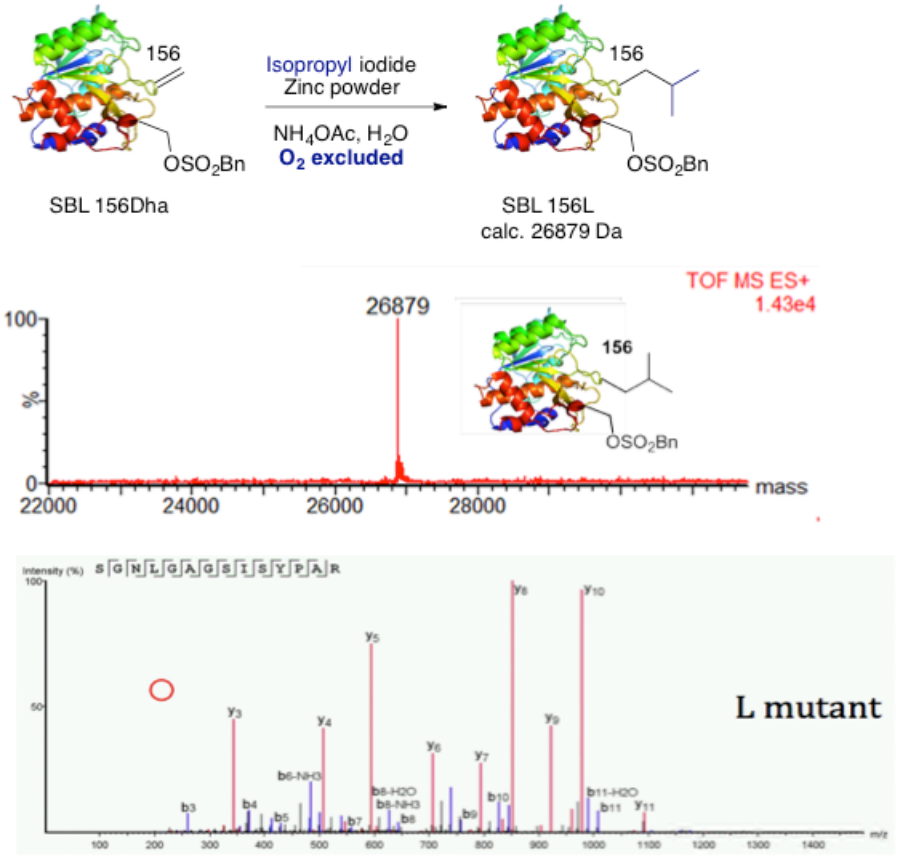


Figure 3.11: Successful chemical mutation of SBL-156Dha → Leu results from performing the reaction in an anaerobic environment. SBL 156L(221PMS) calc. 26879 Da. *Conditions:* SBL-156Dha (200 µL at 0.4 mg/mL, in 500mM NH_4OAc) 4640 eq. zinc powder, 7000 eq. isopropyl iodide, 30 minutes, within glovebox. Top: LC/MS spectrum of product protein. Bottom: LC-MS/MS identification of the modified peptide.

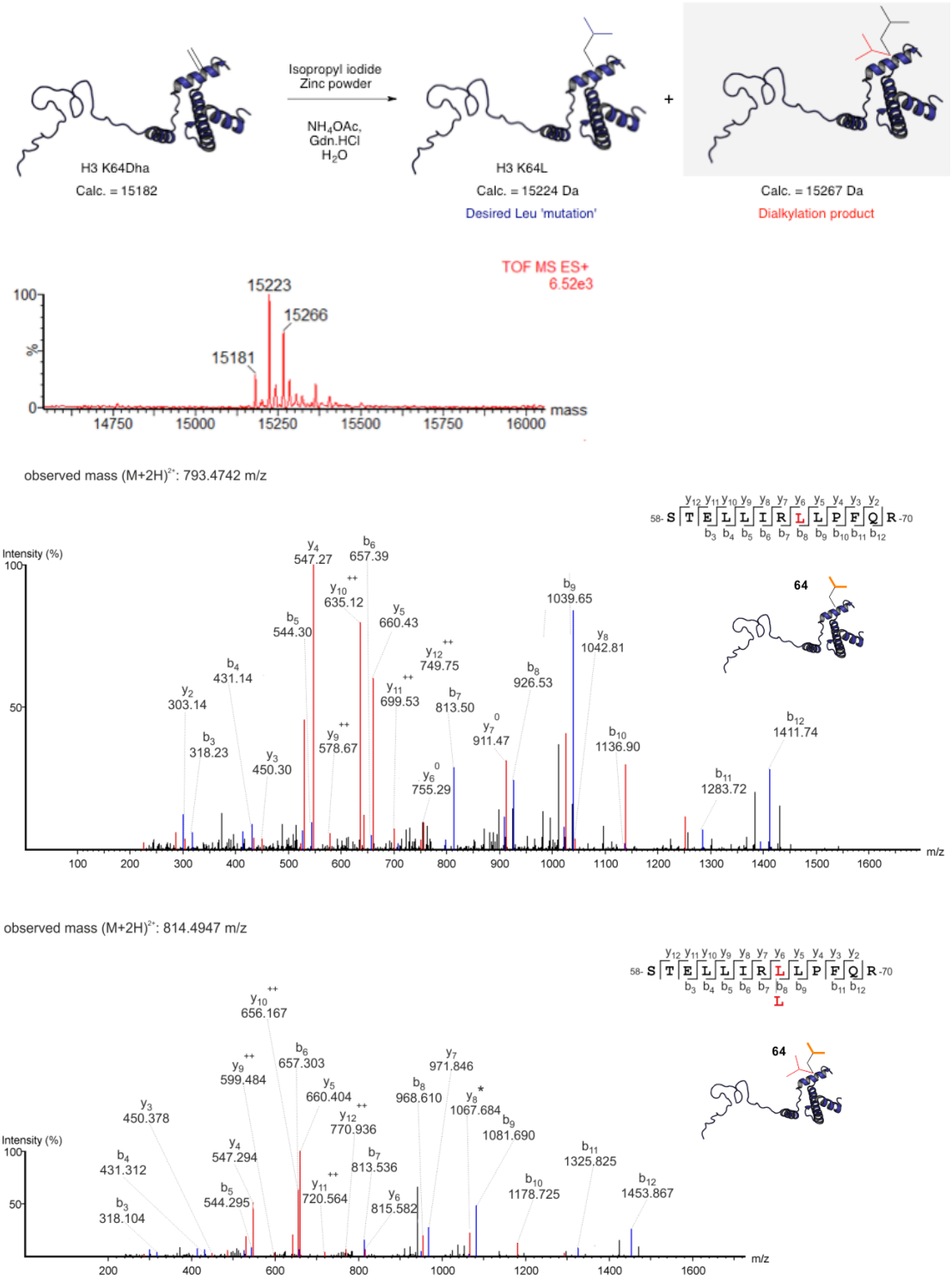


Figure 3.12: Identification of a di-alkylation product upon zinc-mediated chemical mutation of H3-K64Dha. Top: LC/MS of a typical reaction mixture. Bottom: Tryptic digest followed by LC-MS/MS reveals peptides modified with one and two equivalents of isopropyl upon chemical mutation with isopropyl iodide.

The fact that a single leucine product could be observed upon reaction with SBL-156Dha indicates some dependence on protein structure on the lifetime or accessibility of the Cα radical[34,35]. Further, in the case of SBL, the abundance of the dialkylation product was also shown to depend on the alkyl halide precursor employed. For example, primary alkyl halides tested on SBL156Dha gave dialkylation products similar to those observed on histone substrates, while secondary and tertiary halides gave exclusively mono-addition products. *On histone substrates, in contrast, all radical precursors were observed to result in di-alkylation.* As the histone proteins were chemically modified under denaturing conditions (to avoid aggregation of the free histones in solution[36]), this result can be taken as representative of a protein reacted in a 'random coil' state, which one would predict would have the greatest accessibility of the Cα radical. This fact implies that selectivity for mono-addition may be possible for selected protein structures or hindered residue locations, even for this 1st-generation system (see below).

Identification of improved conditions for metal-mediated 'chemical mutation'

The observations, described above, of both oxidative cleavage and disubstitution, provided compelling chemical evidence for the intermediacy of the proposed captodative Cα radical in the zinc-powder mediated reaction. Unfortunately, attempts to directly observe this protein-bound radical by EPR spectroscopy were unsuccessful, presumably due its short lifetime. Attempted spin-trapping[37] with 4-hydroxy-TEMPO[38], a water-soluble derivative of the well-established TEMPO radical trapping agent, was also unsuccessful. However, clear reaction inhibition was observed, even with sub-stoichiometric (relative to zinc and iodide), providing strong support for the proposed radical mechanism. (see **Experimental Section**).

Based on the observed side-reactions and literature precedent from Luche[14] and others[19], the mechanism outlined as **Figure 3.13** was

proposed. This mechanism immediately leads to a number of potential interventions to enhance the desired reaction sequence while eliminating or disfavouring the unwanted side-reactions. In the initial zinc-mediated system, metal-mediated single-electron transfer is essential for both the initial generation of alkyl radicals and, critically, for the second electron transfer that generates an enolate[14] for rapid quenching by (presumably) the aqueous solvent (**Figure 3.13**). In principle, improving the efficiency of the second electron transfer would lead to more rapid quenching of the problematic Cα radical intermediate, thus decreasing the amount of di-alkylated product that can be formed from the unwanted radical-radical coupling reaction.

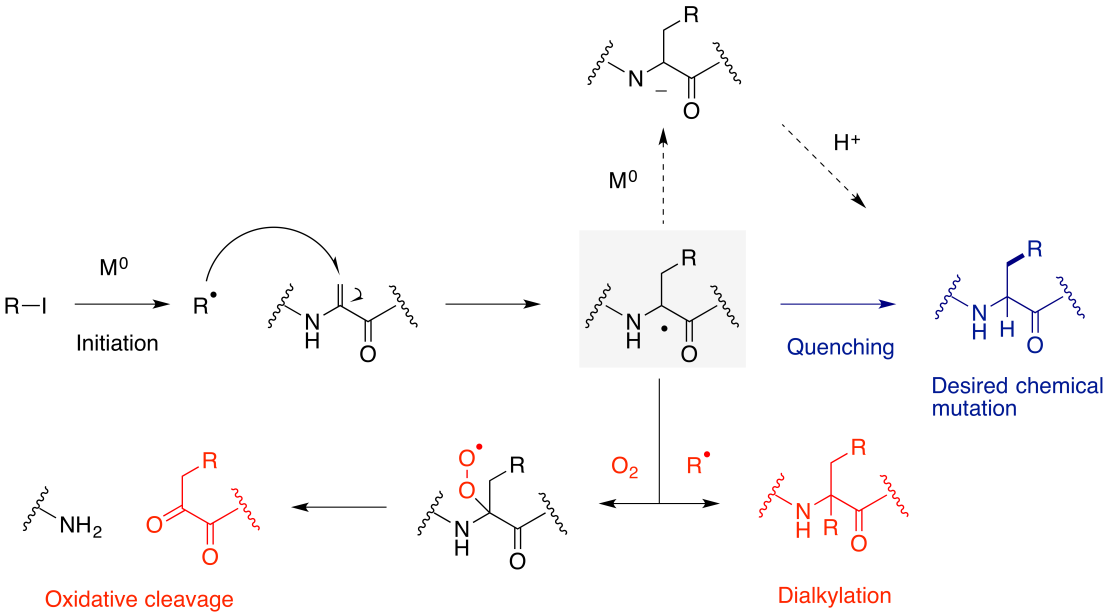


Figure 3.13: Proposed mechanism for metal-mediated chemical mutation, highlighting the quenching pathway (dark blue), mediated by a second SET event.

Investigation of redox potentials[39,40] for aqueous-compatible metal systems led to consideration of indium metal as an alternative candidate for mediating the radical chemistry in our system (standard reduction potentials E°/V, $Zn^{(2/0)}$ –0.76; $In^{(1/0)}$ –0.14; $In^{(2/1)}$ –0.40; $In^{(3/2)}$ –0.49). Indeed, there is literature precedent for the use of In^0 as a radical mediator in Luche-type chemistry[41,42]. Conditions utilized for indium-mediated reactions were adopted from the earlier work with zinc-powder, although initial experimentation suggested that lower equivalents of

indium powder (relative to zinc) were required for similar reaction outcomes, in keeping with the proposed enhanced rate of electron transfer. Pleasingly, tuning of the single-electron donor proved a partially successful strategy. Notably, the use of indium powder in place of zinc allowed successful 'chemical mutation' of SBL-Dha using certain primary iodides that had given large amounts of di-alkylation products when utilized in conjunction with zinc. Lysine and its methylated and dimethylated derivatives could thus be installed site-specifically into the SBL scaffold with complete conversion to the mono-alkylated derivatives (**Figure 3.14a**). This result represents the first 'post-expression' installation of these PTMs in their native form into any protein by chemical means. Even more promising was the first observation of quantitative mono-alkylation on a histone protein (H3-64Dha), leading to installation of ornithine and its trimethylated derivative (**Figure 3.14b**).

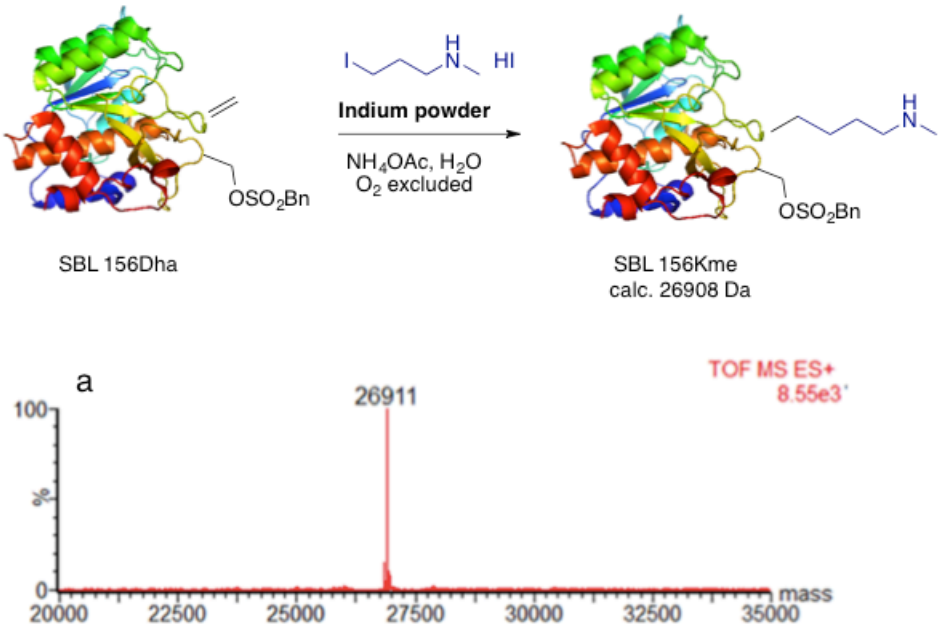


Figure 3.14a: Alteration of M^0 from Zn to In allows successful installation of methyllysine from SBL-156Dha.*Conditions:* SBL-156Dha (200 µL at 1.3 mg/mL, in 500mM NH_4OAc), 5000 eq. indium powder, 10k eq. (3-iodopropyl)methylamine hydroiodide, 30 minutes. *Reaction performed by Robert Vallee.*

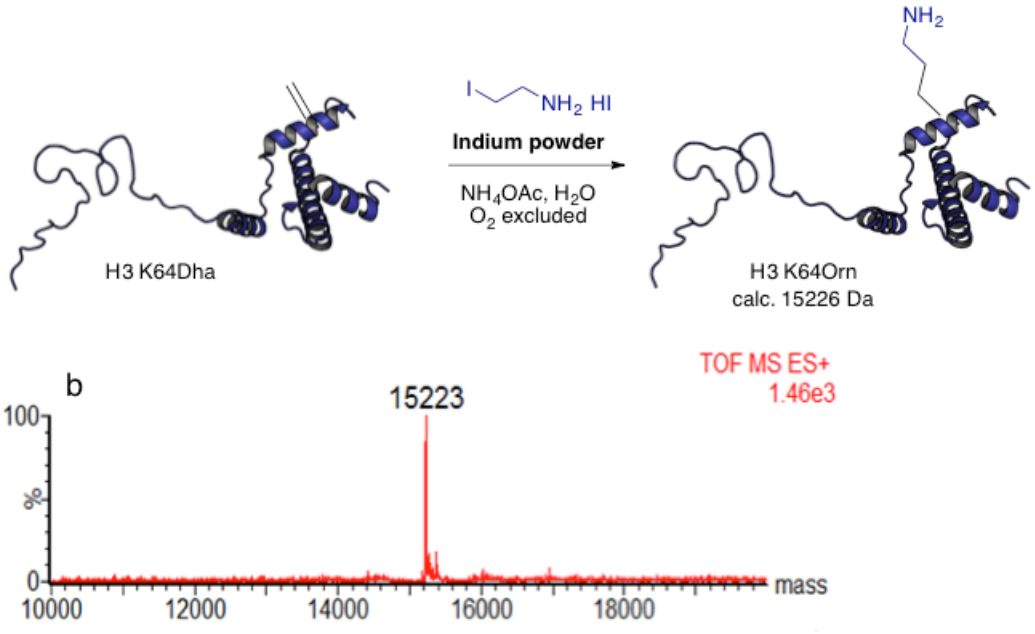


Figure 3.14b: Installation of ornithine to H3-K64Dha *via* indium-mediated chemical mutation. *Conditions:* H3-K64Dha (200 uL at 1 mg/mL, 500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0), 5000 eq. indium powder, 5000 eq. (2-iodoethyl)amine hydroiodide, 1 hr.

Installation of citrulline into histone proteins via indium-mediated chemical mutagenesis

The success of reactions incorporating a (2-iodoethyl)amine motif prompted us to explore the installation of citrulline into histone proteins. Citrulline[43] is a widely observed post-translational modification derived from the enzymatic deimination of arginine by protein arginine deiminase (PADI)4[44,45] enzymes (**Figure 3.15a**). This transformation results in the loss of the positive charge imparted by arginine, which can have important consequences for protein-protein and other molecular interactions mediated by electrostatic contact[46]. In terms of cellular function, protein citrullination has been noted to be an important process in the control of gene expression in the developing embryo[46]. Another context in which citrullination occurs is in auto-immune disease[47,48]. In conditions such as rheumatoid arthritis, extensive citrullination of proteins such as can result in generation of antibodies that target the self-protein.

The iodo-precursor of citrulline, (2-iodoethyl)urea **1** is not commercially available and thus required synthesis. Attempts to synthesise this compound by standard transformations such as the Appel reaction of (hydroxyethyl)urea or the Finkelstein reaction of (2-chloroethyl)urea were unsuccessful. Although product formation was observed in all cases, cyclisation appeared favoured under basic conditions, or indeed with simple manipulations of the compound such as rotary evaporation. It was found that simple heating in the presence of aqueous HI (55%) gave the best result, although cyclisation was still observed as a side-product (**Figure 3.15b**). As every manipulation (chromatography, concentration by rotary evaporation) resulted in an increase in the undesired cyclized byproduct, upon successful reaction, the crude mixture was simply concentrated once and stored immediately in a -20°C freezer for use in protein modification (as the cyclized product contains no iodide it should be innocuous in the chemical mutation reaction).

Pleasingly, installation of citrulline into histone scaffolds was readily accomplished under In^0 mediated conditions (**Figure 3.15C**). Although at certain reagent loadings dialkylation was observed, conditions were readily found by titration of iodide reagents that resulted in the sole mono-alkylation product. Initially, reactions were performed on H3-K27Dha as substrate, H3R26C was expressed, purified and converted to dehydroalanine in order to install citrulline at a naturally occurring site of modification. Pleasingly, utilizing similar conditions to those optimized for the H3-27Dha site, citrulline could also be installed in this natural position. Reaction of a 66 μM solution (1 mg/mL, in pH 6.0, 500 mM NH_4OAc, 5M Gdn.HCl buffer) H3-R26Dha with 2700 equivalents indium powder and 10,000 equivalents (2-iodoethyl)urea resulted in facile conversion to the desired H3-R26cit histone. Citrullination at this residue has been implicated in estrogen receptor activation[49]**;** the chemical procedure reported here provides the first opportunity to further investigate the biology of this modification in *in vitro* biochemical experiments.

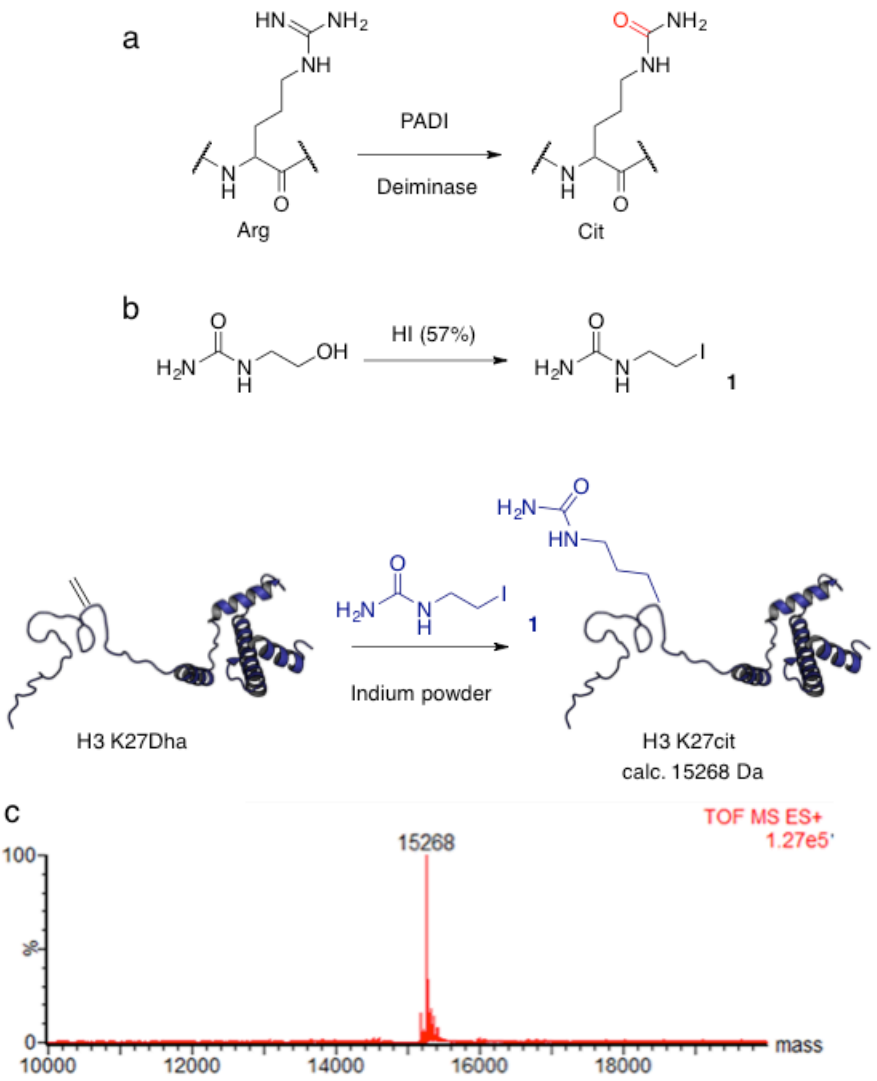


Figure 3.15: Indium-mediated chemical mutation enables the site-specific installation of citrulline (cit) to histone H3. a. In nature, citrulline is formed by deimination of protein arginine residues by PADI enzymes. b. Synthesis of 2-(iodoethyl)urea **1** c. Installation of H3-K27cit from H3-K27Dha *via* indium-mediated chemical mutation. *Conditions:* H3-K27Dha (200 µL at 1 mg/mL, 500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0), 2700 eq. indium powder, 10k eq. (2-iodoethyl)urea, 30 min.

Challenges in metal-mediated 'chemical mutation' and outlook

In this chapter, important progress towards the goal of a generalized chemical procedure for post-translational mutagenesis has been described. Investigation of the previously observed protein fragmentation reaction led to the conclusion that radical oxidative cleavage was occurring at the introduced dehydroalanine residue, subsequent to the initial radical attack. This side-reaction could be eliminated through the use of a positive-pressure anaerobic glovebox environment for the reaction. Further, an additional di-alkylation side-reaction was identified and characterized. It was discovered that this reaction also took place at the dehydroalanine site and results from 'trapping' of an initial Cα radical with a further equivalent of alkyl radical. Important variation in the amount of this side-product was observed between different protein scaffolds and between different iodide alkyl donor reagents. A consideration of the likely reaction mechanism led to the identification of indium powder as an improved metal mediator of the chemical mutation. Notably, these improved conditions allowed the first chemical installation of the native Kme and Kme_2 PTM residues to SBL, and the installation of citrulline to histone H3.

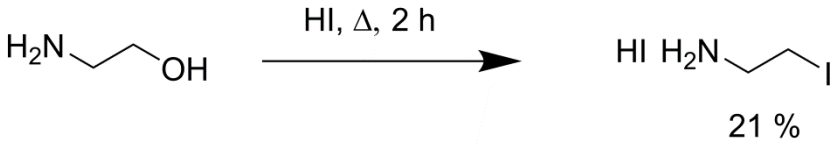
Despite this progress, important obstacles to a truly general chemical mutagenesis reaction protocol remained. Notably, even with the improved reaction conditions using indium metal, di-alkylation products were still observed on histone H3-Dha for the majority of tested iodides, at all reagent loadings tested. Further, very high reagent loadings were necessary for successful reactions, resulting in excesses of metal and iodide that were not compatible with all proteins. Indeed, initial tests with the camelid single chain antibody cAb-Lys resulted in protein destruction (as ascertained by SDS-PAGE and LC-MS) for all metal-mediated conditions attempted[50]. These severe limitations led to the further investigations described in the next chapter.

Experimental Section: Chapter 3

Reagent synthesis

(2-iodoethyl)amine hydroiodide

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in[5]



A solution of ethanolamine (1.00 g, 16.4 mmol) in hydroiodic acid (55 %, 8.0 mL) was heated to 125 ºC and stirred for 2 h. The reaction mixture was cooled to room temperature then on ice to form a dark red residue, which crystallised. The crystals were filtered then recrystallised from $CHCl_3$/EtOH to afford the title product as white flakes (1.02 g, 3.44 mmol, 189-190 ºC, 21 %).

^{1}H NMR *δH* (400 MHz, MeOD) 3.46-3.37 (4H, m, CH_2CH_2)

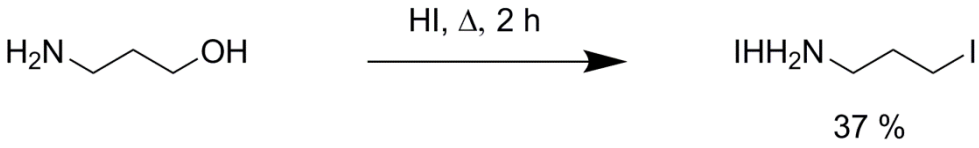
^{13}C NMR *δC* (100 MHz, MeOD) 42.3 (CH_2NH_2) and -3.21 (CH_2I)

IR (υ_{max}/cm^{-1}/KBr) 2995, 1552, 1488, 1435, 1378, 1314, 1206, 1057, 1014, 910 and 860

HRMS m/z (ESI$^+$): Found 171.9615 $[M+H]^+$; C_2H_7NI requires 171.9618.

(3-iodopropyl)amine hydroiodide

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in[5]



To ice cold hydroiodic acid (55 %, 13.0 mL), 3-aminopropa-1-ol (2.0 mL, 26.6 mmol) was added. The reaction mixture was heated to 125 ºC and stirred for 2 h. The reaction mixture was cooled to room temperature to form a dark red residue, which was recrystallised twice from EtOH/$CHCl_3$ to give the title compound as colourless prisms (3.06 g, 9.77 mmol, 163-164 ºC, 37 %).

^{1}H NMR *δH* (400 MHz, MeOD) 3.33 (2H, t, *J* = 7.3, CH_2I), 3.09 (2H, t, *J* = 7.3, CH_2NH_2) and 2.20 (2H, quin, *J* = 7.3, CH_2CH_2I)

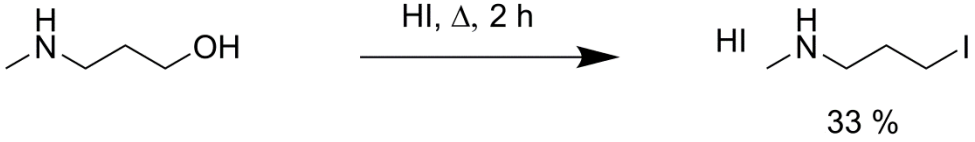
^{13}C NMR *δC* (100 MHz, MeOD) 40.6 (CH_2NH_2), 31.2 (CH_2CH_2I) and 0.1 (CH_2I)

IR (υ_{max}/cm^{-1}/KBr) 2991, 2288, 1860, 1582, 1588, 1309, 1194, 1120 and 1059

HRMS m/z (ESI$^+$): Found 185.9774 $[M+H]^+$. C_3H_9NI requires 185.9775.

(3-iodopropyl)methylamine hydroiodide

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in [5]



33 %

A solution of 3-(methylamino)propan-1-ol (2.00 g, 22.4 mmol) in hydroiodic acid (55 %, 12.0 mL) was heated to 125 ºC and stirred for 2h. The reaction mixture was cooled to room temperature then to -20 ºC overnight, which allowed the dark red residue to crystallise. The crystals were filtered, washed with Et_2O (100 mL) and then recrystallised ($CHCl_3$/EtOH) and washed with Et_2O (100 mL) to afford the title product as yellow needles, which were dried under reduced pressure (2.41 g, 7.39 mmol, 53-55 ºC, 33 %).

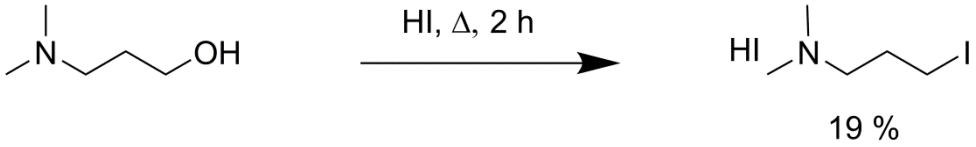
^{1}H NMR *δH* (400 MHz, MeOD) 3.33 (2H, t, *J* = 6.8, CH_2I), 3.16 (2H, t, *J* = 7.8, CH_2NHMe), 2.76 (3H, s, $NHCH_3$) and 2.24 (2H, m, CH_2CH_2I)

^{13}C NMR *δC* (100 MHz, MeOD) 50.1 (CH_2NHCH_3), 33.1 ($NHCH_3$), 30.0 (CH_2CH_2I) and -0.40 (CH_2I)

IR (υ_{max}/cm^{-1}/KBr) 3480, 2935, 2760, 2398, 1457, 1416, 1268, 1203, 1038

HRMS m/z (ESI$^+$): Found 199.9928 $[M+H]^+$; $C_4H_{11}NI$ requires 199.9930

(3-iodopropyl)dimethylamine hydroiodide



19 %

3-dimethylamino-1-propanol (2.0 g, 19.39 mmol) was added to a 25 mL round bottom flask and stirred at room temperature. Hydroiodic acid (55%, unstabilised, 12 mL) was added dropwise to the stirred alcohol, giving a dark-red mixture. The reaction mixture was heated at 125 °C (open air) for 2 hours and then cooled to room temperature. The flask was placed on ice and the product crystallised from the reaction mixture, along with formation of a red solid mass. The solids were filtered and washed with Et_2O. The product was further recrystallised from $CHCl_3$. Remaining iodine was removed by extensive trituration with Et_2O to leave a white solid (1.4 g, 18.2%).

IR (υmax, KBr): 2918, 2686, 1459, 1313, 1187, 1101, 1036, 1004, 959, 843, 727 ^{1}H NMR (400 MHz, MeOD): δ= 2.29 (2H, m, C$\underline{H_2}$CH$_2$I), 2.95 (6H, s, N$\underline{Me_2}$), 3.28-3.33 (4H, m, contains C$\underline{H_2}$NMe$_2$ and C$\underline{H_2}$I). ^{13}C NMR (100 MHz, MeOD): δ = -1.69 (C$\underline{H_2}$I), 27.99 (C$\underline{H_2}$CH$_2$I), 42.39 (N$\underline{Me_2}$), 58.21 (C$\underline{H_2}$NMe$_2$).

LRMS m/z (ESI+): Found 214.0 [M+H]+; $C_5H_{13}IN$ requires 214.07

Protein chemical modification

SBL-156Dha-PMS

Experiments with SBL were performed with Rafal Wiewiora as part of his part II project.

Amino acid sequence of SBL-S156C (BNP' numbering, PDB code for the wild type: 1GCI)

AQSVPWGISRVQAPAAHNRGLTGSGVKVAVLDTGISTHPDLNIRGGASFVPGEP
STQDGNGHGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASGSGSVSSIAQGL
EWAGNNGMHVANLSLGSPSPSATLEQAVNSATSRGVLVVAASGN**C**GAGSISYPA
RYANAMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQSTYPGSTYASLNGTSM
ATPHVAGAAALVKQKNPSWSNVQIRNHLKNTATSLGSTNLYGSGLVNAEAATR

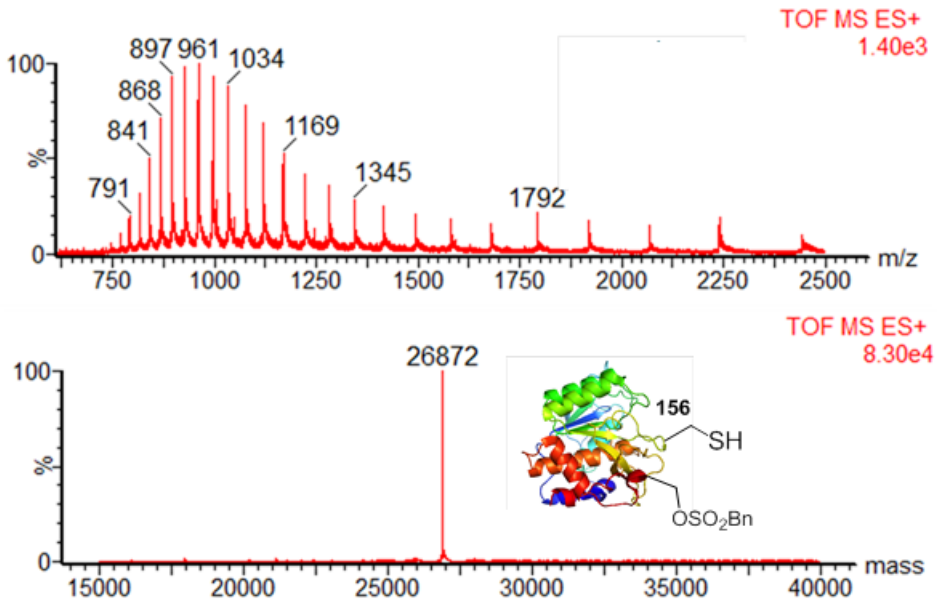
Purification of SBL S156C –PMS[51]

To the crude SBL S156C (5 ml), 400 µl DTT (1 M, in H_2O) was added to give a final concentration of 0.08 M. The solution was shaken at 4 °C for 30 min. Stabilizers from the crude protein preparation were removed by desalting the samples through two PD-10 desalting columns (GE Healthcare, buffer: 20 mM NaOAc, 5 mM $CaCl_2$, pH = 5.2) to yield 7 ml protein extract. A Sepharose HiTrap SP column (5 ml, GE Healthcare) was equilibrated with binding buffer (20 mM NaOAc, 5 mM $CaCl_2$, pH = 5.2). Protein was loaded onto the column and eluted over 20 CV (0-100%) elution buffer (20 mM NaOAc, 200 mM NaCl, pH = 5.2). Fractions were analysed by SDS-PAGE. The protein samples were then denaturated with 50% HCl (0.1 M), incubated for 15 minutes on ice and then diluted with 50 % Laemmli buffer.

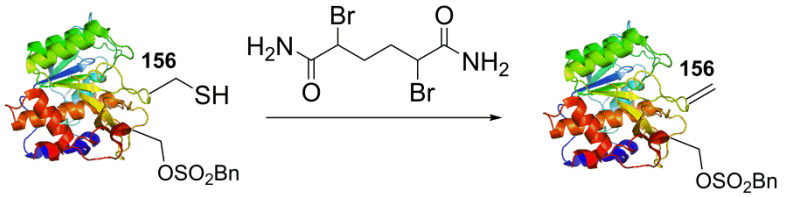
Later, SBL C156 was deactivated by addition of PMSF (phenylmethanesulfonyl fluoride, 100 mM in ethanol, final concentration of 1 mM PMSF in the mixture). The pooled fractions were concentrated using VivaSpin sample concentrators (10 kDa MWCO, 4 °C) and dialyzed against 50 mM sodium phosphate buffer pH 8.0 at 4 °C. After the first purification step, SBL C156-PMS was loaded on a HiLoad 26/60 Superdex 200 prep grade column and eluted with phosphate buffer (50 mM, pH 8.0).Fractions were analysed by SDS-PAGE and mass spectrometry.

ESI-MS for SBL S156C (221PMS) is shown below:

Calculated mass: 26869
Observed mass: 26872



Dehydroalanine formation from SBL 156C-PMS



SBL S156C (221PMS) **SBL S156Dha (221PMS)**

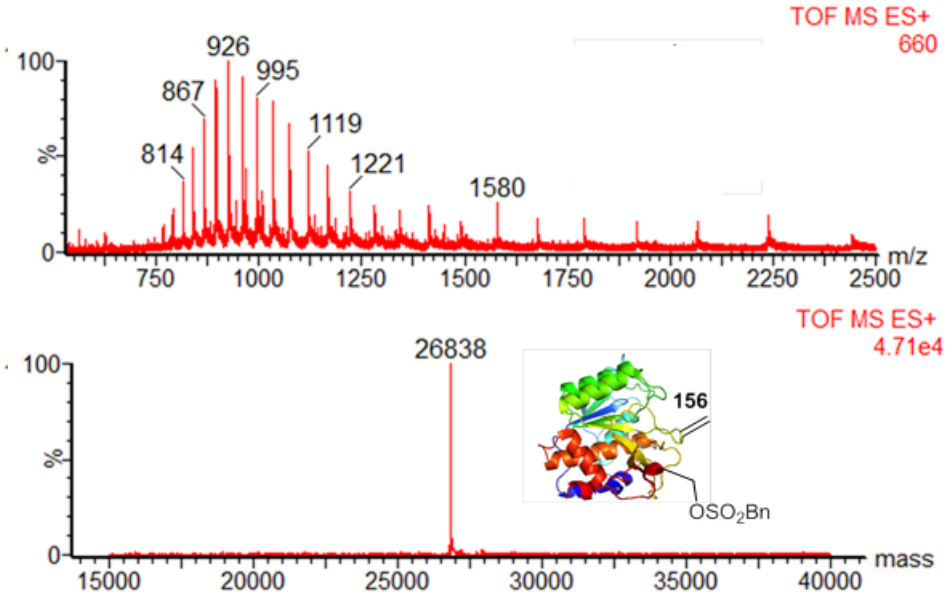
SBL S156C (2.0 mg/mL was treated with DTT (50 equiv., 0.5 M in water) and the solution was shaken for 20 minutes at room temperature. The reduced protein was purified by size exclusion chromatography using PD G25 columns. The protein concentration was checked by NanoDrop before 2,6-dibromohexanediamide (DBHDA, 500 equiv., 0.5 M in DMF) was added. The reaction was shaken for two hours at 37 °C and purified by size exclusion chromatography using PD G25 columns. The concentration of SBL C156Dha was determined by NanoDrop or BCA-assay and increased to 1.5 mg/mL using *VivaSpin* sample concentrators (10 kDa MWCO) if

required. The purified Dha-protein was flash-frozen and stored at – 80 °C until further modification.

ESI-MS for SBL S156Dha-PMS is shown below:

Calculated mass: 26835 (SBL 156Dha 221PMS)

Observed mass: 26838

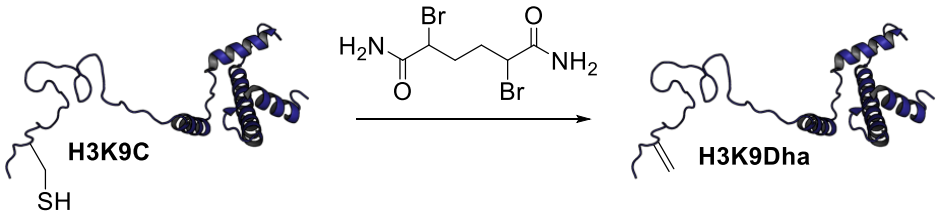


Histone H3 (K9C, C110A)

ARTKQTARCSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR
YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN
LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214
Observed mass = 15214

Histone H3 K9Dha

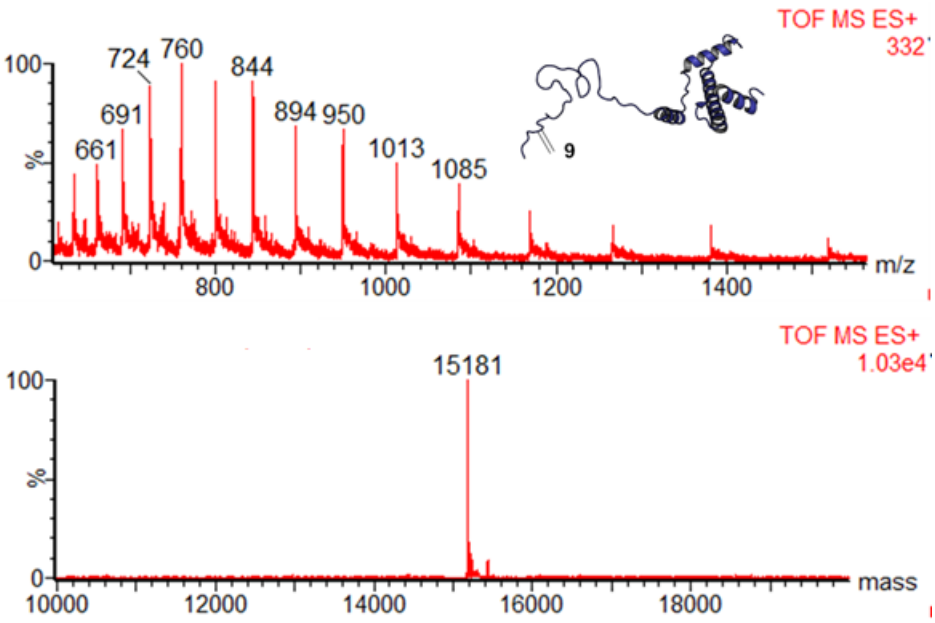


The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass = 15180.7

Observed mass = 15181



Histone H3 (K27C, C110A)

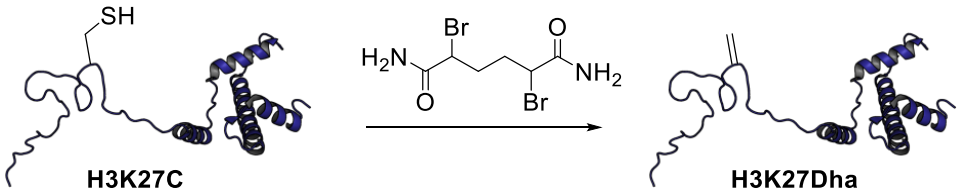
ARTKQTARKSTGGKAPRKQLATKAAR**C**SAPATGGVKKPHRYRPGTVALREIRR

YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN

LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214
Observed mass = 15214

Preparation of H3K27Dha

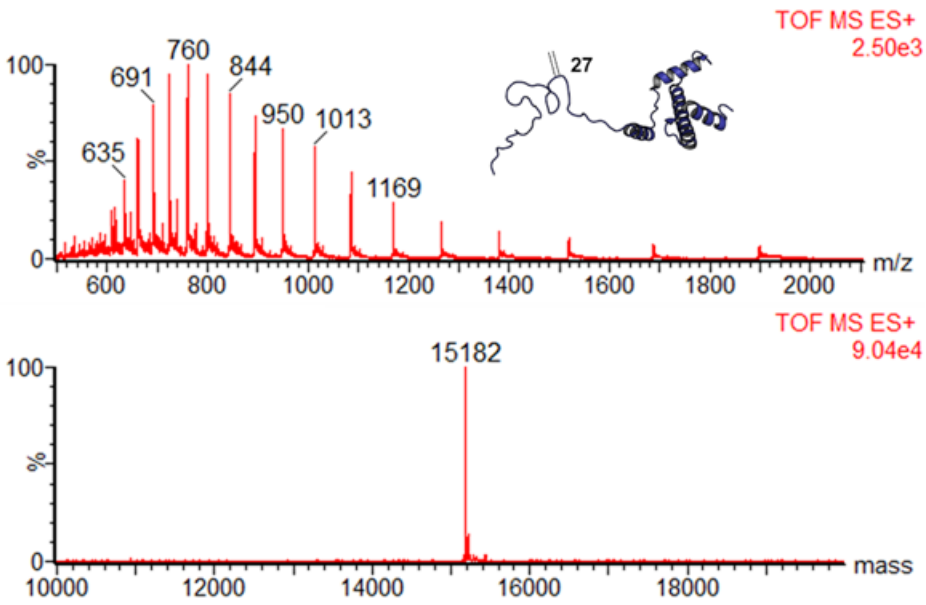


The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass: 15180.7

Observed mass: 15182



Histone H3 (K64C, C110A)

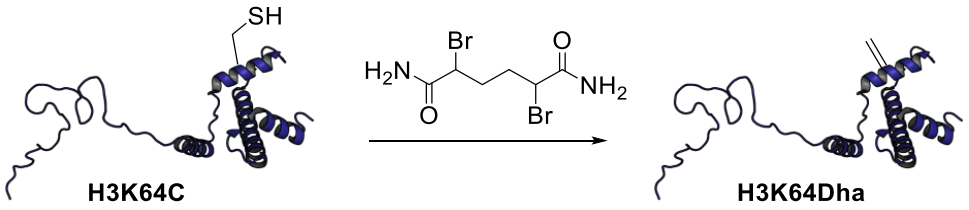
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR

YQKSTELLIR**C**LPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN

LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214
Observed mass = 15214

Preparation of H3K64Dha

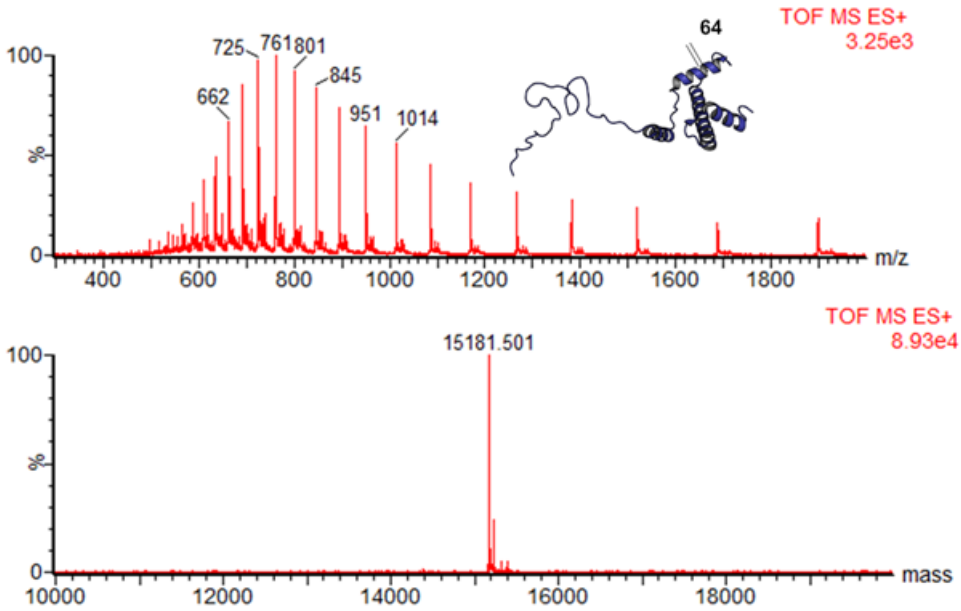


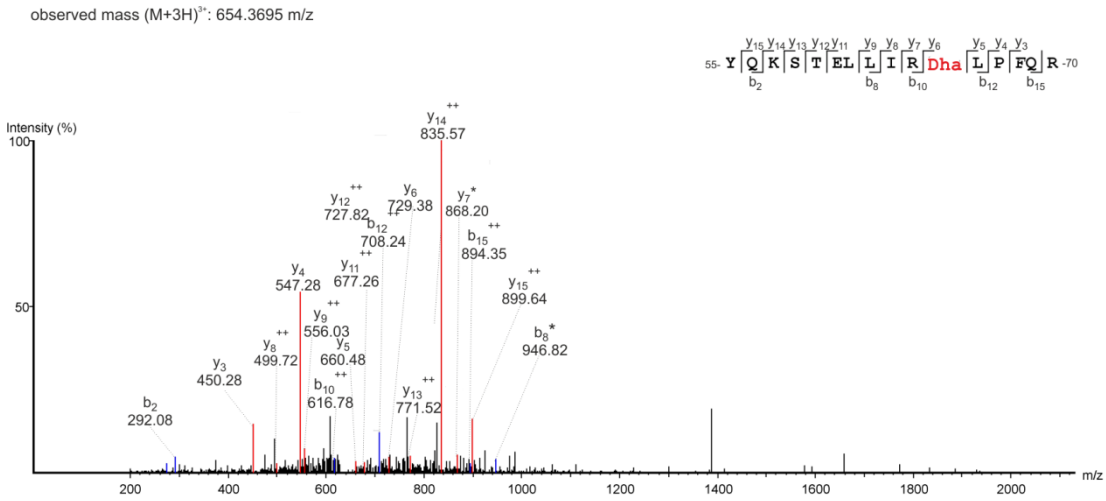
The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass: 15180.7 (H3 K64Dha ,C110A)

Observed mass: 15181.5

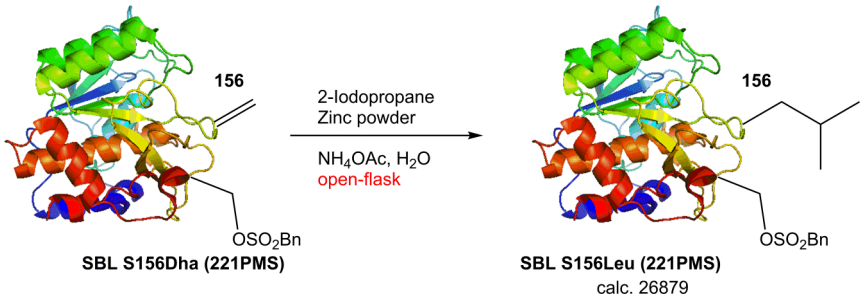




MS/MS spectrum for the peptide $_{55}$YQKSTELLIRDhaLPFQR$_{70}$

Initial observations

Initial attempt at chemical mutation on SBL156Dha (*Experiment performed by Rafal Wiewiora)*



200 µl of a 0.4 mg/ml solution of SBL-156Dha-PMS in NH4OAc buffer (500 mM, pH 6.0) was added to a 1.5 ml Eppendorf tube followed by zinc powder (2 mg, 10000 molar equiv. relative to protein) and 2-iodopropane (2 µl, 7000 equiv. relative to protein). The resulting mixture was briefly vortexed and shaken vigorously at room temperature for 15 min, then briefly centrifuged (18k x g, 2 min) and analysed by LC-MS.

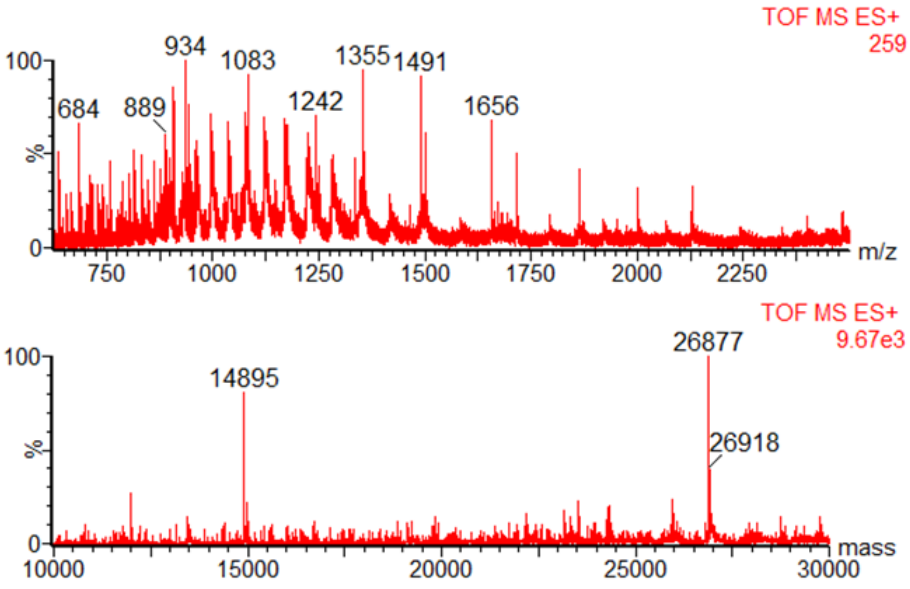
ESI-MS for the reaction mixture is shown below.

Calculated mass: 26879 (SBL156L (221PMS))

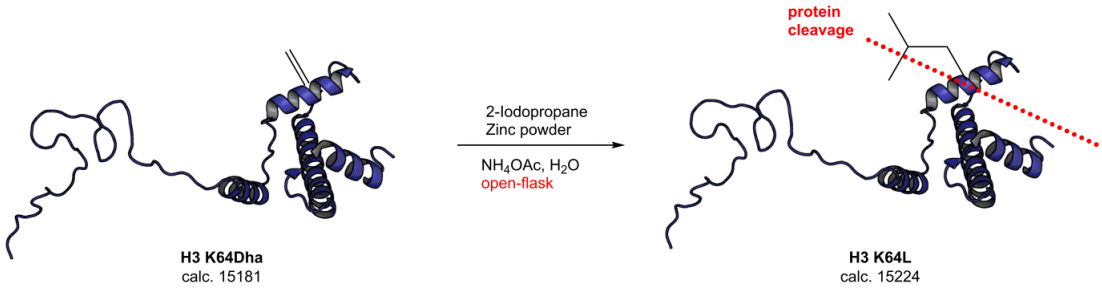
Observed mass: 26877

Calculated mass: 14896 (SBL N terminal fragment –Asn155)

Observed mass: 14895



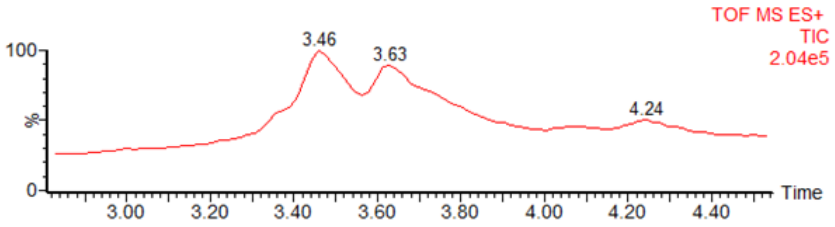
Cleavage of H3K64Dha



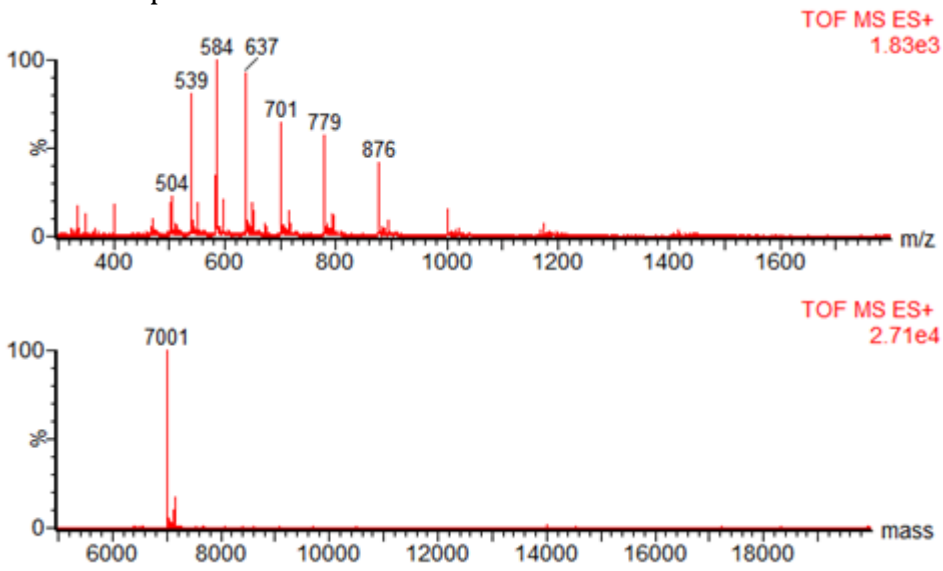
Zinc powder (4 mg, 4640 molar equiv. relative to protein) was added to a 2 mL plastic eppendorf tube. Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added, followed by 20 µL of 10 mg/mL protein stock solution (200 µg, in the same buffer) to give a final protein concentration of 1 mg/mL. 2-iodopropane (5 additions of 0.25 µL, a total of 1300 molar equiv. relative to protein) was added to the protein solution. The reaction was stirred for 30 minutes at room temperature (open air) and then an aliquot was removed for LC-MS analysis.

Using LC/MS method A (see **Chapter 2: General Experimental**), separation of the protein fragments was observed.
LC/MS spectrum for a typical reaction is shown below:

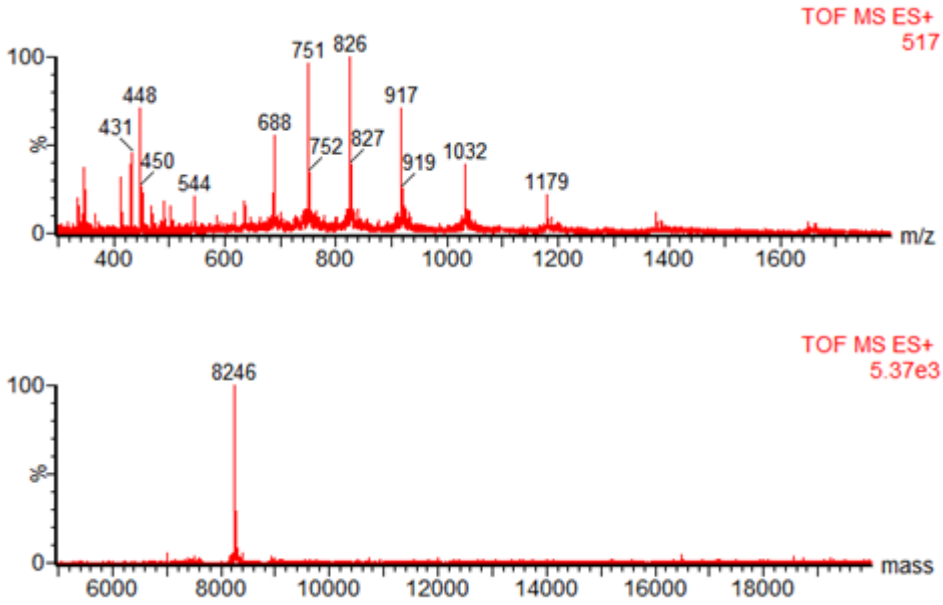
TIC chromatogram:



ESI-MS for peak 3.46 min



ESI-MS for peak 3.63 min



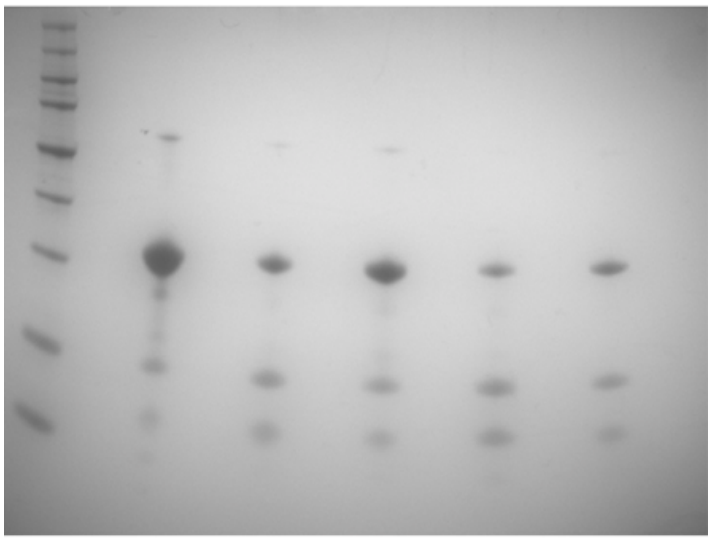
Protein cleavage in an open-flask is not dependent on iodide reagent *(Experiment performed by Rafal Wiewiora)***:**

Five 2 mL Eppendorf tubes, each containing a solution of SBL 156Dha (0.8 mg/mL, 200 µL, 500 mM NH_4OAc pH 6.0), were charged with zinc powder (4 mg per tube, 20,000 molar equiv. relative to each protein) and a small magnetic stir-bar. The iodide of interest was then dissolved in the protein solution. The reaction mixtures were vigorously stirred for 30 minutes at room temperature before being loaded onto a SpinTrap™ G-25 desalting column (pre-equilibrated with the desired buffer), eluted and analysed by LC-MS and SDS-PAGE gel.

Alkyl iodide	Molar equivalents relative to protein	Volume added
Cyclohexyl iodide	20,000	15.5 µL
Cyclopentyl iodide	20,000	13.9 µL
n-propyl iodide	30,000	17.6 µL
t-butyl iodide	15,000	10.7 µL
i-propyl iodide	7000	4 µL

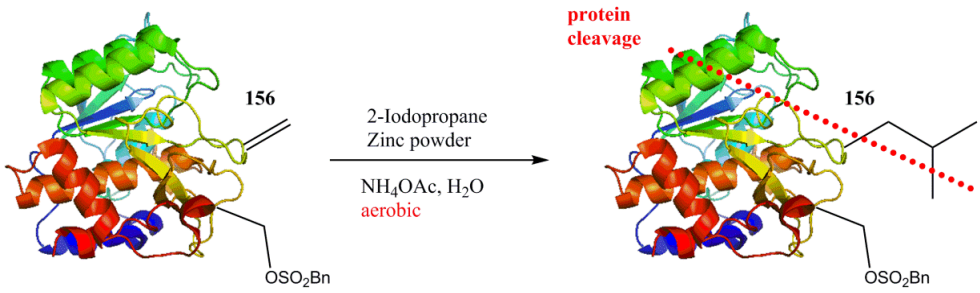
SDS-PAGE analysis:

The eluates obtained above were concentrated to approximately 10 µL using Vivaspin 500 (MWCO 5k) concentrators. The solutions were then treated with 0.8 M HCl at 0°C for 30 min to effect protein denaturation and subsequently diluted 2x with laemmli buffer (containing no β-mercaptoethanol). The samples were run on a SDS-PAGE gel (4-12% Bis-Tris), in 1x MES running buffer. The gel was stained using Coomassie Blue and visualised using a Bio-Rad ChemiDoc MP instrument.



Marker ***c*-hex** ***c*-pent** ***n*-Pr** ***t*-Bu** ***i*-Pr**

Oxygen sensitivity (*Experiment performed by Rafal Wiewiora)*

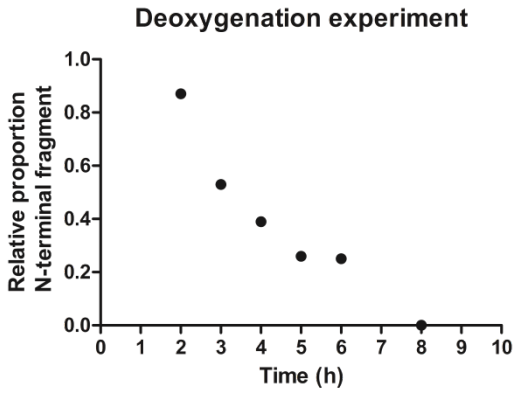


A 2mL Eppendorf tube containing 2 mg solid zinc powder (10,000 molar equiv. relative to protein) within a solution of SBL 156Dha (0.8 mg/mL, 200 µL, 500 mM NH_4OAc pH 6.0) was placed in a N_2 atmosphere glovebox via rapid port and left for the stated number of hours for equilibration with the anaerobic atmosphere.

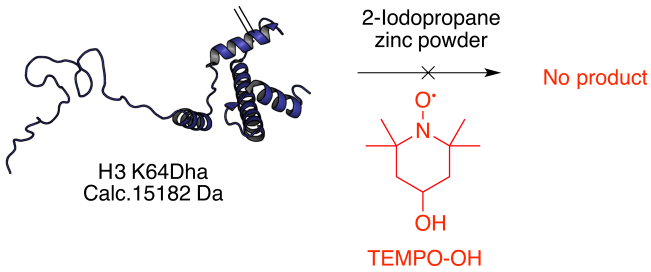
2-iodopropane (2 µL, 10,000 molar equiv. relative to protein) was then added. The reaction mixture was shaken for 30 minutes at room temperature before being loaded onto a SpinTrap™ G-25 desalting column (pre-equilibrated with the desired buffer), eluted and analysed by LC-MS.

In order to quantify protein cleavage, only the N-terminal fragment (calc. 14896) was considered as it results in higher ion intensities ('flys better',

hence improved signal:noise) than the corresponding C-terminal fragment. The heights of the peak were calculated by dividing the intensity of the peak by the total ion count. Both the main peak and a +16 [O] secondary peak were combined in this measurement. The result of this calculation is the percentage of ions resulting from the N-terminal fragment. Given that the reactions and LC-MS analysis were performed in parallel, with the same starting reagents and concentrations, the obtained value should be representative of the amount of N-terminal fragment in the reaction mixture and hence the amount of cleavage that has occurred. A similar approach has been previously employed in our laboratory and strict correlation between the MS intensity data and the true mixture composition was shown[18]. Indeed, the below results were also analysed using gel densitometry (not shown), giving values that were in agreement with the trend demonstrated.



TEMPO-OH inhibition



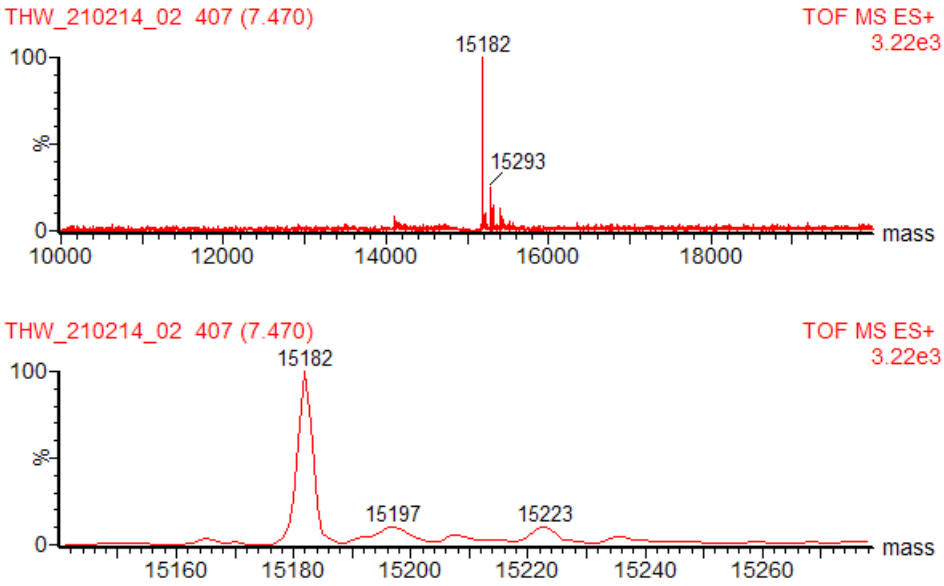
Zinc powder (4 mg, 4640 molar equiv. relative to protein) was added to a 2 mL eppendorf tube. Reaction buffer (500 mM NH4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added, followed by 20 µL of 10 mg/mL H3-K64Dha protein

stock solution (200 µg) to give a final protein concentration of 1 mg/mL. The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was left overnight, open to the inert atmosphere.

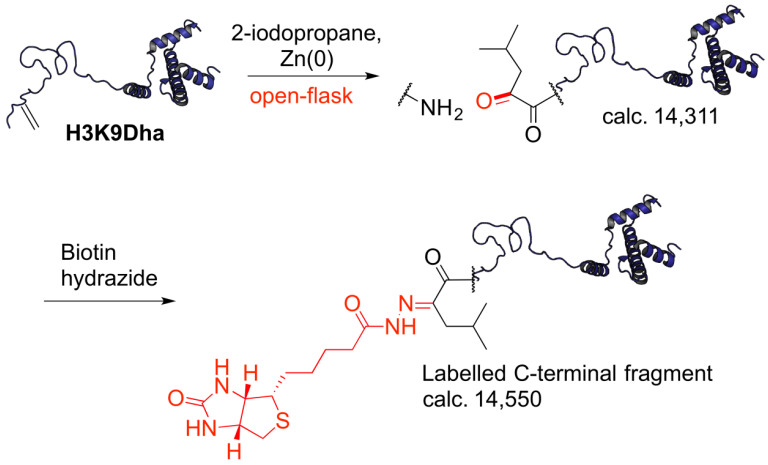
The following day, 4-hydroxy-TEMPO (500 equiv. relative to protein) was added to the protein solution within the glovebox. 2-iodopropane (2.6 µL, 2000 molar equiv. relative to protein) was then added to the protein solution, causing vigorous effervescence. The reaction was stirred, employing a magnetic stirrer and stir-bar, for 30 minutes at room temperature before the reaction mixture was loaded onto a SpinTrap™ G-25 desalting column, pre-equilibrated with the reaction buffer. The reaction mixture was eluted following the manufacturers instructions and then analysed by LC-MS method **A**.

Deconvoluted ESI-MS for the experiment is shown below:

Calculated mass: 15182 (H3 K64Dha, C110A)
15224 (H3 K64L, C110A) (**not observed**)

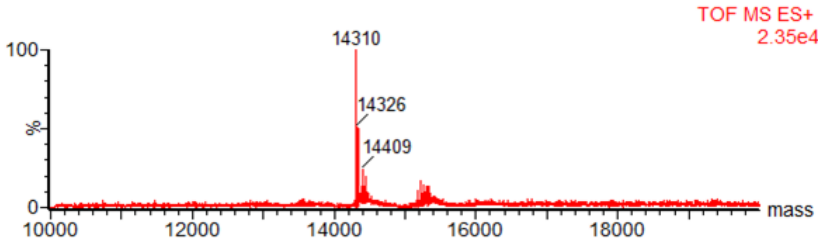


H3K9Dha cleavage and labelling to confirm the existence of *N*-terminal diketone on C-terminal cleavage fragment



Cleavage reaction H3K9Dha

Zinc powder (4 mg, 4640 molar equiv. relative to protein) was added to a 2 mL eppendorf tube. Reaction buffer (500 mM NH4OAc pH 6.0, 5M Gdn.HCl, 180 µL) was added, followed by 20 µL of 10 mg/mL protein stock solution (200 µg) to give a final protein concentration of 1 mg/mL. 2-iodopropane (1.3 µL, 1000 equiv. relative to protein) was added to the protein solution. The reaction was stirred for 30 minutes at room temperature (open air) and then an aliquot was removed for LC/MS analysis. LC/MS spectrum for a typical reaction is shown below:



Calc. C-terminal fragment = 14310.7

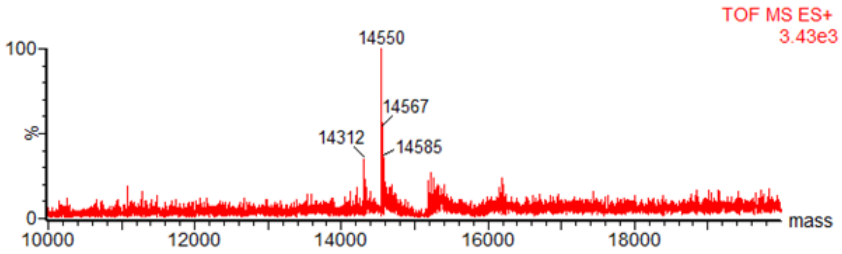
Found 14310

Based on proposed fragment sequence:

H3 9

LSTGGKAPRK QLATKAARKS APATGGVKKP HRYRPGTVAL REIRRYQKST ELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIH AKRVT IMPKDIQLAR RIRGERA

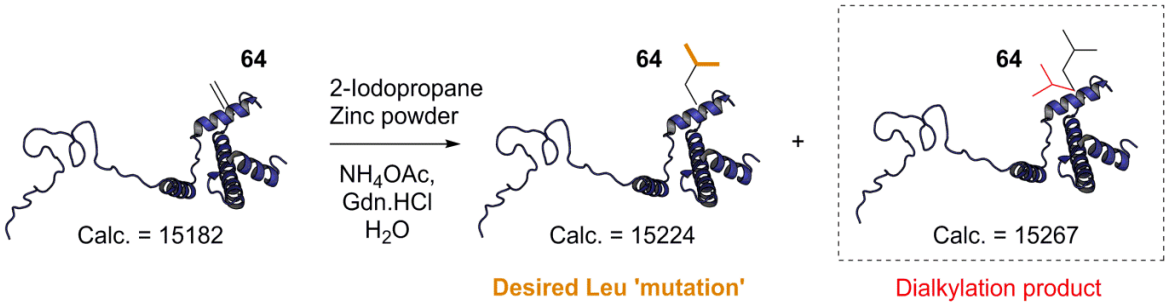
10 µL reaction mixture from the cleavage experiment outlined above (H3K9Dha) was added to 150 µL of pH 8 labelling buffer (100 mM sodium phosphate, 3M Gdn.HCl), to give a final predicted protein concentration of 6.6 µM. 40 µL of a stock solution of biotin hydrazide (50 mM to give a final biotin concentration of 10 mM) was added, along with *m*-phenylenediamine (2.2 mg, 100 mM final concentration) as nucleophilic catalyst. The resultant solution was incubated for 48 hours at 37 degrees, before analysis of an aliquot by LC-MS to confirm labeling:



Calculated mass (Biotin labelled fragment): 14550

Found 14550

Investigation of dialkylation side-product



Reaction procedure:

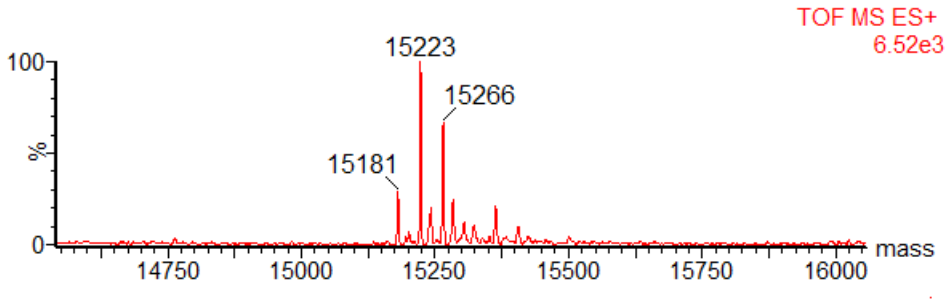
Zinc powder (4 mg, 4640 molar equiv. relative to protein) was added to a 2 mL eppendorf tube. Reaction buffer (500 mM NH4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added, followed by 20 µL of 10 mg/mL protein stock solution (200 µg) to give a final protein concentration of 1 mg/mL. The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was left overnight, open to the inert atmosphere.

The following day, 2-iodopropane (1.3 µL, 1000 molar equiv. relative to protein) was added to the protein solution, causing vigorous effervescence. The reaction was stirred, employing a magnetic stirrer and stir-bar, for 30 minutes at room temperature before the reaction mixture was loaded onto a SpinTrap™ G-25 desalting column, pre-equilibrated with the reaction buffer. The reaction mixture was eluted following the manufacturers instructions and then analysed by LC-MS, with remaining sample used for MSn analysis.

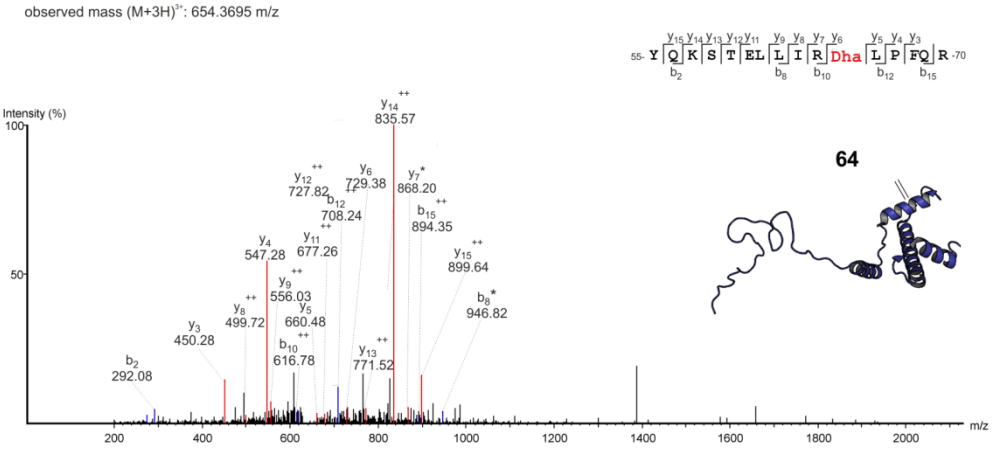
Deconvoluted ESI-MS for the experiment is shown below:

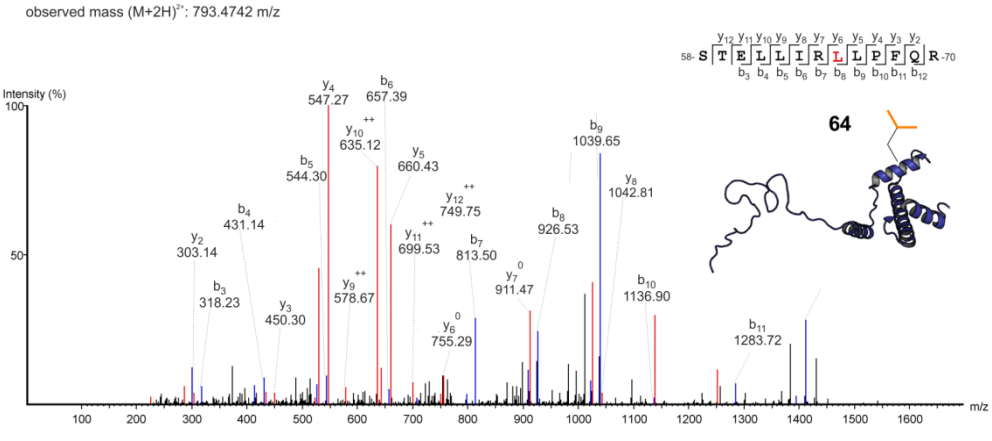
Calculated mass: 15182 (H3 K64Dha, C110A)
15224 (H3 K64L, C110A)
15267 (H3 K64L(iPr), C110A)
Observed mass: 15181 (H3 K64Dha, C110A)
15223 (H3 K64L, C110A)
15266 (H3 K64L(iPr), C110A)

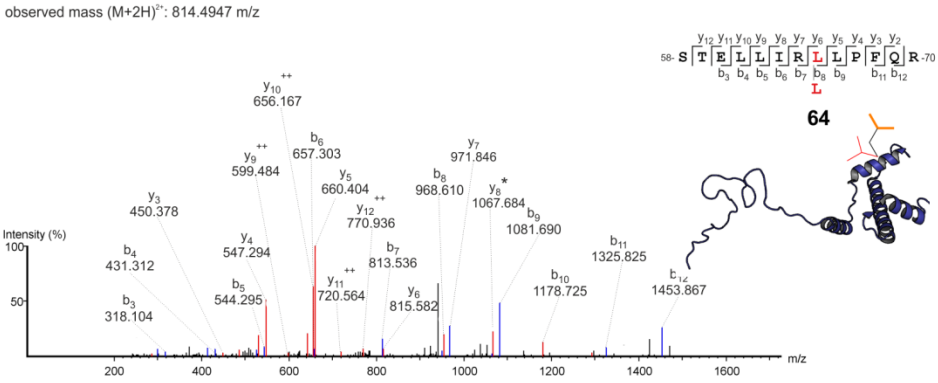
Deconvoluted MS:



MS² Data:







MS/MS spectrum of the peptides 55YQKSTELLIRLLPFQR70 and 58STELLIRLLPFQR70 derived from Histone H3 showing mutation of cysteine 65(64) to Dha (top panel), then Dha to Leucine and 'double Leucine'/L(iPr) (bottom two panels). The labelling of the fragment ions indicated for the peptide sequence are the following: b : N-terminal fragment ion; y : C-terminal fragment ion; ++: doubly charged fragment ion, unlabelled peaks represent singly charged fragments. b*,y*: fragment ion-NH3, b0, y0: fragment ion-H2O. b and y ions fragments are highlighted in blue and red, respectively. The measured masses are indicated for each assigned fragment ion.

General method for 'chemical mutagenesis' using Zn metal on histone H3:

Zn metal (4 mg, 4.64k eq. relative to protein) was added to a 2 mL plastic eppendorf. Reaction buffer (500 mM NH_4OAc, 5M Gdn.Hcl, pH 6.0, 180 µL) was added, followed by 20 µL of 10 mg/mL protein stock solution (200 µg) to give a final protein concentration of 1 mg/mL. The resultant protein solution was then added to an inert atmosphere glovebox *via* rapid port. The solution was left overnight, open to the inert atmosphere.

The following day, *i*-PrI (1.3 µL, 1k eq. relative to protein) was added to the protein solution. The reaction was stirred for 30 minutes at room temperature and then the reaction mixture was loaded onto a PDG25 spin-trap, pre-equilibrated with the reaction buffer. LC/MS spectrum for a typical reaction is shown below:

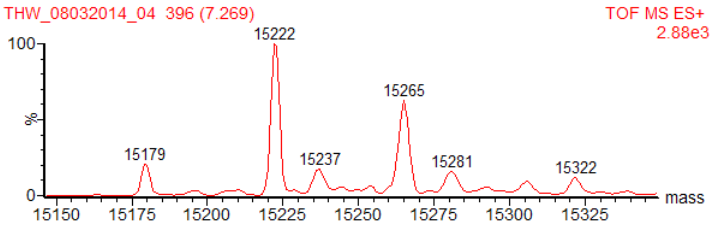
Attempted optimisation to decrease 'double addition' side reaction

(Representative series of reactions)

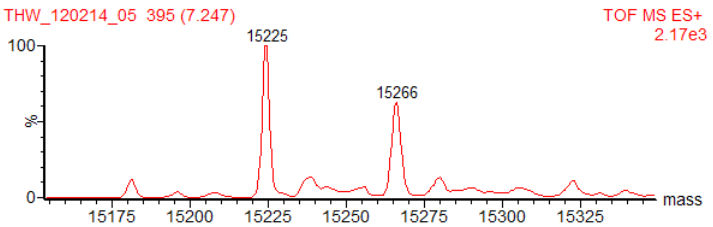
Entry	**Zinc** (eq. relative to protein)	**Isopropyl iodide** (eq.)	**Additive**	**Outcome** (s.m: mono: di-addition)
1	4.6k	1.2k	None	12:55:33
2	4.6k	2.4k	None	0:62:38
3	2.3k	2.4k	None	12:55:33
4	4.6k	6k isopropyl **bromide**	None	70:30:0
5	4.6k	3* 6k isopropyl bromide	None	70:30:0
6	4.6k	1.2k	CuI, 4.6k	35:43:22 (protein degradation

				observed)
7	4.6k	2.4k	1k $(TMS)_3SiH$ + 1mM β - mercaptoethanol	0:69:31
8	4.6k	2.4k	5k $(TMS)_3SiH$ + 1mM β - mercaptoethanol	13:61:26
9	4.6k	1.2k	1k Ethylpiperidine hypophosphite (EPHP)	0:76:24
10	4.6k	1.2k	2k EPHP	0:70:30
11	**4k In**	**1k**	**None**	**0:84:16**

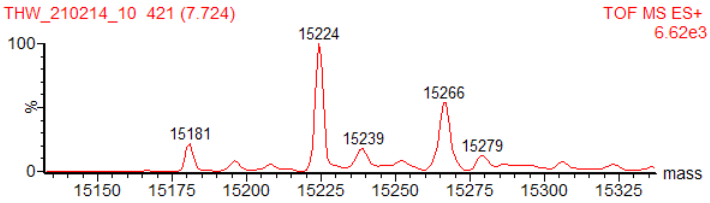
Entry 1



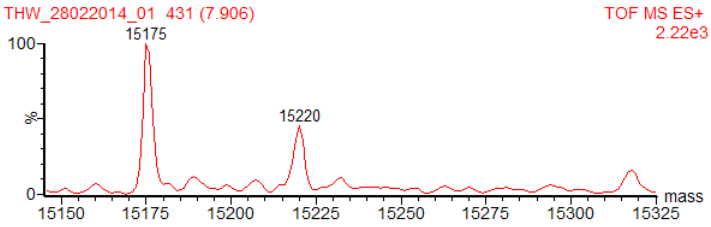
Entry 2



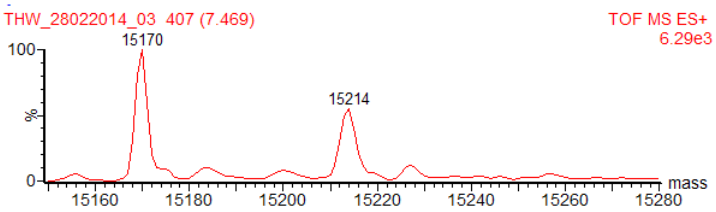
Entry 3



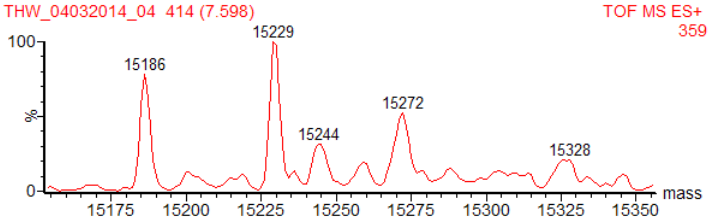
Entry 4



Entry 5

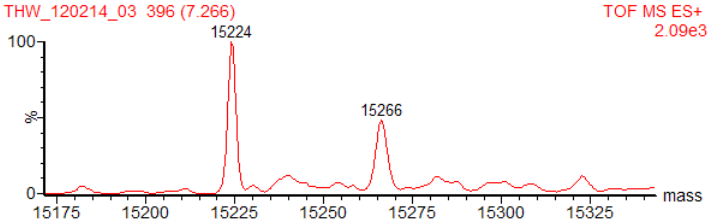


Entry 6

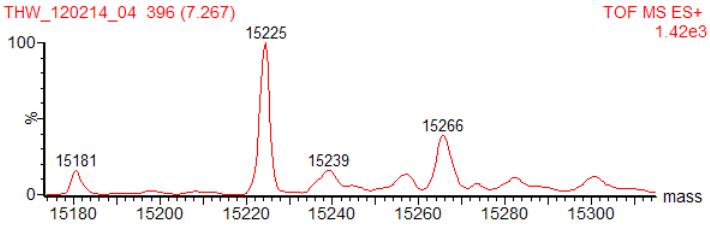


Note – calibration incorrect for above 3 e.g.s – doesn't alter calculation of peak ratios

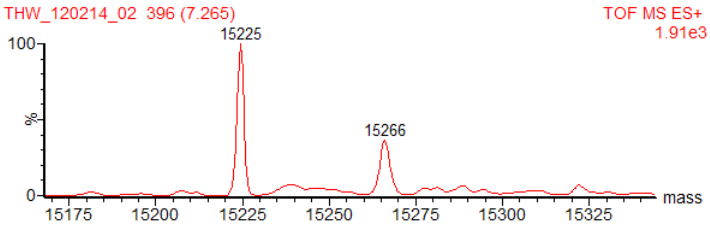
Entry 7



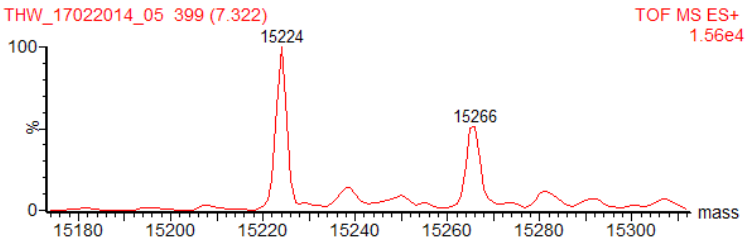
Entry 8



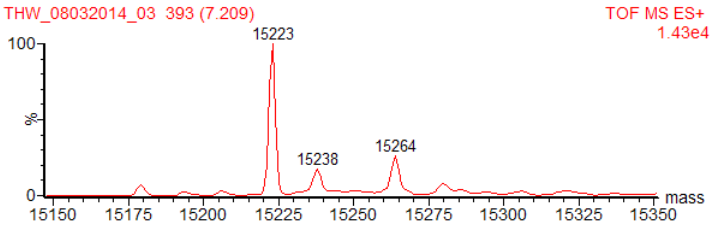
Entry 9



Entry 10



Entry 11

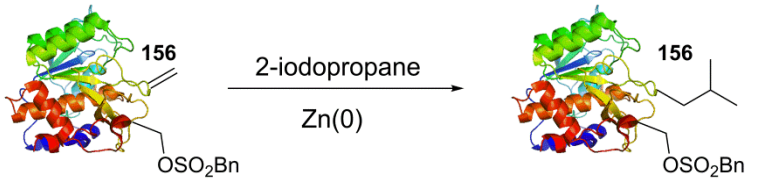


Successful chemical mutations

Chemical mutagenesis from SBL-156Dha-PMS

(SBL *experiments performed by Rafal Wiewiora or Robert Vallee)*

Installation of leucine from SBL-156Dha-PMS



SBL S156Dha (221PMS) **SBL S156L (221PMS)**

A solution of SBL S156Dha-PMS (200 µL at 0.4 mg/mL) in ammonium acetate buffer (500 mM, pH 6.0) was transferred to a N_2 atmosphere glovebox and left overnight at room temperature for equilibration with the anaerobic atmosphere.

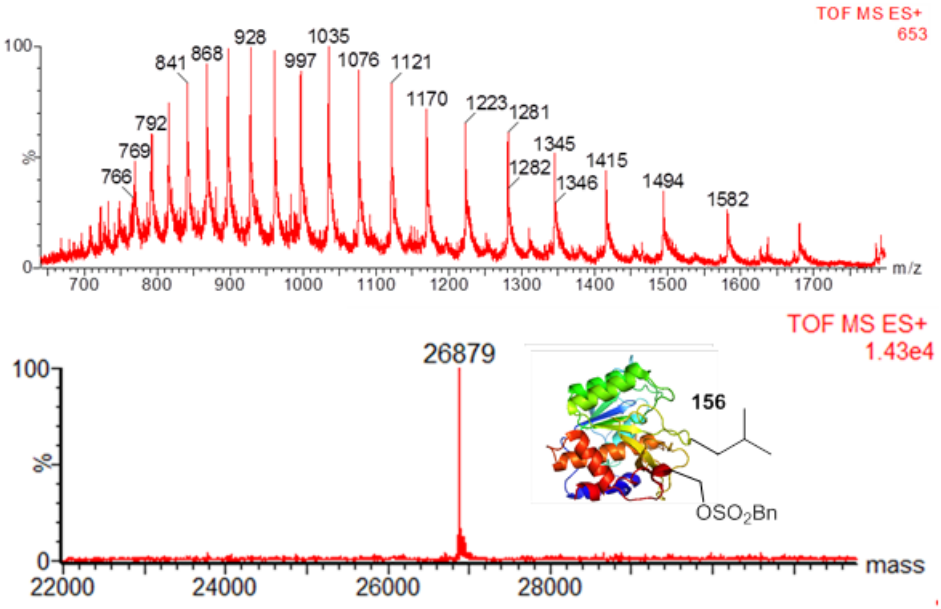
The following morning, within the glovebox, the degassed protein solution was combined with 2 µL of 2-iodopropane. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate eppendorf tube, containing 2 mg powdered zinc. Effervescence and the formation of a grey slurry are observed.

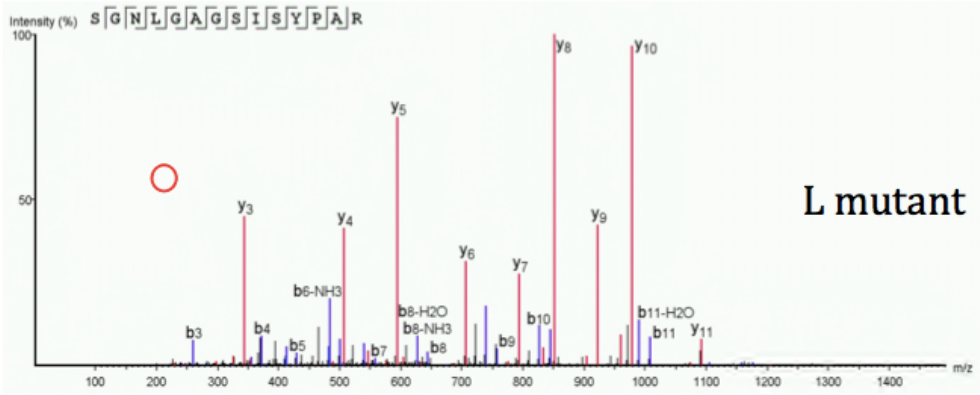
The reaction was stirred at room temperature for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS (protein mass spectrometry method c.).

ESI-MS for the modified protease is shown below:

Calculated mass: 26879 (SBL 156L 221PMS)

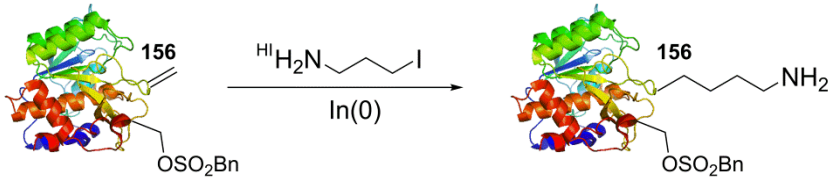
Observed mass: 26879





MS/MS spectrum of the peptide $_{151}$SGNLGAGSISYPAR$_{164}$

Installation of lysine from SBL-156Dha-PMS

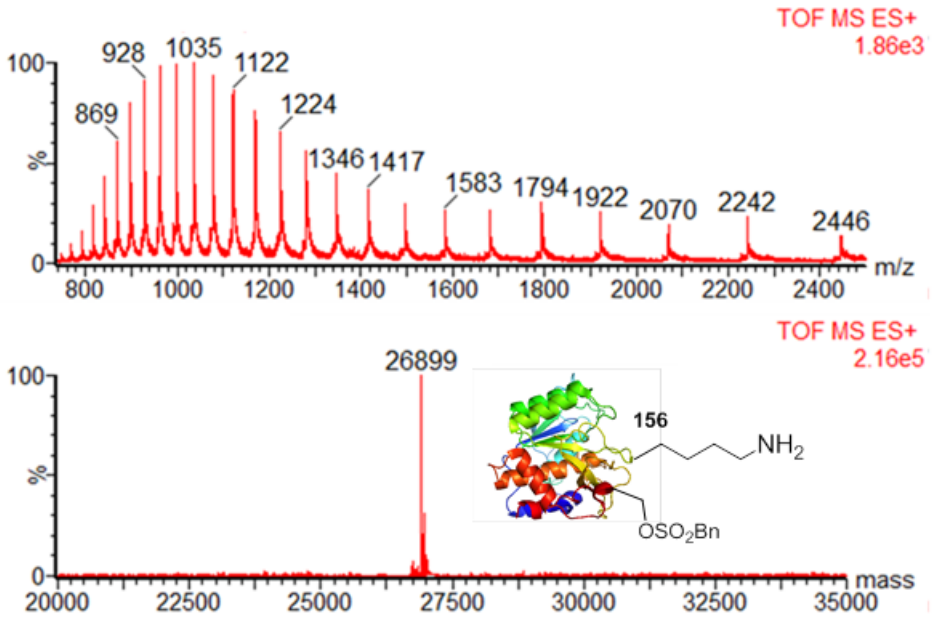


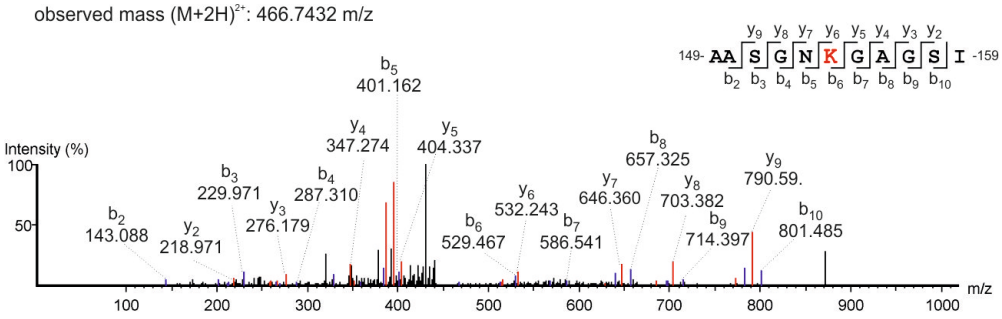
SBL S156Dha (221PMS) **SBL S156K (221PMS)**

A solution of SBL S156Dha-PMS (200 µL at 1.3 mg/mL) in ammonium acetate buffer (500 mM, pH 6.0) was transferred to a N_2 atmosphere glovebox and left overnight at room temperature for equilibration with the anaerobic atmosphere.

The following morning, within the glovebox, the degassed protein solution was combined with 30 mg of 3-iodopropan-1-aminium iodide **S1**. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate eppendorf tube, containing 5 mg powdered indium. Effervescence and the formation of a grey slurry were observed.
The reaction was stirred at room temperature for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

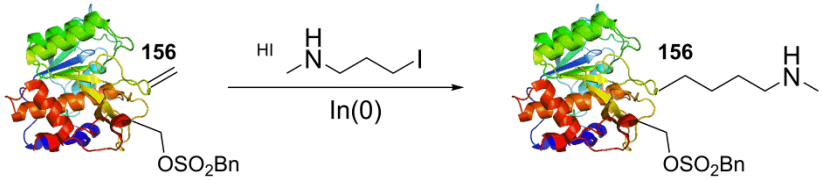
ESI-MS for the modified protease is shown below:
Calculated mass: 26894 (SBL 156K 221PMS)
Observed mass: 26899





MS/MS spectrum for the peptide $_{149}$AASGNKGAGSI$_{159}$

Installation of monomethyllysine from SBL-156Dha-PMS



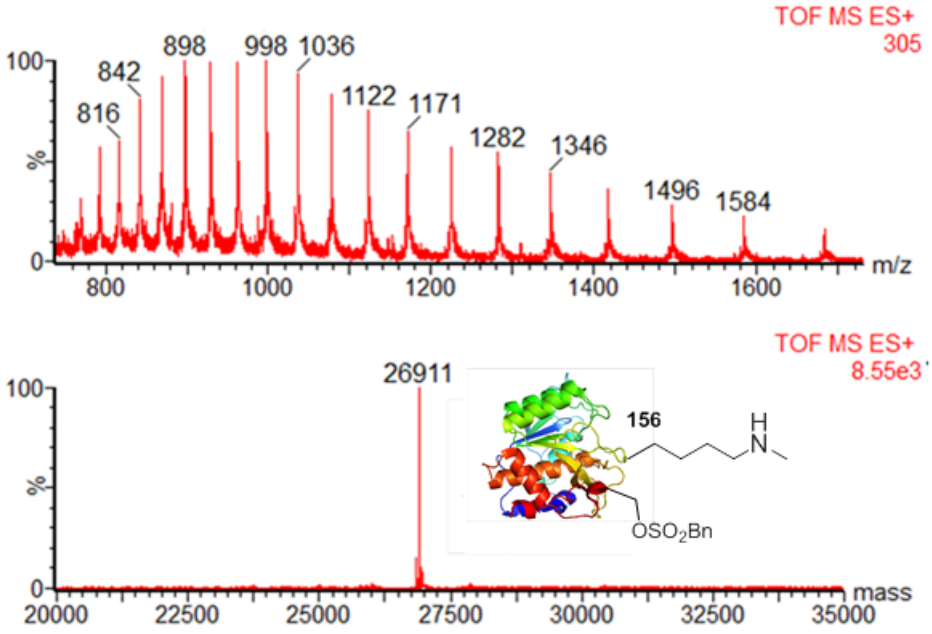
SBL S156Dha (221PMS) **SBL S156Kme (221PMS)**

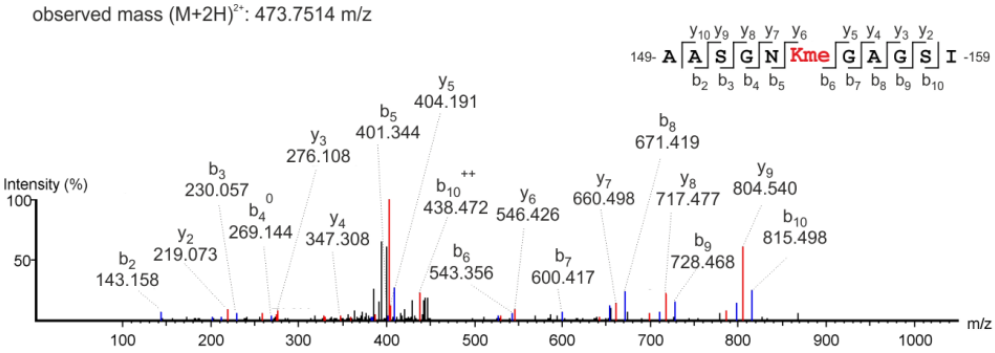
A solution of SBL S156Dha-PMS (200 µL at 1.3 mg/mL) in ammonium acetate buffer (500 mM, pH 6.0) was transferred to a N_2 atmosphere glovebox and left overnight at room temperature for equilibration with the anaerobic atmosphere.

The following morning, within the glovebox, the degassed protein solution was combined with 30 mg of (3-iodopropyl)methylamine hydroiodide **S4**. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate eppendorf tube, containing 5 mg powdered indium. Effervescence and the formation of a grey slurry were observed.

The reaction was stirred at room temperature for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

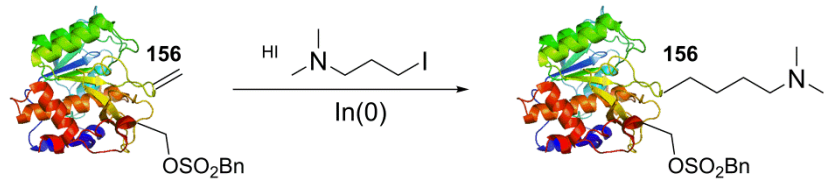
ESI-MS for the modified protease is shown below:
Calculated mass: 26908 (SBL 156Kme 221PMS)
Observed mass: 26911





MS/MS spectrum for the peptide $_{149}$AASGNKmeGAGSI$_{159}$

Installation of dimethyllysine from SBL-156Dha-PMS



SBL S156Dha (221PMS) **SBL S156Kme$_2$ (221PMS)**

A solution of SBL S156Dha-PMS (200 µL at 1.3 mg/mL) in ammonium acetate buffer (500 mM, pH 6.0) was transferred to a N_2 atmosphere

glovebox and left overnight at room temperature for equilibration with the anaerobic atmosphere.

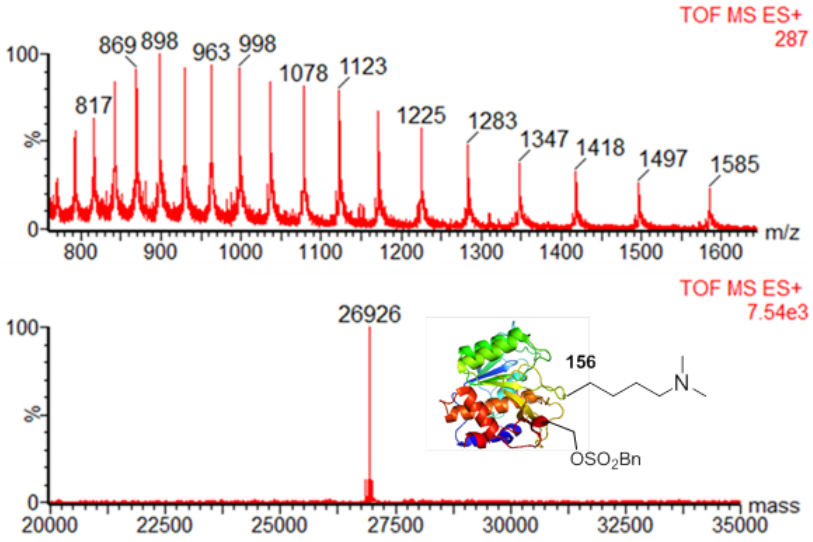
The following morning, within the glovebox, the degassed protein solution was combined with 33 mg of (3-iodopropyl)dimethylamine hydroiodide **S5**. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate eppendorf tube, containing 5 mg powdered indium. Effervescence and the formation of a grey slurry were observed.

The reaction was stirred at room temperature for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

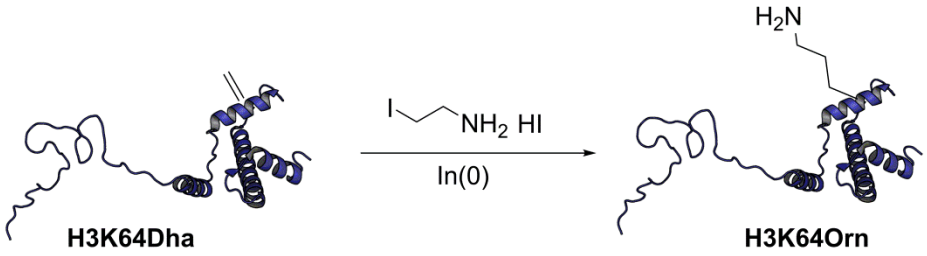
ESI-MS for the modified protease is shown below:

Calculated mass: 26922 (SBL 156Kme$_2$ 221PMS)

Observed mass: 26926



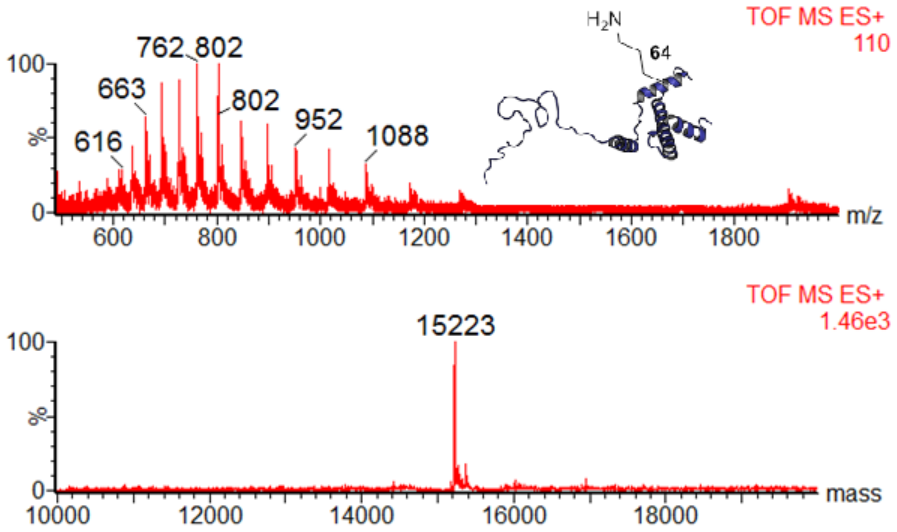
Installation of ornithine H3 K64Orn



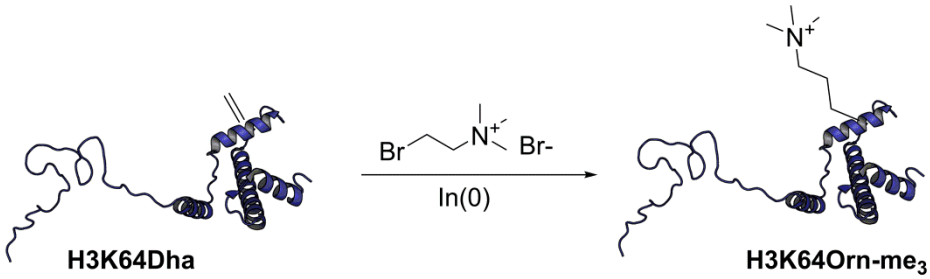
Indium powder (6mg, Sigma) was added to a 200 µL eppendorf tube. Reaction buffer (500 mM NH_4OAc, pH 6.0, 5M Gdn.HCl, 180 µL) was then added, followed by 20 µL of 10 mg/mL H3K64Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then ported into a N_2 atmosphere glovebox and equilibrated to anaerobic conditions (cap open) overnight at room temperature.

The following morning, within the glovebox, 20 mg (2-iodoethyl)amine hydroiodide **S2** was combined with the protein/Indium to give an orange slurry, which began to effervesce and clarify upon magnetic stirring, as the reaction progressed. After 1 hour stirring at room temperature, the reaction was removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass (H3K64Orn C110A): 15226
Observed mass: 15223



Installation of trimethylornithine H3 K64Orn-me$_3$



Indium powder (6mg, Sigma) was added to a 200 µL eppendorf tube. Reaction buffer (500 mM NH_4OAc, pH 6.0, 5M Gdn.HCl, 180 µL) was then added, followed by 20 µL of 10 mg/mL H3K64Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then ported into a N_2 atmosphere glovebox and equilibrated to anaerobic conditions (cap open) overnight at room temperature.

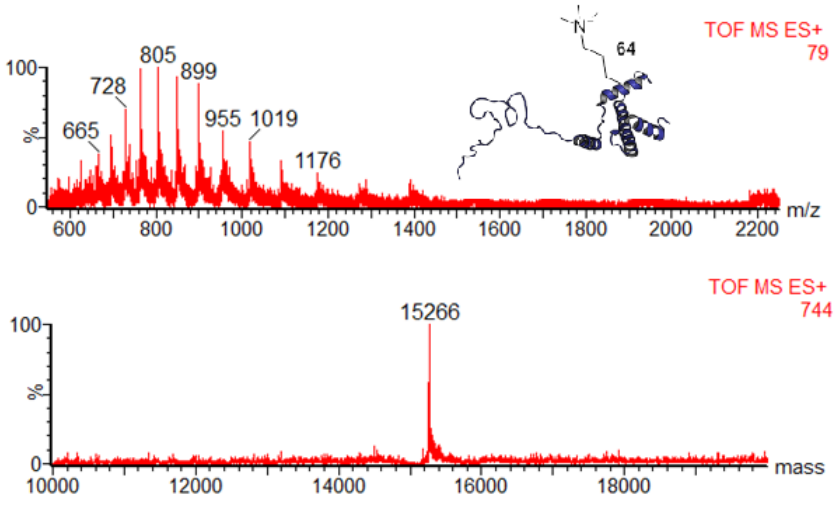
The following morning, within the glovebox, 16 mg (2-Bromoethyl)trimethylammonium bromide **S2** was combined with the protein/Indium to give a pale solution, which began to effervesce and clarify upon magnetic stirring, as the reaction progressed. After two hours stirring at room temperature, the reaction was removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-

equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

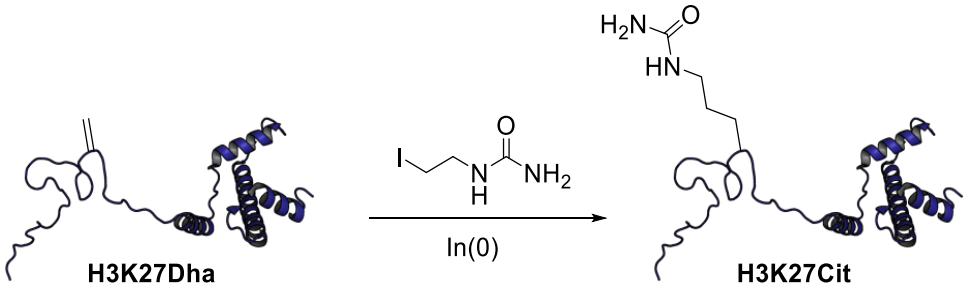
ESI-MS for the modified histone is shown below:

Calculated mass (H3K64Orn C110A): 15270

Observed mass: 15266



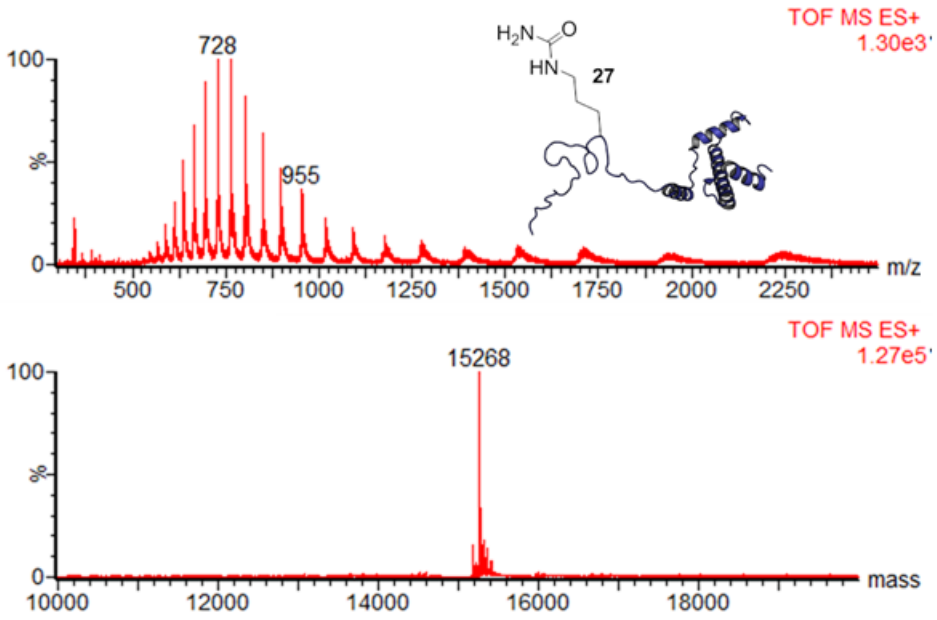
Installation of citrulline H3 27cit



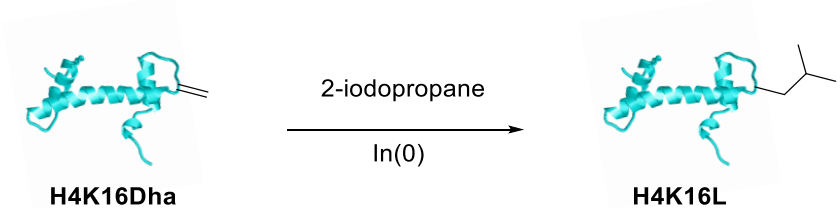
Indium powder (4mg, Sigma) was added to a 200 µL eppendorf tube. Reaction buffer (500 mM NH_4OAc, pH 6.0, 5M Gdn.HCl, 180 µL) was then added, followed by 20 µL of 10 mg/mL H3K27Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then ported into a N_2 atmosphere glovebox and equilibrated to anaerobic conditions (cap open) overnight at room temperature.

The following morning, within the glovebox, 28 mg (2-iodoethyl)urea was combined with the protein/Indium to give a red slurry, which began to effervesce and clarify upon magnetic stirring, as the reaction progressed. After 30 minutes at room temperature, the reaction was removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below
Calculated mass: 15267.8 (H3K27Cit C110A)
Observed mass: 15268



Introduction of leucine from dehydroalanine H4 16L



Indium powder (2 mg, Sigma) was added to a 200 µL eppendorf tube. Reaction buffer (500 mM NH_4OAc, pH 6.0, 5M Gdn.HCl, 180 µL) was then

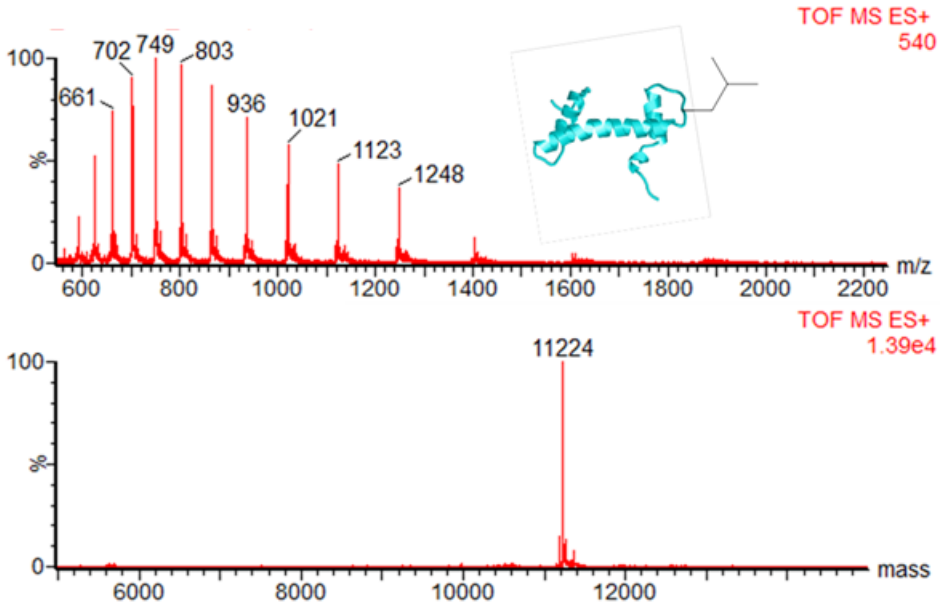
added, followed by 20 µL of 5 mg/mL H3K64Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 0.5 mg/mL. The resultant protein solution was then ported into a N_2 atmosphere glovebox and equilibrated to anaerobic conditions (cap open) overnight at room temperature.

The following morning, within the glovebox, 7 µL 2-iodopropane was combined with the protein/Indium to give a pale solution, which began to effervesce and clarify upon magnetic stirring, as the reaction progressed. After 30 minutes stirring at room temperature, the reaction was removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below.

Calculated mass: 11221

Observed mass: 11224



References: Chapter 3

1 Neet, K. E. & Koshland, D. The conversion of serine at the active site of subtilisin to cysteine: a" chemical mutation". *Proceedings of the National Academy of Sciences* **56**, 1606-1611 (1966).

2 Gross, E. in *Protein Crosslinking* 131-153 (Springer, 1977).

3 Gulzar, M. S., Morris, K. B. & Gani, D. Control of the regioselectivity of N-nucleophile addition to N-carbonyl protected dehydroalanines: a model for the ammonia-lyase enzymes. *Journal of the Chemical Society, Chemical Communications*, 1061-1062 (1995).

4 Gerstberger, S. Part II Thesis thesis, University of Oxford.

5 Chalker, J. *Reaction Engineering for Protein Modification: Tools for Chemistry and Biology.* DPhil Thesis thesis, University of Oxford, (2011).

6 Crich, D. & Davies, J. W. Free-radical addition to di-and tripeptides containing dehydroalanine residues. *Tetrahedron* **45**, 5641-5654 (1989).

7 Beckwith, A. L. & Chai, C. L. Diastereoselective radical addition to derivatives of dehydroalanine and of dehydrolactic acid. *Journal of the Chemical Society, Chemical Communications*, 1087-1088 (1990).

8 Kessler, H., Wittmann, V., Köck, M. & Kottenhahn, M. Synthesis of C‐Glycopeptides via Free Radical Addition of Glycosyl Bromides to Dehydroalanine Derivatives. *Angewandte Chemie International Edition in English* **31**, 902-904 (1992).

9 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).

10 Crich, D., Davies, J., Negron, G. & Quintero, L. Preparation of α-amino acids by radical addition to a dehydroalanine derivative. *Journal of chemical research. Synopses*, 140-141 (1988).

11 Giese, B. Formation of CC bonds by addition of free radicals to alkenes. *Angewandte Chemie International Edition in English* **22**, 753-764 (1983).

12 Petrier, C., Dupuy, C. & Luche, J. Conjugate additions to α, β-unsaturated carbonyl compounds in aqueous media. *Tetrahedron letters* **27**, 3149-3152 (1986).

13 Delair, P. & Luche, J.-L. A new sonochemical carbonyl cross-coupling reaction. *Journal of the Chemical Society, Chemical Communications*, 398-399 (1989).

14 Luche, J., Allavena, C., Petrier, C. & Dupuy, C. Ultrasound in organic synthesis 17 mechanistic aspects of the conjugate additions to α-enones in aqueous media. *Tetrahedron letters* **29**, 5373-5374 (1988).

15 Bernardes, G. J., 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).

16 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Chromatin Protocols*, 1-16 (1999).

17 Orphanides, G. & Reinberg, D. A unified theory of gene expression. *Cell* **108**, 439-451 (2002).

18 Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post‐Translational Modifications. *Angewandte Chemie International Edition* **51**, 1835-1839 (2012).

19 Hawkins, C. L. & Davies, M. J. Generation and propagation of radical reactions on proteins. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1504**, 196-219 (2001).

20 Friedman, M. Chemistry, biochemistry, nutrition, and microbiology of lysinoalanine, lanthionine, and histidinoalanine in food and other proteins. *Journal of Agricultural and Food Chemistry* **47**, 1295-1319 (1999).

21 Fields, R. & Dixon, H. B. Micro method for determination of reactive carbonyl groups in proteins and peptides, using 2, 4-dinitrophenylhydrazine. *Biochemical Journal* **121**, 587-589 (1971).

22 Matthews, M. L. *et al.* Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nature Chemistry* (2016).

23 Thom, J., Anderson, D., McGregor, J. & Cotton, G. Recombinant Protein Hydrazides: Application to Site-Specific Protein PEGylation. *Bioconjugate Chemistry* **22**, 1017-1020, doi:10.1021/bc2001374 (2011).

24 Lempens, E. H., Helms, B. A., Merkx, M. & Meijer, E. Efficient and chemoselective surface immobilization of proteins by using aniline‐catalyzed oxime chemistry. *ChemBioChem* **10**, 658-662 (2009).

25 Dirksen, A., Dirksen, S., Hackeng, T. M. & Dawson, P. E. Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. *Journal of the American Chemical Society* **128**, 15602-15603 (2006).

26 Rashidian, M. *et al.* A Highly Efficient Catalyst for Oxime Ligation and Hydrazone–Oxime Exchange Suitable for Bioconjugation. *Bioconjugate Chemistry* **24**, 333-342, doi:10.1021/bc3004167 (2013).

27 Gross, E. & Morell, J. L. The number and nature of α, β-unsaturated amino acids in nisin. *FEBS letters* **2**, 61-64 (1968).

28 Chan, W. C., Bycroft, B. W., Lian, L.-Y. & Roberts, G. C. Isolation and characterisation of two degradation products derived from the peptide antibiotic nisin. *Febs Letters* **252**, 29-36 (1989).

29 Spande, T., Witkop, B., Degani, Y. & Patchornik, A. Selective cleavage and modification of peptides and proteins. *Advances in protein chemistry* **24**, 97-260 (1970).

30 Knight, Z. A. *et al.* Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotech* **21**, 1047-1054 (2003).

31 Cohen, S. L., Price, C. & Vlasak, J. β-Elimination and peptide bond hydrolysis: two distinct mechanisms of human IgG1 hinge fragmentation upon storage. *Journal of the American Chemical Society* **129**, 6976-6977 (2007).

32 Xu, L., Farthing, A. K., Shi, Y.-J., Meinke, P. T. & Liu, K. Conversion of Nocathiacin I to Nocathiacin Acid by a Mild and Selective Cleavage of Dehydroalanine. *The Journal of Organic Chemistry* **72**, 7447-7450, doi:10.1021/jo071115p (2007).

33 Whitaker, J. R., Feeney, R. E. & Sternberg, M. M. Chemical and physical modification of proteins by the hydroxide ion. *Critical Reviews in Food Science & Nutrition* **19**, 173-212 (1983).

34 Rauk, A. *et al.* Effects of structure on αC– H bond enthalpies of amino acid residues: Relevance to H transfers in enzyme mechanisms and in protein oxidation. *Biochemistry* **38**, 9089-9096 (1999).

35 Rauk, A. & Armstrong, D. A. Influence of β-sheet structure on the susceptibility of proteins to backbone oxidative damage: preference for αc-centered radical formation at glycine residues of antiparallel β-sheets. *Journal of the American Chemical Society* **122**, 4185-4192 (2000).

36 Diggle, J. H., Mcvittie, J. D. & Peacocke, A. R. The Self‐Association of Chicken‐Erythrocyte Histones. *European Journal of Biochemistry* **56**, 173-182 (1975).

37 Wright, P. J. & English, A. M. Scavenging with TEMPO• to identify peptide-and protein-based radicals by mass spectrometry: Advantages of spin scavenging over spin trapping. *Journal of the American Chemical Society* **125**, 8655-8665 (2003).

38 Lam, M. A., Pattison, D. I., Bottle, S. E., Keddie, D. J. & Davies, M. J. Nitric oxide and nitroxides can act as efficient scavengers of protein-derived free radicals. *Chemical research in toxicology* **21**, 2111-2119 (2008).

39 Milazzo, G., Sharma, V. & Caroli, S. Tables of standard electrode potentials Wiley. *New York, NY* (1978).

40 Bard, A. J., Parsons, R. & Jordan, J. *Standard potentials in aqueous solution*. Vol. 6 (CRC press, 1985).

41 Miyabe, H., Ueda, M., Nishimura, A. & Naito, T. Indium-mediated intermolecular alkyl radical addition to electron-deficient CN bond and CC bond in water. *Organic letters* **4**, 131-134 (2002).

42 Shen, Z.-L., Yeo, Y.-L. & Loh, T.-P. Indium– copper and indium–silver mediated Barbier–Grignard-type alkylation reaction of aldehydes using unactivated alkyl halides in water. *The Journal of organic chemistry* **73**, 3922-3924 (2008).

43 György, B., Tóth, E., Tarcsa, E., Falus, A. & Buzás, E. I. Citrullination: a posttranslational modification in health and disease. *The international journal of biochemistry & cell biology* **38**, 1662-1677 (2006).

44 Anzilotti, C., Pratesi, F., Tommasi, C. & Migliorini, P. Peptidylarginine deiminase 4 and citrullination in health and disease. *Autoimmunity reviews* **9**, 158-160 (2010).

45 Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *Bioessays* **25**, 1106-1118 (2003).

46 Christophorou, M. A. *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104-108 (2014).

47 Suzuki, A. *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nature genetics* **34**, 395-402 (2003).

48 Chang, X.-t. *et al.* Localization of peptidylarginine deiminase 4 (PADI4) and citrullinated protein in synovial tissue of rheumatoid arthritis. *Rheumatology* **44**, 40-50 (2005).

49 Zhang, X. *et al.* Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor α target gene activation. *Proceedings of the National Academy of Sciences* **109**, 13331-13336, doi:10.1073/pnas.1203280109 (2012).

50 Gunnoo, S. B. *Site-specific chemical modification of antibodies for the modulation of function* DPhil Thesis thesis, University of Oxford, (2013).

51 Stabile, M. R. *et al.* Probing the specificity of the S1 binding site of M222 mutants of subtilisin B. lentus with boronic acid inhibitors. *Bioorganic & Medicinal Chemistry Letters* **6**, 2501-2506 (1996).

Chapter 4: The discovery of borohydride-mediated 'chemical mutation'

Introduction

As described in the previous chapter, protein chemical mutation employing zinc or indium powder as radical initiators resulted in significant levels of undesired di-alkylation side-products. Although substitution of indium for zinc proved partially successful, supporting the hypothesis of an improved rate of quenching of the problematic Cα radical (grey box, **Figure 4.1**) through tuning of redox potential, this did not prove a general solution. As a consequence, alternative methods for 'quenching' of this troublesome intermediate were sought. In theory, direct hydrogen atom ("radical hydride") transfer would enable an alternative 'quenching' process not dependent on the two-step mechanism (single-electron reduction followed by protonation) required for the M^0 – mediated system. Hydrogen atom transfer could also enable the building of a suitable radical chain reaction[1-3], which would improve the overall efficiency of the reaction. Analysis of the proposed chain cycle (highlighted in blue, **Figure 4.1**), however, highlights the challenges inherent in this approach. A successful chain reaction would require selective increase of the rate of the key hydrogen atom transfer ($\mathbf{k_1}$) over the rates of not only the di-alkylation resultant from radical-radical coupling ($\mathbf{k_2}$), but also over the rate of direct reduction of the initial alkyl radical ($\mathbf{k_3}$).

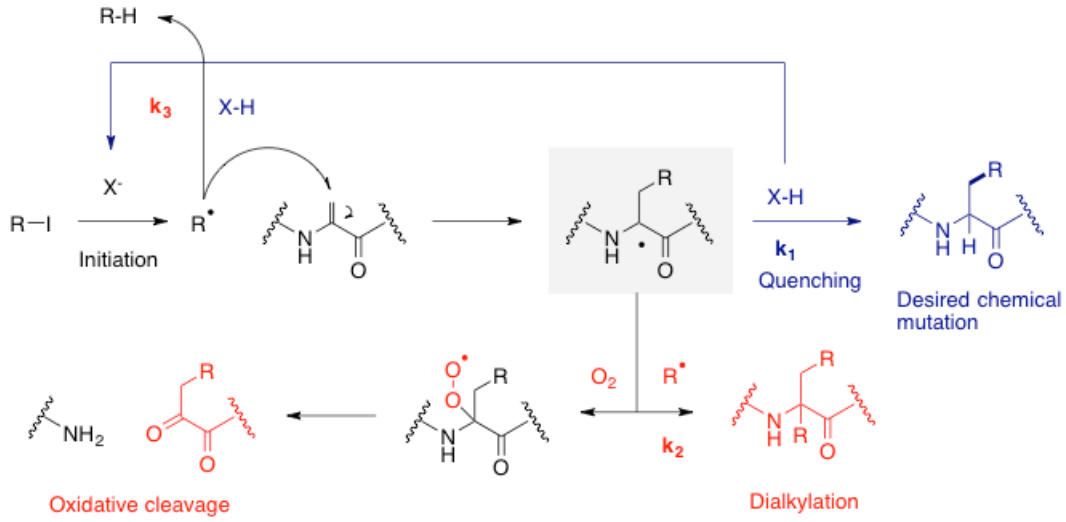


Figure 4.1: Proposed chain-cycle for quenching of the problematic Cα radical via hydrogen atom transfer

Identification of sodium borohydride as a useful radical 'mediator' for chemical mutation

Initial experiments focused on the addition of known hydrogen atom donors to the M^0-mediated reaction system. Silanes and thiols have been used extensively as hydrogen transfer reagents in small-molecule radical chemistry[4]. Critically, these reagents have also been reported to be compatible with aqueous conditions. Initial experiments employing zinc metal and tris(trimethylsilyl) silane[5-7] as a water-soluble H-atom transfer reagent were unsuccessful, with no robust effect on the quantity of di-alkylation observed (see **Experimental Section, Chapter 3, p. 101**). Supplementing the silane with a small molecule thiol to effect polarity transfer in hydrogen atom transfer was also unsuccessful. Finally, the use of zinc (or indium) metal in tandem with *N*-Ethylpiperidine hypophosphite (EPHP)[8-10], an alternative water-soluble H-atom transfer reagent, was also notably unsuccessful, with no effect observed on the di-alkylation ratio (see **Experimental Section, Chapter 3, p. 102**).

An alternative strategy was thus envisaged, in which the active metal species (M^0) is generated in-situ by reduction of a suitable precursor[11-14]. This strategy should allow improved control of radical initiation and quenching via step-wise control over the amount of the active metal species present. Based on the initial success of the indium-mediated system, $InCl_3$ was chosen as a precursor reagent, which we envisaged could be converted to the active In^0 species *in-situ* by reduction, for example, with sodium borohydride ($NaBH_4$). Indeed, this system has been reported by Baba et al.[15] to be useful for the dehalogenation of alkyl halides. These authors proposed a mechanism involving *in-situ* generation of $InCl_2H$ as a radical hydride source. In the context of chemical mutation, it was envisaged that *in-situ* generation of $InCl_2H$ could act to quench the problematic Cα radical.

Initial exploration of the $InCl_3$/$NaBH_4$ system yielded promising results. After initial optimization, the reaction of H3-K64Dha (66 μM (1 mg/mL), in pH 6.0, 500 mM NH_4OAc, 5M Gdn.HCl buffer) with 2000 equivalents (3-iodopropyl)amine hydroiodide **1** in the presence of 2000 equivalents $InCl_3$ and 1000 equivalents $NaBH_4$ afforded the desired Dha→K chemical mutation as the essentially sole product (**Figure 4.2**). This result represented the first installation of a lysine-type side-chain to a histone protein without concurrent di-alkylation.

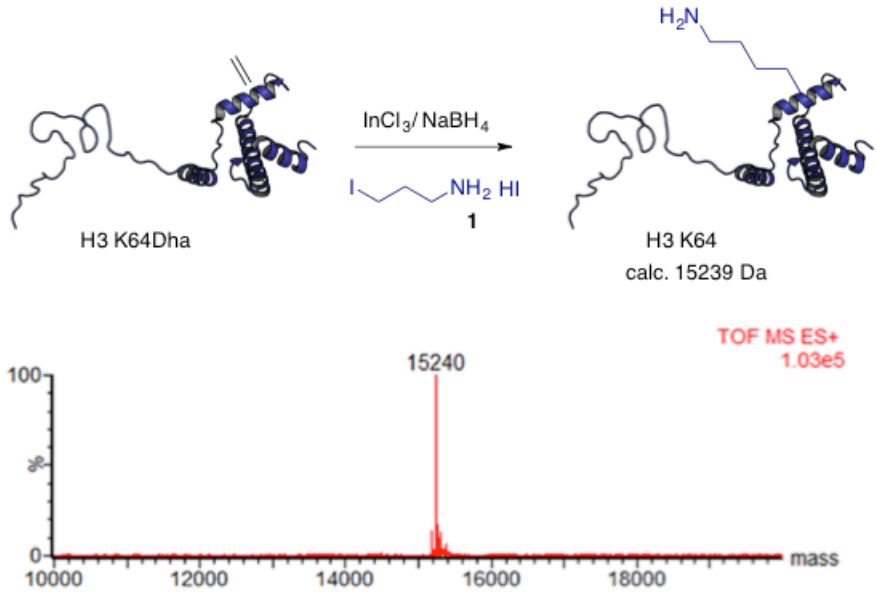


Figure 4.2: Successful chemical mutation of H3-K64Dha → H3-K64 using the $InCl_3/NaBH_4$ system. *Conditions:* 1 mg/mL (66 µM, pH 6.0 500 mM NH_4OAc, 5M Gdn.HCl) H3-K64Dha, 1000 eq. NaBH4, 2000 eq. $InCl_3$, 2000 eq. (3-iodopropyl)amine iodide, 30 minutes, 4 °C)

Monomethyl-arginine, a relatively unexplored naturally occurring PTM[16], could also be installed into the unnatural H3-K64Dha site with quantitative conversion to the mono-addition product (**Figure 4.3a,** further details of our exploration into arginine methylation (including syntheses of the requisite iodo-reagents) are reported in **Chapter 5**). Previous attempts using this primary iodide and M^0 –mediated conditions had resulted in intractable mixtures of remaining starting material in addition to mono- and dialkylation products (**Figure 4.3b**).

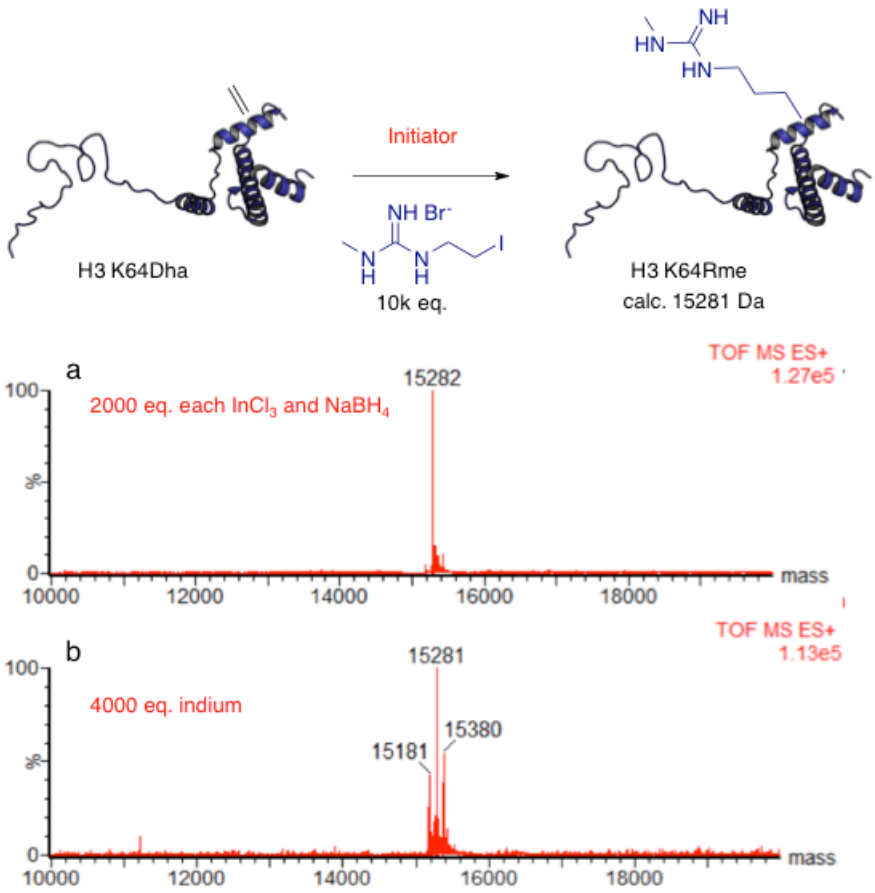
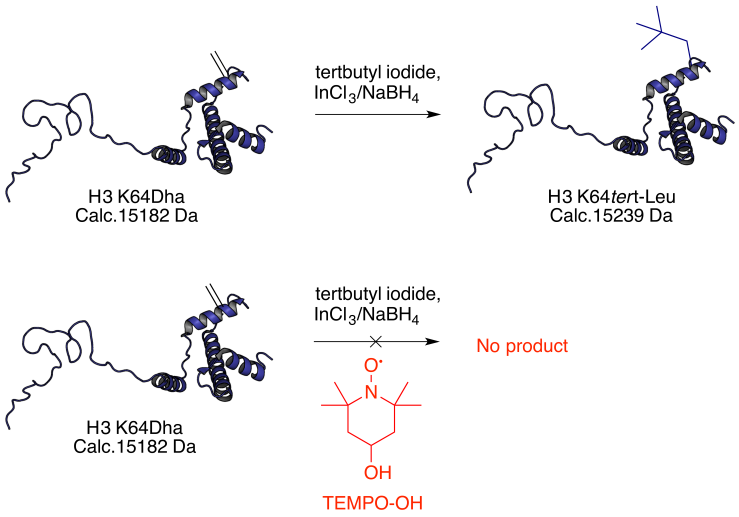


Figure 4.3: Comparison of product profile for chemical mutation → Rme using either a. $InCl_3/NaBH_4$ or b. Indium powder

These promising results obtained with the $InCl_3/NaBH_4$ reaction inspired further investigation into the reaction selectivity and mechanism. Reactions run in the presence of TEMPO-OH resulted in reaction inhibition, consistent with a radical mechanism (**Figure 4.4a**). Further, continuous wave EPR spectroscopy of a flash-frozen reaction mixture containing $InCl_3$, $NaBH_4$ and isopropyl iodide resulted in a spectrum highly consistent with a simulation of the expected alkyl radical spin system (**Figure 4.4b**).

a



b

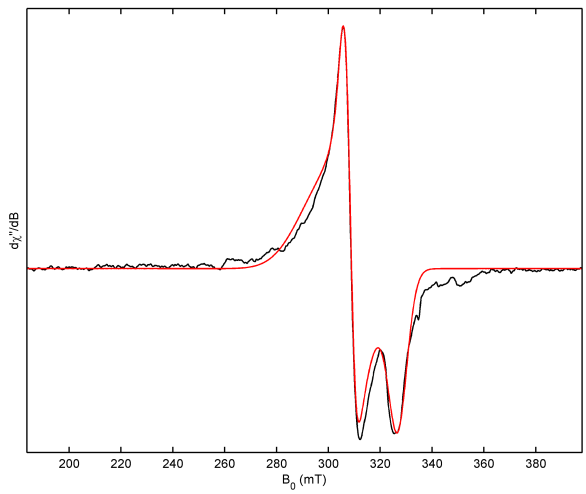


Figure 4.4: Evidence for a radical mechanism for $InCl_3/NaBH_4$ –mediated chemical mutation. a. The reaction is inhibited by sub-stoichiometric TEMPO-OH. b. Continuous wave electron paramagnetic resonance spectrum of $InCl_3/NaBH_4$ mixed with 2-iodopropane in 500 mM NH_4OAc buffer to generate an organoradical. The spectrum was recorded on a Bruker EMX at a temperature of 100 K, 5 mW microwave power and 10 G modulation amplitude under non saturated conditions. The red line indicates a spin simulation performed using the EasySpin software package for MatLab.

It is conceivable that NaBH4, or other hydride-containing species in the reaction manifold, could act to reduce the protein-Dha directly to alanine[17]. To test this possibility, and to confirm the site of the radical addition, an incomplete reaction solution was treated with 2-mercaptoethanol in basic buffer[18] (see **Experimental Section** for details). Any dehydroalanine-

containing species should rapidly form a thioether covalent adduct with 2-mercaptoethanol under these conditions. Indeed, complete adduct formation with the remaining Dha was observed, indicating that no Dha was reduced during the reaction. Further, the initial product peak was not observed to shift, strongly suggesting that the site of radical modification is indeed at the expected Dha site.

Control experiments were then undertaken to assess the necessity of each component of this complex reaction mixture. No reaction was observed in the absence of $NaBH_4$. In contrast, and to our surprise, the desired reaction was still observed in the absence of $InCl_3$. Although borohydrides have traditionally been viewed as sources of nucleophilic hydride, rare studies have previously suggested possible radical behavior, including hydrogen atom transfer in radical chain processes[11,19-23]. A recent report from Kawamoto et al.[24] utilized sodium cyanoborohydride in concert with AIBN to generate alkyl radicals from alkyl iodides for conjugate addition reactions with acrylate acceptors (**Figure 4.5**). These authors proposed a polar effect wherein hydrogen transfer to the Cα-ester radical from cyanoborohydride is preferred over direct transfer to alkyl-type radicals. Strikingly, in our hands $NaBH_4$ proved to be a more efficient reagent (judged by fewer equivalents of iodide required) relative to all of the previous systems investigated. Relative to the previous systems, di-alkylation products were greatly reduced and could be readily eliminated for even primary halides by simple titration of iodide loading, a feat that did not prove possible for the M^0 systems.

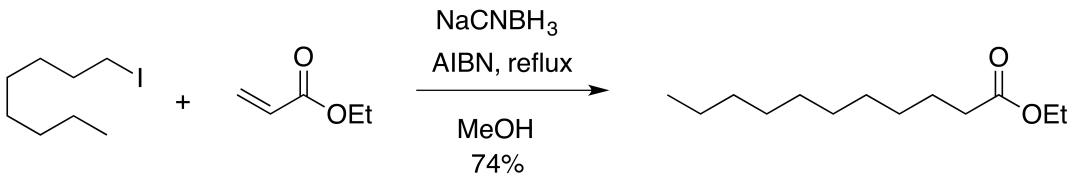


Figure 4.5: Conjugate addition reaction of alkyl radicals mediated by sodium cyanoborohydride. Adapted from [24]

Initial investigation into reaction selectivity and mechanism

Further investigations were undertaken to assuage any doubts about competing hydride-mediated processes. The reaction of H3-K64Dha with *tert*-butyl iodide was chosen as a model reaction, as tertiary iodides are unlikely to react *via* SN_2-type mechanisms and are instead favoured substrates for radical generation. Quantitative conversion to the desired unnatural '*tert*-butylalanine' at position 64 was observed by treating H3-K4Dha (66 μM, ~1 mg/mL, in pH 6.0, 500 mM NH_4OAc, 5M Gdn.HCl buffer) with 2000 equivalents *tert*-butyliodide and 2000 equivalents $NaBH_4$, under anaerobic conditions (**Figure 4.6a**). Reaction at the intended dehydroalanine site was confirmed by tryptic digest followed by 'peptide mapping' via LC/MS-MS (**Figure 4.6b**). Notably, this reaction was completely inhibited by the addition of sub-stoichiometric (relative to iodide and borohydride, 100 eq. relative to protein-Dha) 4-hydroxy-TEMPO, strongly implying a radical mechanism (**Figure 4.6c**). Finally, EPR spectroscopy conducted on the reaction of $NaBH_4$ and isopropyl iodide (the substrate used previously for EPR detection of radicals in the $InCl_3/NaBH_4$ system) presented some evidence for radical generation, albeit at lower concentrations relative to the metal-mediated system. This observed lower efficiency of radical generation could perhaps be partly responsible for the reduced quantities of di-alkylation and hence 'cleaner' product profiles noted for this system.

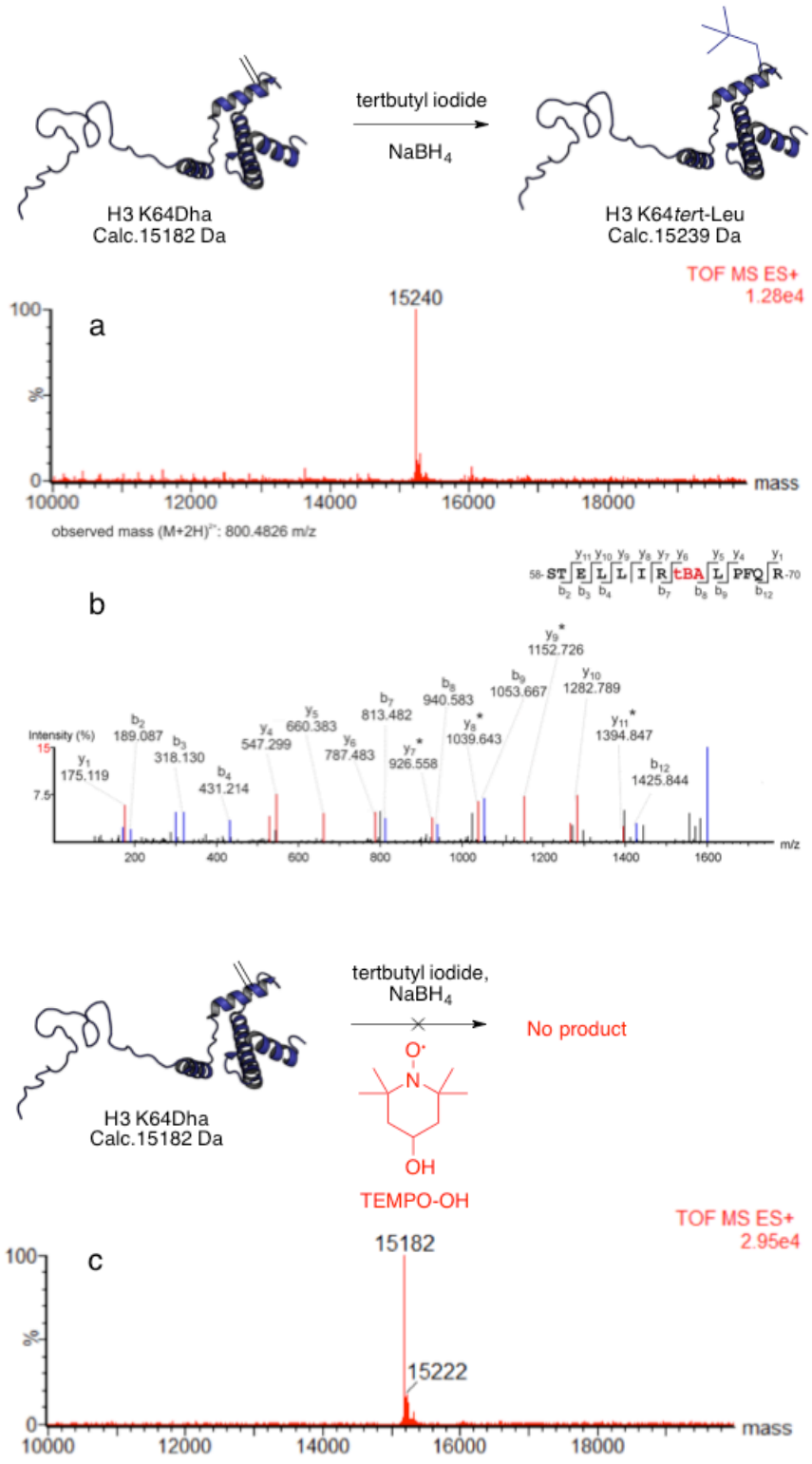


Figure 4.6: Initial investigations into borohydride-mediated chemical mutation. a. Reaction of $NaBH_4$ and *tert*-butyl iodide with H3-K64Dha affords the *tert*-butyl alanine containing protein product. b. Site of the $NaBH_4$-mediated reaction is confirmed by tryptic digest followed by LC-MS/MS analysis. c. The reaction is inhibited by sub-stoichiometric TEMPO-OH

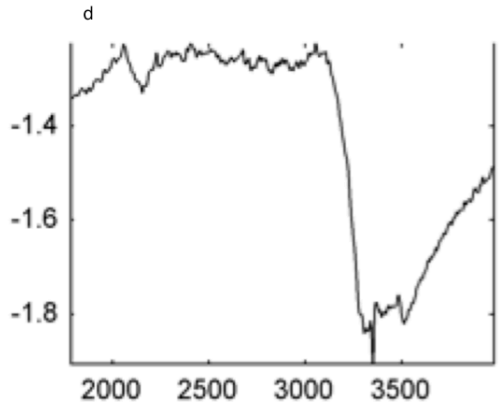


Figure 4.6d: EPR spectroscopy suggests a low concentration of alkyl radicals in the $NaBH_4$-mediated system

Next, the possibility of competing hydride-mediated reduction of dehydroalanine to alanine was investigated. Indeed, incubation of protein-bound dehydroalanine with $NaBH_4$ has been used previously to effect this transformation[17]. During initial attempts at optimizing reactions performed with $NaBH_4$ and alkyl iodide, persistence of starting material was observed under some conditions, especially with primary, bulky and/or insoluble iodides. It was suspected that this reaction-incompetent, persistent 'starting material' was in fact alanine resultant from hydride-mediated reduction. To test this idea, aliquots of a reaction performed in the absence of iodide (H3K27-Dha + $NaBH_4$ only) were tested for remaining dehydroalanine by addition of dithiothreitol (DTT) as a representative thiol nucleophile (see **Experimental Section**). In the presence of dehydroalanine, quantitative conversion to the thia-Michael product is observed, while incomplete conversion indicates the presence of alanine. Indeed, this experiment revealed complete reduction to alanine occurred on a time frame of ~1hr under the conditions employed, which were representative of a typical 'chemical mutation' (66 μM H3-Dha, 1000 eq. $NaBH_4$). This result indicates a 'kinetic window' for successful reactions, a fact that may limit the applicability of chemical mutation for the installation of bulky or insoluble halides that have slower rates of radical generation and/or addition to dehydroalanine.

The chemoselectivity of the borohydride-mediated reaction was addressed next. For these reactions, the primary iodide (3-iodopropyl)dimethylamine hydroiodide **2** was utilized, as primary halides are the most reactive in the SN_2-type reactions that could be expected to occur at nucleophilic protein residues such as lysine, an abundant residue in histone H3. Reaction of H3-K27Dha with **2** yielded H3K27me_2, a known PTM regulated by the polycomb complex PRC2[25]**.** This reaction was readily optimized using borohydride-mediated conditions to give quantitative conversion to the desired mono-alkylated product (**Figure 4.7a**). The modified protein was subjected to trypsin digestion and peptide mapping by LC-MS/MS. This analysis readily identified Kme_2 at the desired K27 site (**Figure 4.7b**), with the modification observed in all of the different peptide products containing this site. Using the PEAKS™ software[26], a search was then undertaken for modification at nucleophilic residues (+86.16 Da mass shift, **Figure 4.7c**). Pleasingly, no such modification was observed. Although the overall protein sequence coverage was admittedly poor, as is often the case for the highly Lys/Arg rich histone H3, multiple lysines were present in the identified peptides, and none bore the expected product of nucleophilic attack on the iodide reagent. Further, a similar analysis with the protein cAbLys, in which sequence coverage was almost complete, also detected no modification at any residue other than the introduced dehydroalanine (this result is discussed further in the context of the protein scope, below. Full sequence maps are displayed in the **Experimental Section**). Finally, as a stringent test of the reaction selectivity for dehydroalanine, H3-K9C, containing no dehydroalanine, was reacted with **2** and sodium borohydride in 'high pH' buffer (as a stringent test for SN2-type reactions, 100 mM sodium phosphate pH 8.0, 5M Gdn.HCl). No modification was observed (**Figure 4.7d**), again consistent with the desired radical alkylation mechanism*.

**It must be noted that over-alkylation via SN_2 type reactions is not chemically impossible under these reaction condition, and has indeed been observed under some circumstances. However, the short-time frame, low-temperature,*

low pH relative to the pKa of protein nucleophiles and (relatively) low concentrations of reagents employed in our optimized protocols greatly mitigates against these competing processes, explaining their absence in the control reactions above.

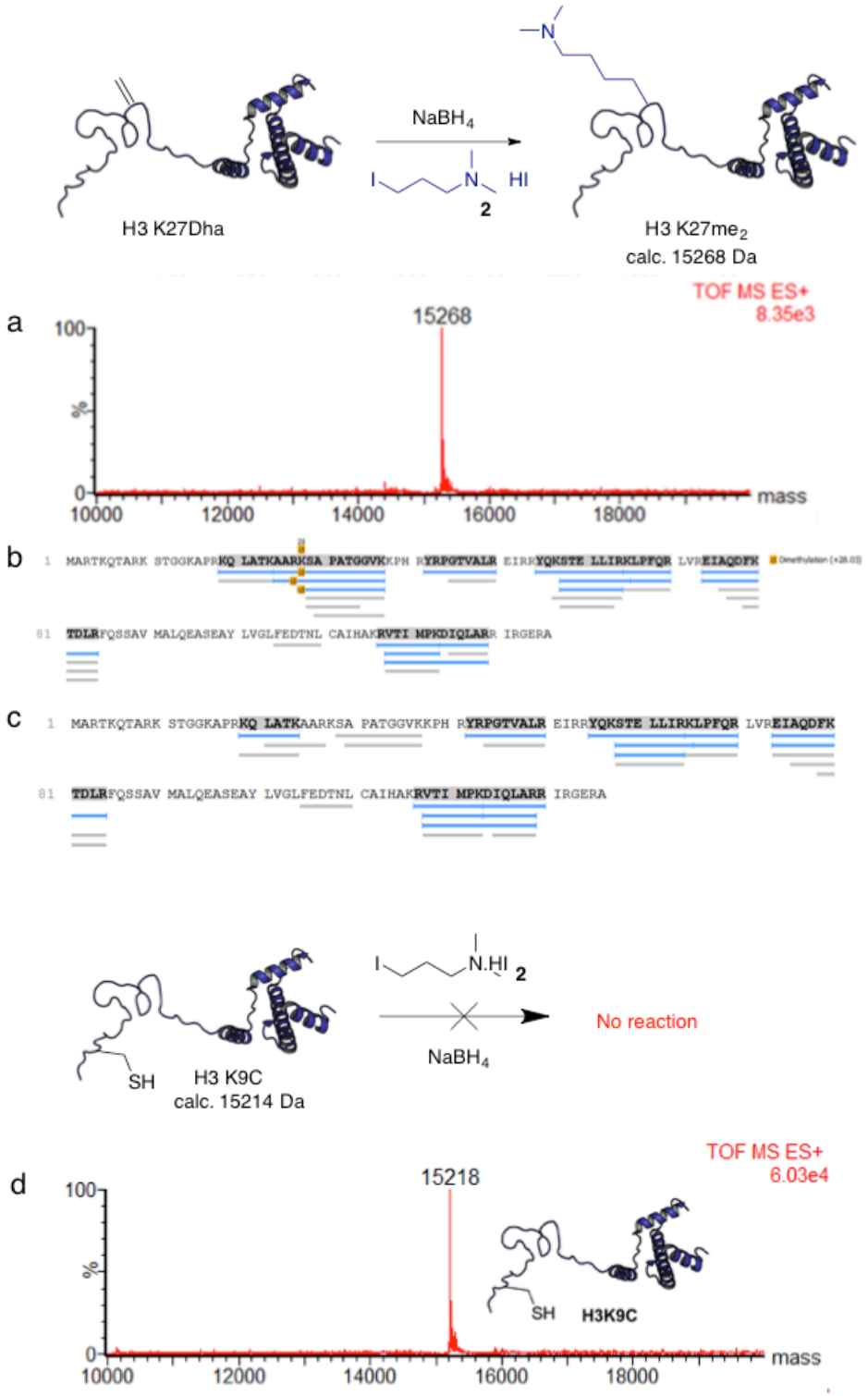


Figure 4.7: Chemoselectivity of borohydride-mediated chemical mutation. a. Reaction of H3-K27Dha with $NaBH_4$ and (3- iodopropyl)dimethylamine hydroiodide **2** yields the desired mono-alkylated product. b. Tryptic digest followed by LC-MS/MS confirms the site of modification (b.) and the lack of competing alkylation at nucleophilic sites (c.). d. Reaction does not take place in the absence of Dha.

Side-chain scope accessible *via* the borohydride-mediated 'chemical mutation' protocol

With promising conditions for chemical mutation in hand, we turned to exploring the side-chain scope and protein diversity that could be accessed with the method (see **Appendix Figure 1** for full scope of chemical mutations conducted in the Davis laboratory to date). Of pre-eminent interest was introduction of naturally occurring PTMs, particularly in the histone context, but attention was also directed towards potentially useful unnatural residues.

Conditions were readily obtained for installation of both lysine and its methyl and dimethyl congeners by simple variation of the iodide loading. It was discovered that such 'titration' of iodide loading with a fixed quantity of sodium borohydride was the most facile procedure for optimizing these reactions. It is recommended that this titration be performed for each new batch of protein-Dha or iodide, as slight variation in protein concentration or purity of reagents can have a large effect at the low protein concentrations typically employed. Reactions typically were complete within ~15-30 minutes, with LC-MS employed to confirm the success of reaction. Installation of trimethyllysine and a ^{13}C-methyl labeled variant was also possible – this is the subject of the next chapter (**Chapter 5**). The iodide precursors for these chemical mutations to lysine derivatives were most conveniently prepared by simple treatment of the corresponding alcohol, typically commercially available, with aqueous hydroiodic acid (HI, 55%) at >100 °C in an open flask[27] (see **Experimental Section**). Concentration of the reaction mixture by evaporation generally enables facile crystallization of the desired product, as the hydroiodide salt, upon removal of heat.

H3-K27Dha was chosen as the protein scaffold for the majority of reactions dedicated towards exploring the side-chain scope of borohydride-mediated chemical mutation. This mutant was empirically selected as it provided the most suitable peptides for identification/characterization following tryptic digest and LC-MS/MS, which was the main technique employed for identification of the installed modifications. Successful installation of lysine, Kme and Kme_2 side- chains into H3-K27Dha from the requisite iodides, mediated by $NaBH_4$, was readily confirmed by such analysis (**Figure 4.8**).

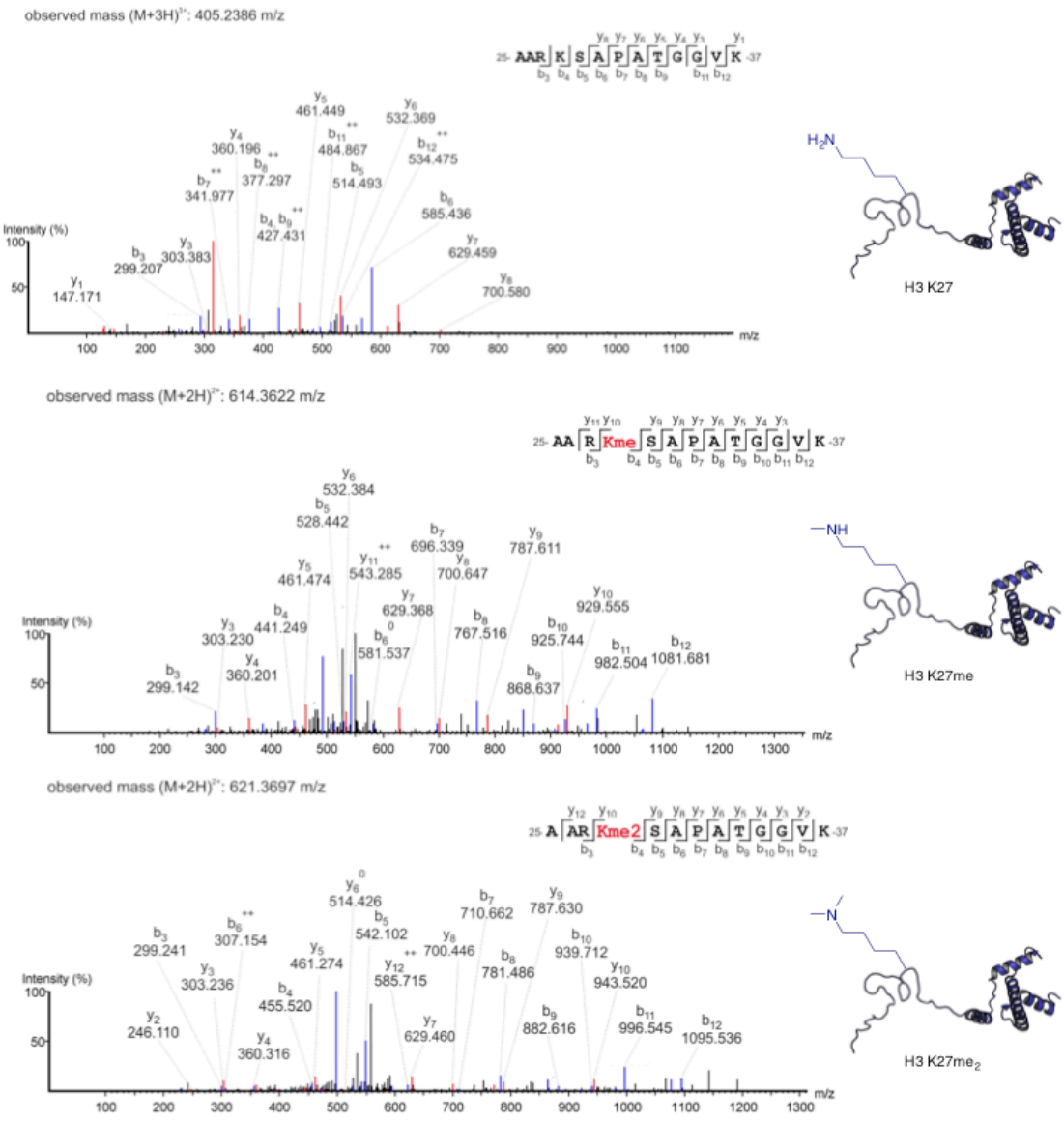


Figure 4.8: Successful installation of lysine, Kme and Kme_2 side- chains into H3-K27Dha from the requisite iodides, mediated by $NaBH_4$, is confirmed by tryptic digest followed by LC-MS/MS analysis

In terms of naturally occurring PTMs, methylglutamine (Qme) was also readily installed at this (unnatural) site (**Figure 4.9**) and confirmed by LC-MS/MS. Methylated glutamine is a recently discovered[28] PTM that has been identified on histone H2A as a regulator of FACT (facilitator of chromatin transcription) complex binding at a ribosomal DNA site. As our method is currently the only reported chemical or biological method for

incorporation of this residue into whole proteins, it may play a critical role in elucidation of further function for this PTM. Additionally, installation of modified arginine residues is also readily accomplished with our method, as has been briefly alluded to above; further details of this application are reported in **Chapter 5.**

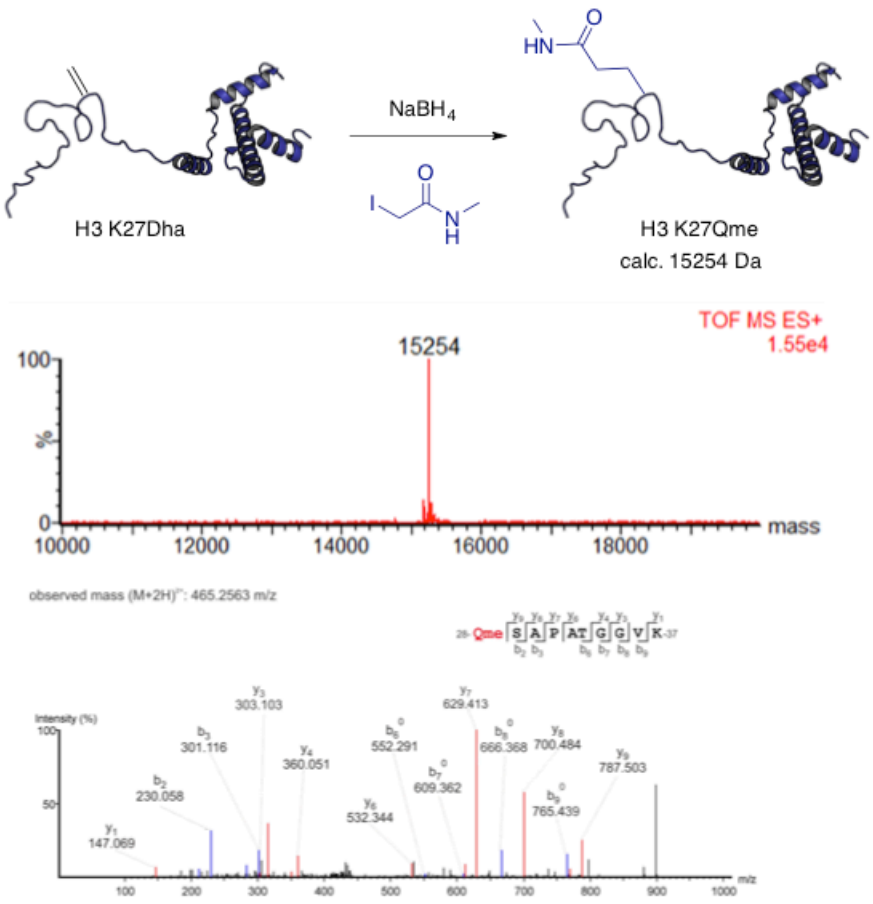


Figure 4.9: Installation of methylated glutamine (Qme) from H3-K27Dha. Top: LC/MS of reaction mixture. Bottom: LC-MS/MS of identified Qme-containing peptide *Conditions:* 1 mg/mL (66 μM, pH 6.0 500 mM NH_4OAc, 5M Gdn.HCl) H3-K27Dha, 500 eq. NaBH4, 500 eq. 2-iodo-*N*-methylacetamide, 30 minutes, 4 °C)

Unnatural alkyl residues could also be readily introduced into histone scaffolds by the developed protocol (**Figure 4.10**). Notably, tertiary, secondary and primary alkyl iodides are all readily utilised, including those with biologically relevant hydroxyl and amine functionality. The ready availability of short-chain alkyl iodides enabled systematic exploration of alkyl chain length and branching pattern. Thus, not only could leucine be

incorporated by reaction with isopropyl iodide but a series of systematic variants could also be installed: 'demethyl-Leu' (smaller), 'tert-Leu' (bulkier, also described above) and sec-Leu (extended). Notably, a CF_3-containing analogue of 'demethyl-Leu', 'CF3-norvaline' was readily installed by reacting H3-K27Dha with $NaBH_4$ and commercially available trifluoro-2-iodoethane (**Figure 4.10**). Trifluoromethyl groups are profoundly useful side-chains for biophysical characterization of proteins by NMR[29-34]; therefore this reaction may find application in this area. In general, the introduction of different alkyl side chains may be useful for modulating binding interfaces, particularly in the context of hydrophobic interactions where extension of classical mutagenesis to incorporate unnatural variants may enable enhanced binding affinities.

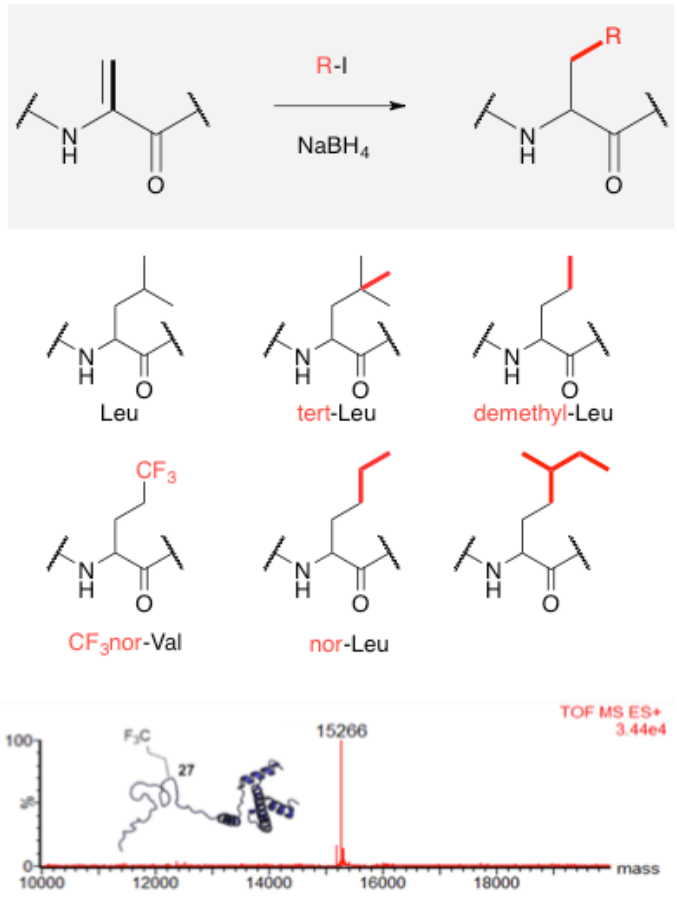


Figure 4.10: Alkyl side-chains installed to H3-K27Dha via borohydride-mediated chemical mutation. Top: Side-chains installed. Bottom: Installation of CF_3-norvaline. See **Experimental Section** for details.

Installation of glycosylated residues

In cases where a suitable iodide precursor is difficult to synthesise, 'mimics' of natural modifications may provide a more synthetically accessible alternative. Further, unnatural analogues of natural PTMs may be *more* useful than the natural form for biological experiments, for example in the case of hydrolytically unstable modifications[35]. The installation of phosphoserine analogues by chemical mutation is described in **Chapter 6.** Chemical glycosylation of expressed proteins can provide access to glycoproteins that are difficult to produce by conventional biological methods for non-canonical residue incorporation, such as stop-codon suppression[36]. Access to chain-extended analogues of both O- and N-linked glycosylated residues were readily accessible on histone proteins via our method (**Figure 4.11**). Further work on protein glycosylation by post-translational mutagenesis, including synthetic schemes for the precursors, can be found in *Wright et. al,*[37] and in the DPhil. thesis of Benjamin Bower.

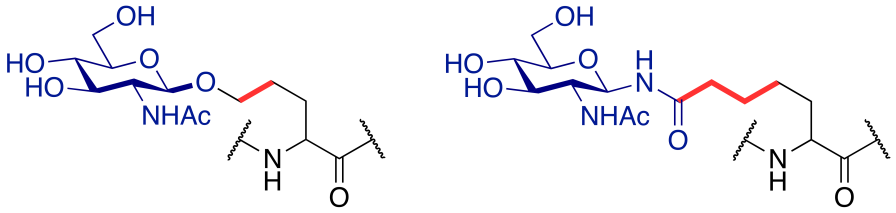


Figure 4.11: Glycosylated residues accessible *via* chemical mutation

Challenging residues

The developed method is, of course, not without limitations. Based on an empirical survey of the reaction scope, key factors enabling quantitative conversion with a given halide 'donor' are the aqueous solubility of the reagent and, critically, the 'philicity' of the resultant alkyl radical. Extending alkyl chains beyond propyl- to butyl- and beyond results in reduced conversions, presumably due to decreased solubility in the aqueous phase with increased hydrocarbon content. Radical polarity effects[4,38,39] are a crucial and often neglected element of reactivity in radical reactions. In the case of protein chemical mutation, 'nucleophilic'-type radical reagents

appear better matched to the electrophilic Dha centre. Thus, attempted chemical mutations with benzyl bromide and iodomethane, reagents for generation of (relatively) electrophilic benzyl[40,41] and methyl[38,42,43] radicals, resulted in poor conversions (**Figure 4.12**). Intriguingly, 'di-alkylation' products appeared to be equally favoured when conducting reactions with these radicals, perhaps due to an enhanced rate of radical-radical coupling. Attempted reactions with aryl iodides or bromides yielded only protein-Dha starting material.

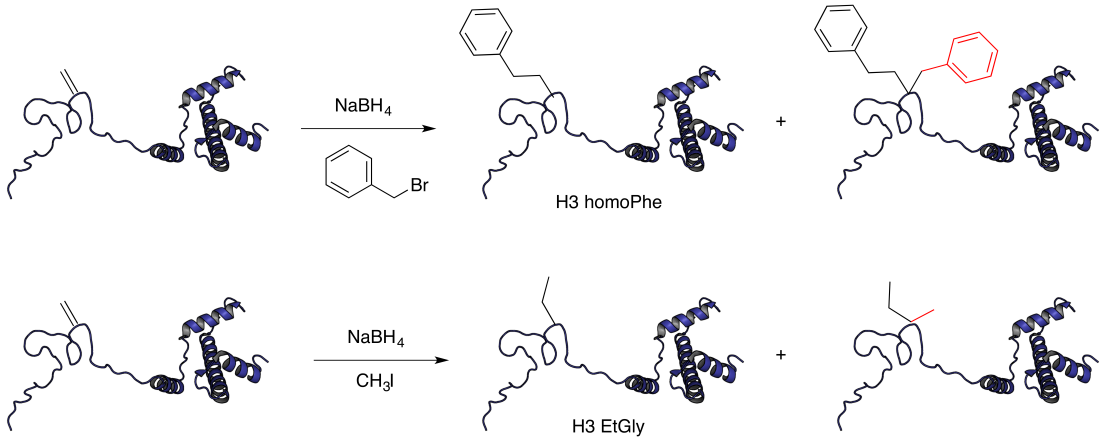


Figure 4.12: Challenging reactions. Attempted chemical mutation with 'electrophilic' radicals affords mixtures of starting material, mono- and di-alkylated products.

Assessment of protein scope and retention of protein structure

To date, nine different proteins have been successfully modified in our laboratory by the chemical mutation protocols described in this thesis (see **Appendix Figure 1**).

In addition to SBL and histone H3 and H4, two more challenging protein targets, cAbLys3[44,45] and p38α kinase[46], were initially selected to further explore the protein scope of the borohydride-mediated reaction.

cAbLys modification

The mammalian single-chain antibody cAbLys was chosen as a target as it had previously been shown to be recalcitrant towards metal-mediated modification, with protein degradation being the only result of previous attempts at chemical mutation. Notably, this protein possesses a distinct

folded structure (unlike the histone proteins) in solution, enabling assessment of the effect of the developed reaction conditions on protein tertiary structure.

The protein was expressed in mammalian cells by collaborators (see *Wright et al.*[37] for details) and provided as the A104C mutant for dehydroalanine formation. Cysteine to dehydroalanine conversion was readily performed following published procedures[45]. In brief, a 63 μM solution of cAbLys A104C in PBS (pH 8.0) was subjected to reduction with dithiothreitol (1000 eq.) to ensure the presence of cysteine in its reduced form. Following passage through a short desalting column, the reduced protein solution was treated with DBHDA (1000 eq.) at 37 °C for 5 hours, at which point LC-MS analysis indicated quantitative conversion to the dehydroalanine– containing protein. Subsequent desalting provided the desired protein for the key chemical mutation step.

With cAbLys-A104Dha in hand, attention turned towards reaction with representative iodides in chemical mutation reactions. Lysine derivatives were examined first. Indeed, lysine and its dimethylated derivative, Kme_2, were readily installed via $NaBH_4$- mediated reaction with the appropriate iodides (**Figure 4.13a**). cAbLys-A104Kme_2 was chosen for subsequent characterization. Tryptic digest followed by LC-MS/MS confirmed the site-selectivity of this modification – no residues other than the desired 104Dha were modified (**Figure 4.13b**).

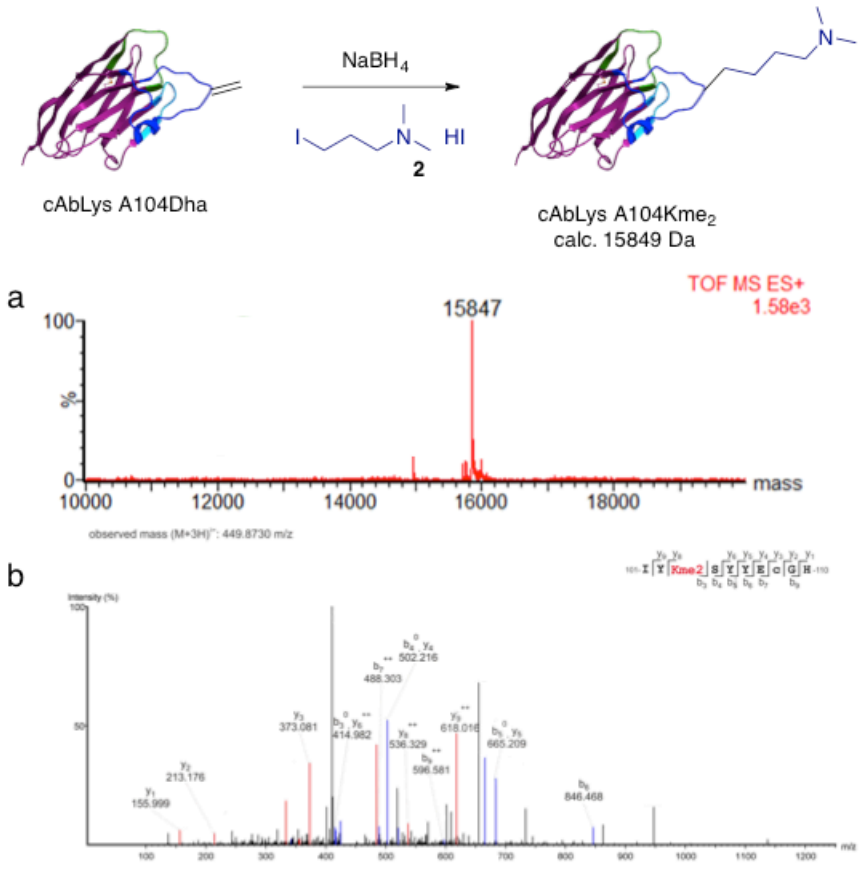


Figure 4.13: Installation of Kme_2 to cAbLys A104Dha. a. LC/MS of product mixture. b. LC-MS/MS of Kme_2 containing peptide. *Conditions:* 0.5 mg/mL cAbLys A104Dha (PBS, pH 8.0), 1000 eq. $NaBH_4$, 2000 eq. (3- iodopropyl)dimethylamine hydroiodide, 30 min, 4 °C.

Next, attention was turned to the question of retention of protein structure during the two-step chemical mutation procedure. cAbLys possesses a β-strand rich variable-domain immunoglobulin fold characteristic of camelid single-chain antibodies[44]. The site here modified is in the loosely structured antigen-binding loop and is thus not expected to lead to gross alterations (via allosteric mechanisms) in secondary or tertiary structure as a consequence of side-chain alteration alone. As a consequence, any disruption to protein folding should represent the effects of the reaction conditions themselves. Circular dichroism (CD) spectroscopy is a rapid, albeit low resolution, method for assessing structural changes in proteins[47]. Circularly polarized light is used to elicit n→p* and p→p* transitions of the amide bonds in the protein[48]. Critically, the intensity of these transitions is

influenced by the geometry of the peptide backbone, rendering far-UV CD useful as a probe of protein secondary structure[47,48]. Three samples: cAbLys-A104C (the starting mutant, 'unmodified'), cAbLys-A104-Dha and cAbLys-A104Kme$_2$, were prepared for CD analysis (10 mM pH 8.0 NaPi buffer). The spectra are reported as **Figure 4.14**. It can be seen that no gross alteration in secondary structure has occurred, as the spectra are highly similar and representative of a folded β-sheet rich protein (negative band ~218 nm, positive ~196 nm).

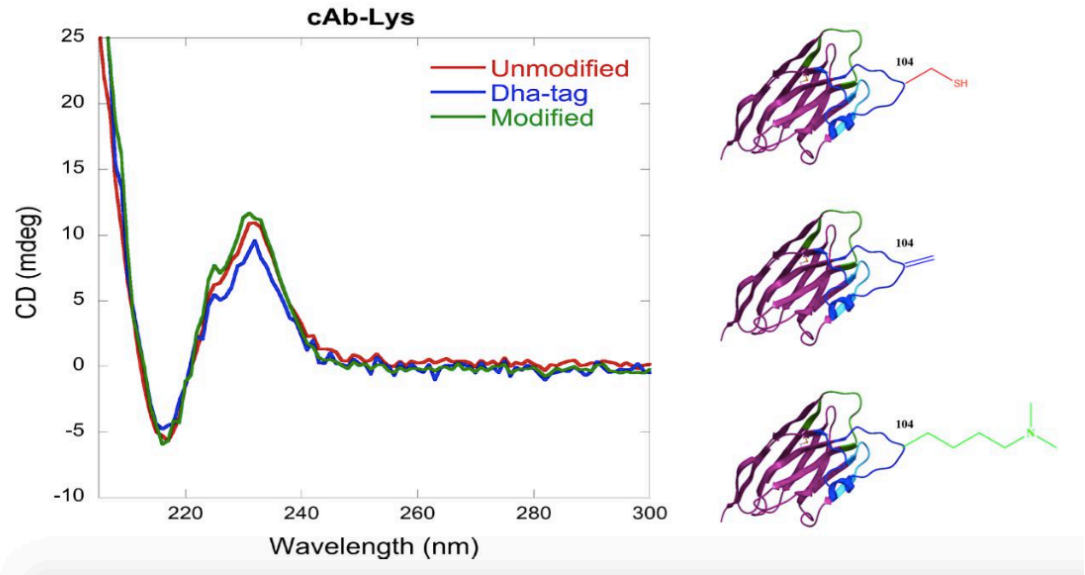


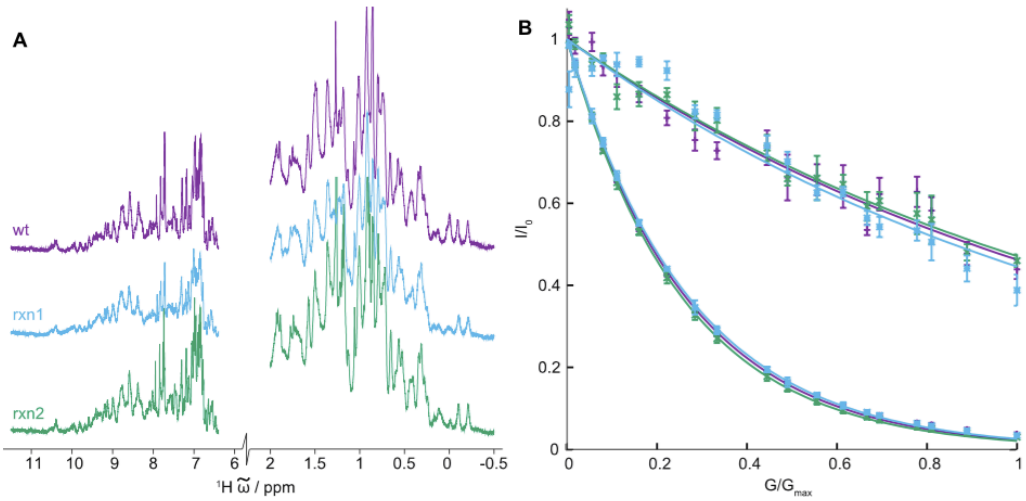
Figure 4.14: CD analysis reveals no gross change in cAbLys secondary-structure during the course of chemical mutation

As CD spectroscopy does not provide information as to protein tertiary structure per se, an alternative method was sought to assess any change in overall folding following chemical modification. Pulsed field gradient NMR spectroscopy can be used to measure the translational diffusion of molecules, including proteins, in solution[49-51]. The diffusion coefficient is correlated with the size and shape of the protein and thus can report on unfolding or oligomerisation events that may arise due to chemical modification.

Wild-type cAbLys, A104Dha and A104Kme$_2$ protein samples were prepared at 100 μM (in 10 mM PBS pH 7.4, 2 mM EDTA, 2 mM NaN3, 5%

D_2O) concentration for ^1H NMR analysis. 1H 1D Watergate[52] NMR spectra were acquired, revealing highly similar spectra for the three protein samples (**Figure 4.15a,** experiment performed by Andrew Baldwin). Notably, analysis of the amide and aliphatic regions revealed retention of the folded state, with abundant resonances observed in the region 8.5-9 ppm (amide resonances), and below 0 ppm (aliphatic region). The presence of such resonances is taken as a diagnostic 'fingerprint' of a folded protein[53]. Moreover, the similar intensity patterns observed in the three spectra also suggest that the overall protein fold is largely conserved. Subtle differences in certain peak intensities were observed, suggesting that the installed modifications lead to modulations in local protein dynamics, a finding likely to have functional relevance but not unexpected given the inherent 'mutational' nature of the modification.

The backbone amide resonance (8-8.15 ppm) was chosen for diffusion analysis by pulsed field gradient stimulated echo experiments[54]. The EDTA present in the buffer was used as an internal control. In this experiment, the signal intensity is expected to decrease as a function of the applied gradient strength. Fitting of the obtained integrated intensities to the Sketsjal-Tanner[55] equation enabled estimation of the diffusion coefficient for each species. The results are shown in **Figure 4.15b** and diffusion coefficients tabulated in **Figure 4.15c.** The decay curves obtained from the wild-type, Dha and Kme_2 modified proteins are highly similar and provide consistent diffusion coefficients. Given the precision of the measurements, these results indicate that no change in the overall fold or oligomerisation state occurred as a consequence of the chemical modification procedures.



C	Amide (8-8.15 ppm) D_T / cm^2s^{-1}	EDTA (3.605-3.615 ppm) D_T / cm^2s^{-1}
wt (A104)	$7.6 \times 10^{-7} \pm 5 \times 10^{-8}$	$3.70 \times 10^{-6} \pm 5 \times 10^{-8}$
A104Dha	$7.4 \times 10^{-7} \pm 3 \times 10^{-8}$	$3.84 \times 10^{-6} \pm 2 \times 10^{-8}$
$A104Kme_2$	$7.9 \times 10^{-7} \pm 7 \times 10^{-8}$	$3.62 \times 10^{-6} \pm 2 \times 10^{-8}$

Figure 4.15: 1D protein NMR reveals overall retention of fold during chemical mutation of cAbLys (Figure courtesy of Andrew Baldwin). a. ^{1}H 1D watergate NMR spectra of cAbLys wild-type (A104), A104Dha ('rxn1') and $A104Kme_2$. b. Pulsed field gradient stimulated echo experiments reveal overall conservation of fold and oligomerisation state following chemical mutation. c. Tabulated translational diffusion coefficients.

p38α modification

As a further extension of the developed methodology to a more complex protein target, p38α kinase was chosen as a target for modification. p38α is a mammalian MAP (mitogen-activated protein) kinase that plays a central role in a number of key signaling pathways[56]**.** Previous work in our laboratory has demonstrated chemical control over p38α activation state using a 'tag-and-modify' strategy[46]. In that work, a cysteine mutant of the key residue Thr180 was prepared and dehydroalanine installed via bis-alkylation elimination using 2,5-dibromohexane diacetamide (DBHDA). Thia-Michael addition of thiophosphate enabled installation of phosphocysteine (pCys) as a mimic of threonine phosphorylation. Notably, this synthetic mimic of the activated native protein was active and could be used *in-vitro* to phosphorylate ATF2, a representative protein substrate of the activated kinase.

As no chemical method exists for installation of the native phosphothreonine at this site, it would be of interest to compare the pCys

mimic previously installed by thia-Michael addition to the C-linked phosphomimics developed in this work (discussed further in **Chapter 6**). As a prelude to this investigation, the stability of p38α to chemical mutation, particularly towards the key radical-mediated C-C bond formation step, was explored. Two mutants, A172C and T180C, arising from the previous study[57], were used in these experiments. A small amount of A172C, containing the introduced cysteine at a non-functional site, was provided (courtesy of co-workers Phin Chooi and Seb Galan) for initial experiments. Dehydroalanine was successfully introduced using the previously reported procedure (**Figure 4.16a**), providing p38α-A172Dha for the key chemical mutation reaction. Initial studies revealed the importance of incubation of this protein at 4°C during degassing to maintain protein stability. Pleasingly, after brief optimization, conditions were found for the successful chemical installation of lysine via reaction with (3-iodopropyl)amine hydroiodide mediated by $NaBH_4$ (**Figure 4.16b)**. Further, the site of modification was confirmed by tryptic digest of the modified protein, followed by LC-MS/MS (**Figure 4.16C**).

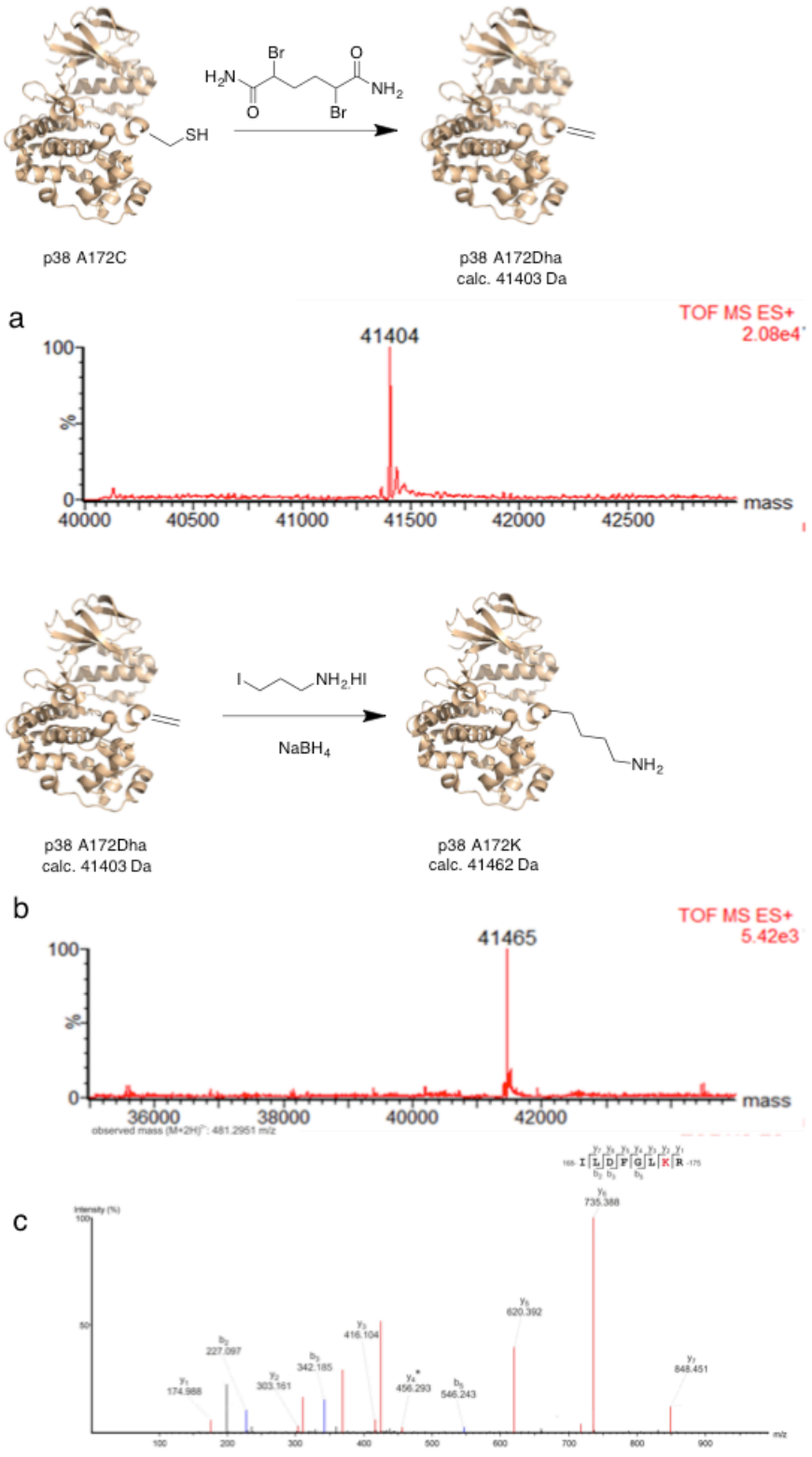


Figure 4.16: Chemical modification of p38α A172Dha. a. Conversion of A172C to dehydroalanine using DBHDA. b. Installation of lysine *via* borohydride-mediated chemical mutation. c. Confirmation of the site of modification using LC-MS/MS. See **Experimental Section** for details.

Following this success, attention was directed towards chemical modification of the T180C mutant, with an eye towards installation of C-C linked phosphomimics at this key site of natural post-translational phosphorylation. The requisite p38α-T180C protein was expressed as a GST-tagged variant and purified from lysate by GST-affinity chromatography[57]. Cleavage of the GST-tag was accomplished by incubating the protein with thrombin, followed by repeated passage through the GST-affinity column to re-bind the cleaved tag. The flow-through, containing cleaved p38α, was concentrated and loaded onto a size-exclusion (SEC) column for further purification by FPLC. Elution from the SEC column provided pure p38α-T180C (see **Experimental Section** for further details of protein expression and purification). Dehydroalanine formation proceeded smoothly following the reported procedure.

With the requisite p38α-T180Dha in hand, attention turned towards the key C-C bond forming radical reaction. The conditions optimized for reaction at 172Dha were readily transferrable and, after brief optimization, lysine could be readily installed at the 180 site by reaction of the protein-Dha with (3-iodopropyl)amine hydroiodide mediated by $NaBH_4$ (**Figure 4.17a**). The effects of this model chemical modification on the secondary structure and folding of p38α were then briefly investigated by CD spectroscopy. For this experiment, p38α-T180C was compared with the chemically mutated p38α-T180K species. Notably, no gross change in protein secondary structure was observed by CD analysis (**Figure 4.17b**), indicating overall retention of folding during the two-step chemical modification procedure.

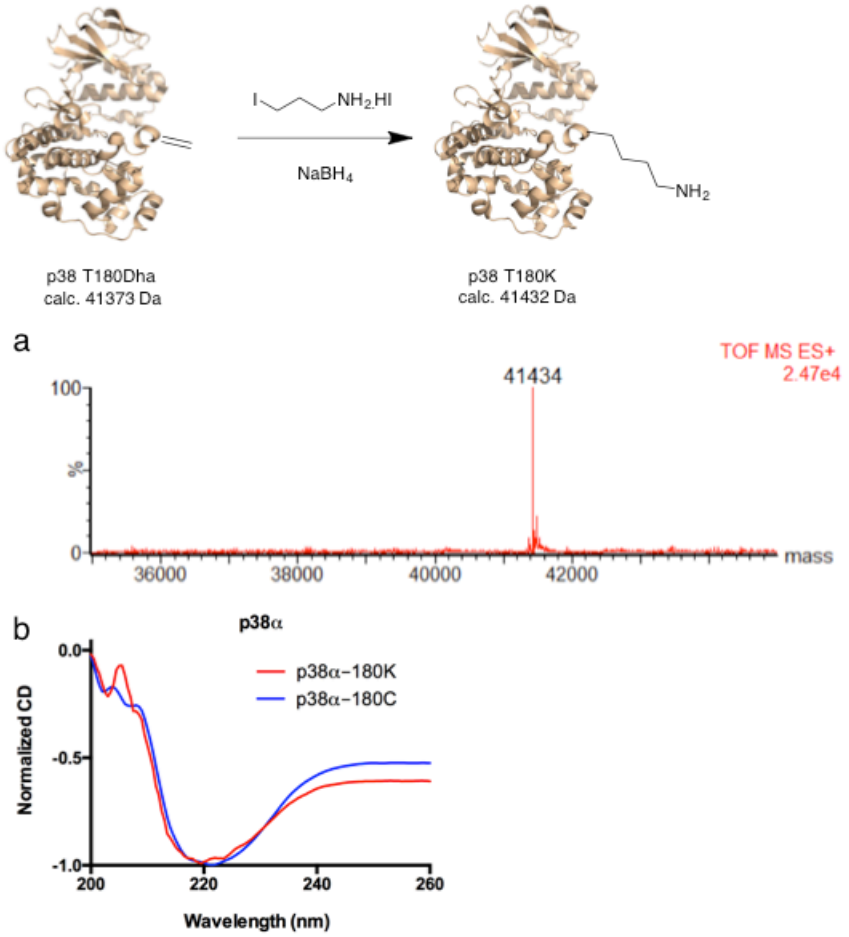


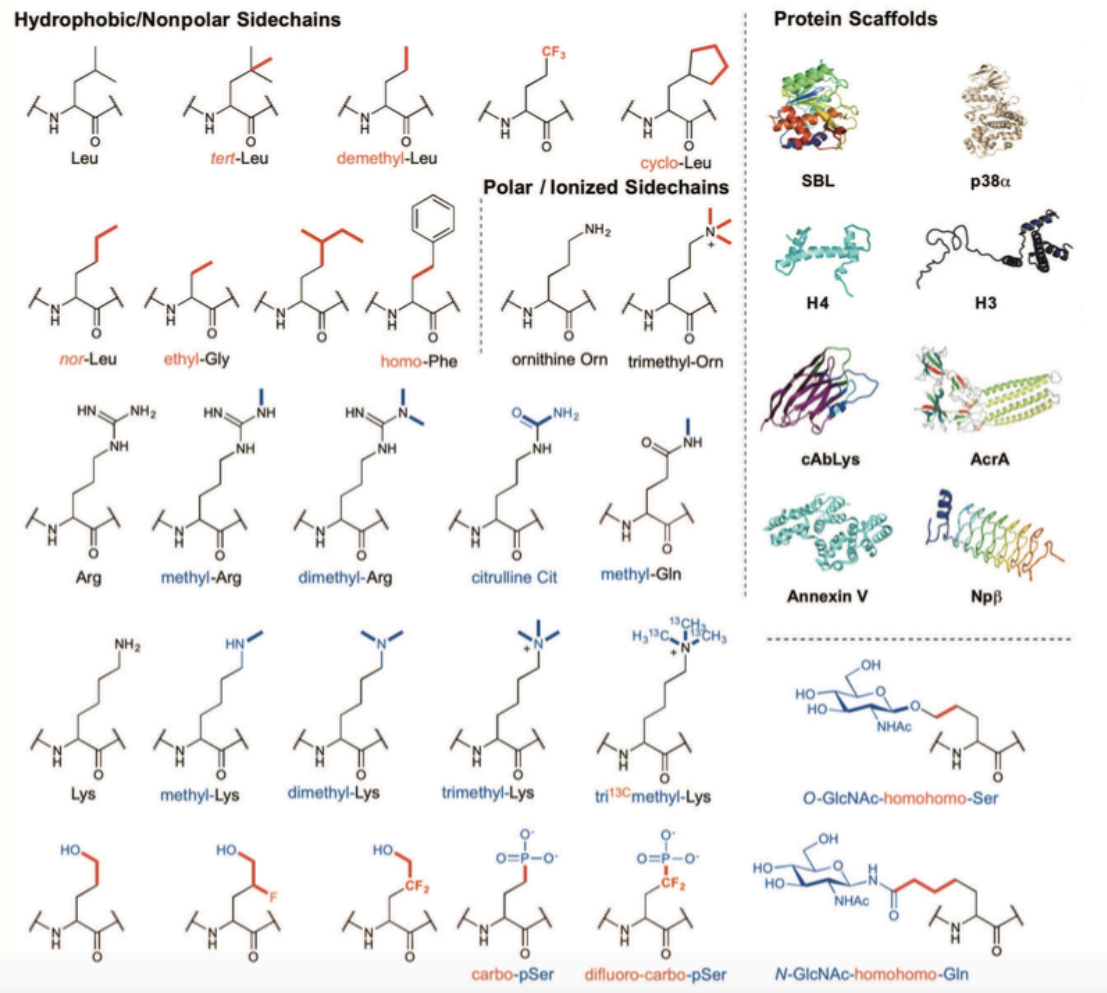
Figure 4.17: Chemical mutation of p38a T180Dha. a. Installation of lysine *via* borohydride mediated chemical mutation. See **Experimental Section** for details. b. Normalised CD analysis reveals overall retention of secondary structure following chemical mutation to lysine. CD was normalized to concentration.

Summary and outlook: Chapter 4

In this chapter, effective reaction conditions for the quantitative 'chemical mutation' of four different proteins were developed. Strikingly, the use of sodium borohydride in aqueous buffer was found to be the optimal reagent to mediate the desired chemistry. Initial investigations into the mechanism of this unusual reaction suggest a radical mechanism, although the question of the initiating species remains open, as unlike in previous reports of borohydride-mediated radical chemistry[24], no external initiator was added in these experiments. Notably, the developed conditions appear strictly selective for dehydroalanine and appear benign with regard to protein

secondary and tertiary structure. These facts bode well for application of the developed chemistry to provide modified proteins for biochemical experiments into protein *function,* the subject of the next two chapters.

Appendix Figure: Scope of borohydride-mediated chemical mutation (from[37])

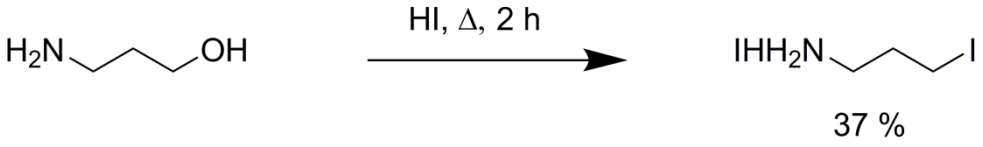


Experimental Section: Chapter 4

Reagent synthesis

(3-iodopropyl)amine hydroiodide

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in [27]



To ice cold hydroiodic acid (55 %, 13.0 mL), 3-aminopropa-1-ol (2.0 mL, 26.6 mmol) was added. The reaction mixture was heated to 125 ºC and stirred for 2 h. The reaction mixture was cooled to room temperature to form a dark red residue, which was recrystallised twice from EtOH/$CHCl_3$ to give the title compound as colourless prisms (3.06 g, 9.77 mmol, 163-164 ºC, 37 %).

^{1}H NMR δH (400 MHz, MeOD) 3.33 (2H, t, J = 7.3, CH_2I), 3.09 (2H, t, J = 7.3, CH_2NH_2) and 2.20 (2H, quin, J = 7.3, CH_2CH_2I)

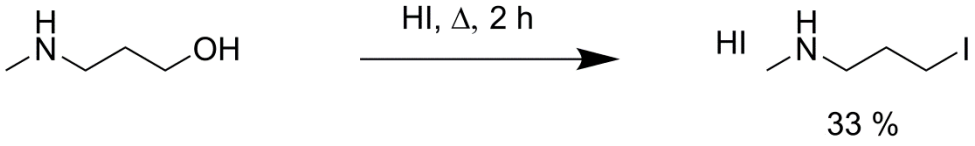
^{13}C NMR δC (100 MHz, MeOD) 40.6 (CH_2NH_2), 31.2 (CH_2CH_2I) and 0.1 (CH_2I)

IR (υ_{max}/cm^{-1}/KBr) 2991, 2288, 1860, 1582, 1588, 1309, 1194, 1120 and 1059

HRMS m/z (ESI^+): Found 185.9774 $[M+H]^+$. C_3H_9NI requires 185.9775.

(3-iodopropyl)methylamine hydroiodide

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in [27]

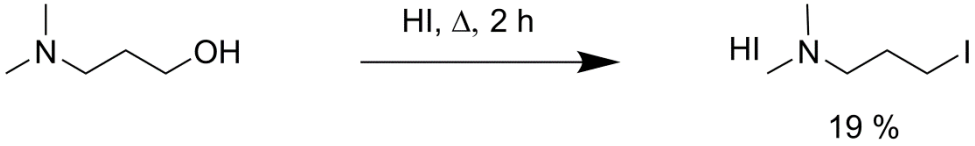


A solution of 3-(methylamino)propan-1-ol (2.00 g, 22.4 mmol) in hydroiodic acid (55 %, 12.0 mL) was heated to 125 ºC and stirred for 2h. The reaction mixture was cooled to room temperature then to -20 ºC overnight, which allowed the dark red residue to crystallise. The crystals

were filtered, washed with Et_2O (100 mL) and then recrystallised ($CHCl_3$/EtOH) and washed with Et_2O (100 mL) to afford the title product as yellow needles, which were dried under reduced pressure (2.41 g, 7.39 mmol, 53-55 ºC, 33 %).

^{1}H NMR *δH* (400 MHz, MeOD) 3.33 (2H, t, *J* = 6.8, CH_2I), 3.16 (2H, t, *J* = 7.8, CH_2NHMe), 2.76 (3H, s, $NHCH_3$) and 2.24 (2H, m, CH_2CH_2I)
^{13}C NMR *δC* (100 MHz, MeOD) 50.1 (CH_2NHCH_3), 33.1 ($NHCH_3$), 30.0 (CH_2CH_2I) and -0.40 (CH_2I)
IR (υ_{max}/cm^{-1}/KBr) 3480, 2935, 2760, 2398, 1457, 1416, 1268, 1203, 1038
HRMS m/z (ESI$^+$): Found 199.9928 $[M+H]^+$; $C_4H_{11}NI$ requires 199.9930

(3-iodopropyl)dimethylamine hydroiodide



19 %

3-dimethylamino-1-propanol (2.0 g, 19.39 mmol) was added to a 25 mL round bottom flask and stirred at room temperature. Hydroiodic acid (55%, unstabilised, 12 mL) was added dropwise to the stirred alcohol, giving a dark-red mixture. The reaction mixture was heated at 125 °C (open air) for 2 hours and then cooled to room temperature. The flask was placed on ice and the product crystallised from the reaction mixture, along with formation of a red solid mass. The solids were filtered and washed with Et_2O. The product was further recrystallised from $CHCl_3$. Remaining iodine was removed by extensive trituration with Et_2O to leave a white solid (1.4 g, 18.2%).
IR (υmax, KBr): 2918, 2686, 1459, 1313, 1187, 1101, 1036, 1004, 959, 843, 727 ^{1}H NMR (400 MHz, MeOD): δ= 2.29 (2H, m, $C\underline{H_2}CH_2I$), 2.95 (6H, s, $N\underline{Me_2}$), 3.28-3.33 (4H, m, contains $C\underline{H_2}NMe_2$ and $C\underline{H_2}I$). ^{13}C NMR (100 MHz, MeOD): δ = -1.69 ($\underline{C}H_2I$), 27.99 ($\underline{C}H_2CH_2I$), 42.39 ($N\underline{Me_2}$), 58.21 ($\underline{C}H_2NMe_2$).
LRMS m/z (ESI+): Found 214.0 [M+H]+; $C_5H_{13}IN$ requires 214.07

Protein chemical modification

Histone H3 (K27C, C110A)

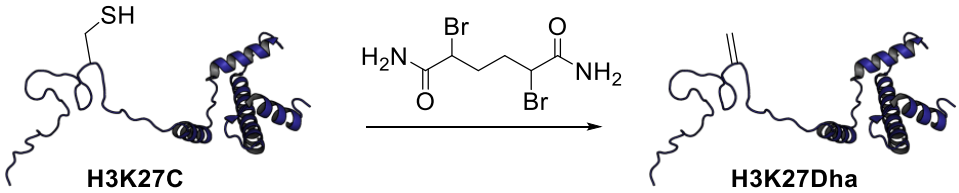
ARTKQTARKSTGGKAPRKQLATKAAR**C**SAPATGGVKKPHRYRPGTVALREIRR

YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN

LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214; Observed mass = 15214

Preparation of H3K27Dha

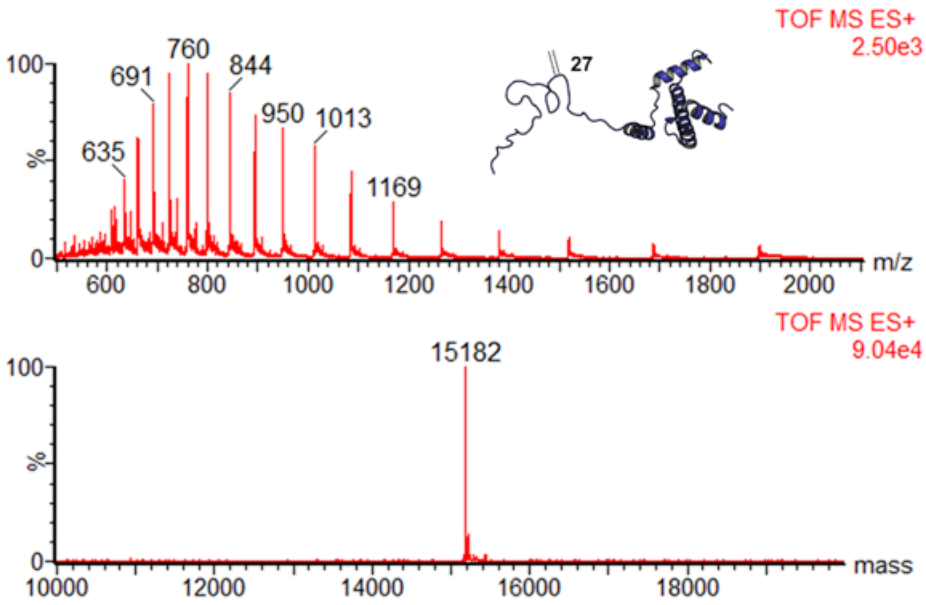


The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass: 15180.7

Observed mass: 15182



Histone H3 (K64C, C110A)

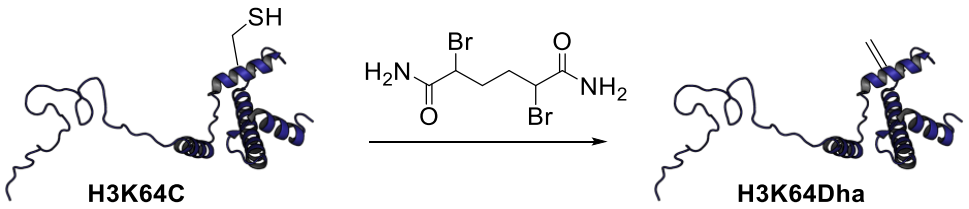
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR

YQKSTELLIR**C**LPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN

LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214; Observed mass = 15214

Preparation of H3K64Dha

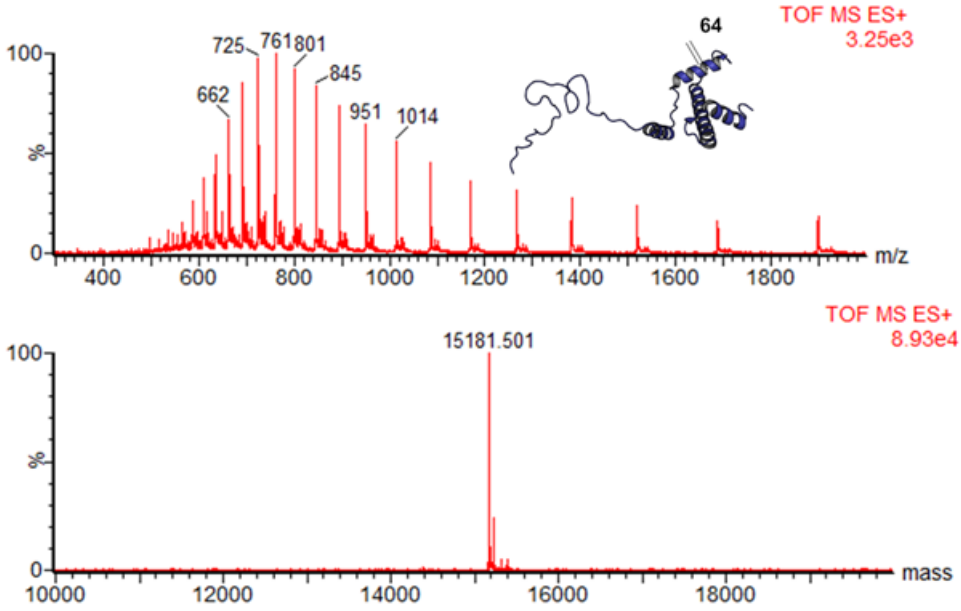


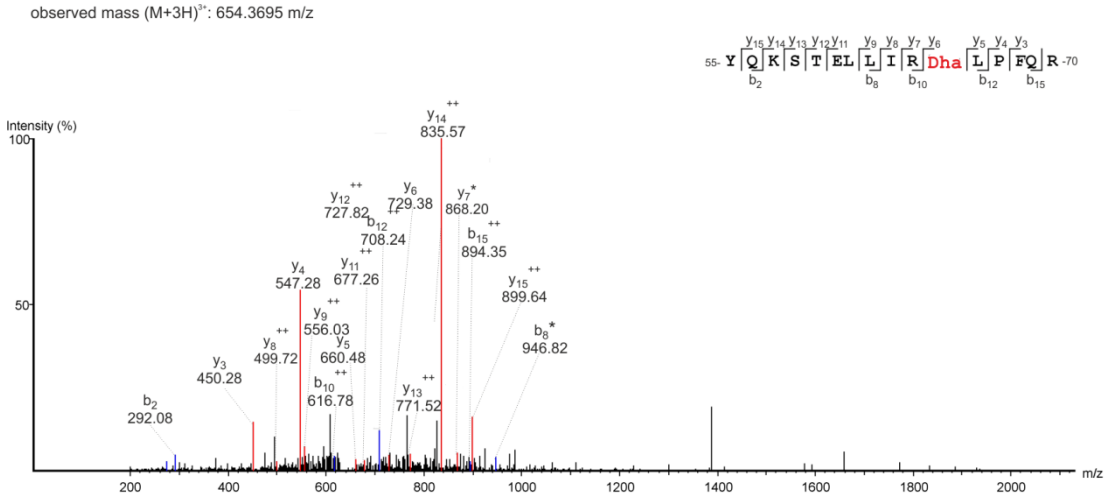
The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass: 15180.7 (H3 K64Dha ,C110A)

Observed mass: 15181.5





MS/MS spectrum for the peptide $_{55}$YQKSTELLIRDhaLPFQR$_{70}$

p38α kinase A172Dha

For expression and purification of p38α A172C, please refer to **supplementary reference[57].**

Protein sequence:

```
        10          20          30          40          50
60
GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
RVAVKKLSRP

        70          80          90         100         110
120
FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE FNDVYLVTHL
MGADLNNIVK

       130         140         150         160         170
180
SQKLTDDHVQ FLIYQILRGL KYIHSADIIH RDLKPSNLAV NEDSELKILD
FGLCRHTDDE

       190         200         210         220         230
240
MTGYVATRWY RAPEIMLNWM HYNQTVDIWS VGCIMAELLT GRTLFPGTDH
IDQLKLILRL

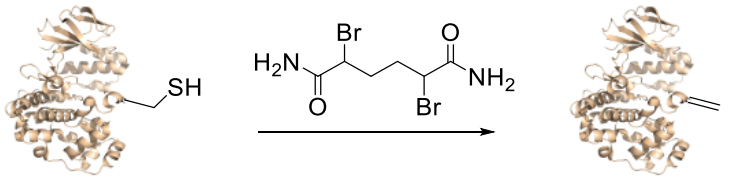
       250         260         270         280         290
300
VGTPGAELLK KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM
LVLDSDKRIT

       310         320         330         340         350
360
AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
VPPPLDQEEMES
```

Calculated mass (p38α A172C): 41437

Observed mass: 41438

Preparation of p38α A172Dha



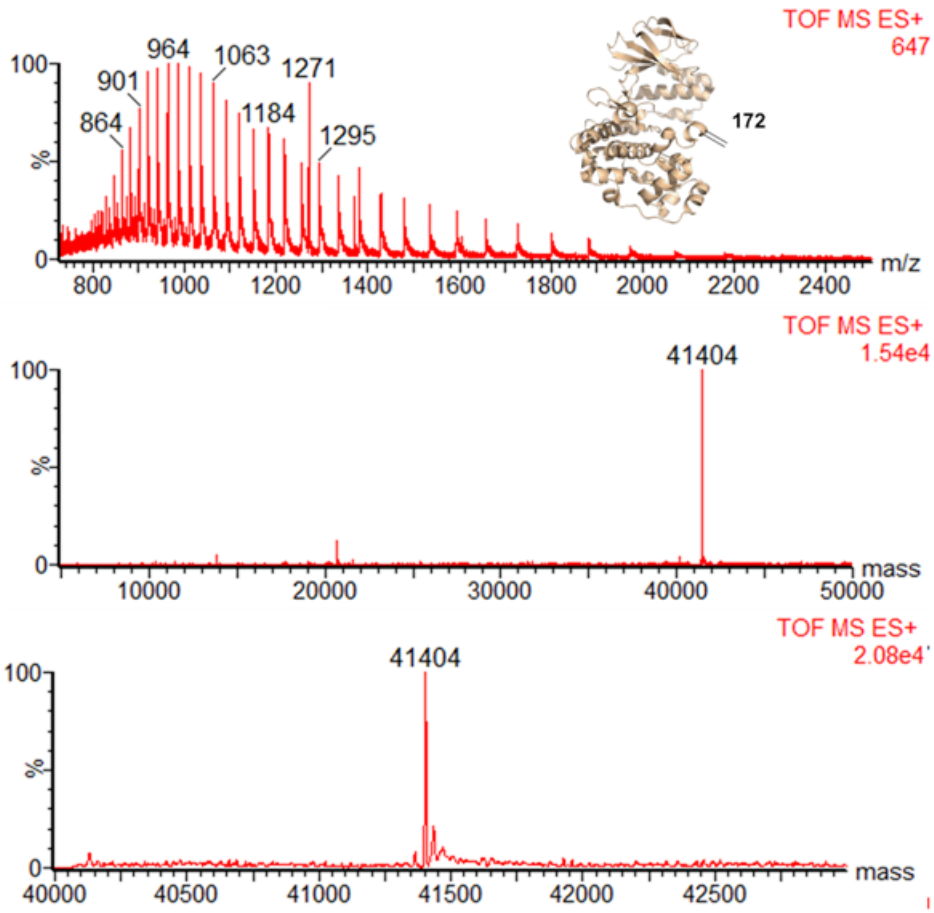
p38A172C **p38A172Dha**

Procedure was adapted from reference[57]

A freshly thawed aliquot of p38α -A172C (100 µL of a 3.5 mg/mL solution, 8.4 nmol, 1.0 equiv.) in storage buffer (50mM HEPES pH 7.8, 50mM NaCl, 5% glycerol, 0.5mM TCEP) was loaded onto a SpinTrap™ G-25 desalting column and eluted with p38 reaction buffer (50mM HEPES pH 7.8).

A suspension of DBHDA (50 µL of a 2.6 mg/mL suspension in reaction buffer, 0.42 µmol, 100 equiv.) was added and the reaction shaken at 37 °C, 600 rpm for 4 h. Analysis of the samples by LC-MS showed quantitative conversion to the desired Dha product after 4 h. The reaction mixture was loaded onto a SpinTrap™ G-25 desalting column and eluted with 50mM HEPES pH 7.5.

ESI-MS for the modified kinase is shown below:
Calculated mass: 41,403 (p38α A172Dha)
Observed mass: 41,404


TOF MS ES+
647
964
1063
901
1271
1184
864
1295
172
100
%
0
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
TOF MS ES+
1.54e4
41404
100
%
0
10000
20000
30000
40000
50000
mass
TOF MS ES+
2.08e4
41404
100
%
0
40000
40500
41000
41500
42000
42500
mass

p38α kinase T180Dha

For expression and purification of p38α T180C, please refer to **supplementary reference[57].**

Protein sequence:

tag
GSMSQERPTF YRQELNKTIW EVPERYQNLS PVGSGAYGSV CAAFDTKTGL
RVAVKKLSRP

```
         70          80          90         100         110
120
FQSIIHAKRT YRELRLLKHM KHENVIGLLD VFTPARSLEE FNDVYLVTHL
MGADLNNIVK

       130         140         150         160         170
180
SQKLTDDHVQ FLIYQILRGL KYIHSADIIH RDLKPSNLAV NEDSELKILD
FGLARHTDDE

       190         200         210         220         230
240
MCGYVATRWY RAPEIMLNWM HYNQTVDIWS VGCIMAELLT GRTLFPGTDH
IDQLKLILRL

       250         260         270         280         290
300
VGTPGAELLK KISSESARNY IQSLTQMPKM NFANVFIGAN PLAVDLLEKM
LVLDSDKRIT

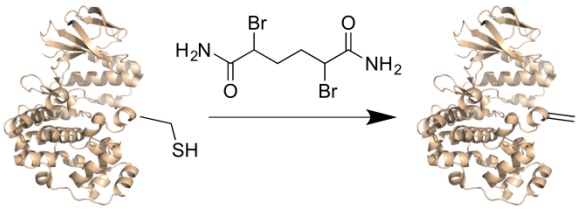
       310         320         330         340         350
360
AAQALAHAYF AQYHDPDDEP VADPYDQSFE SRDLLIDEWK SLTYDEVISF
VPPPLDQEEM
```

ES

Calculated mass (p38α T180C): 41407

Observed mass: 41409

Preparation of p38α T180Dha



p38 T180C **p38 T180Dha**

Procedure was adapted from reference[57].

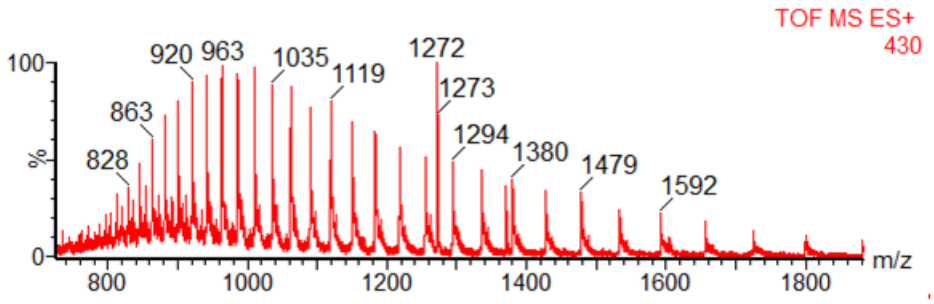
A freshly thawed aliquot of p38α -T180C (100 µL of a 3.5 mg/mL solution, 8.4 nmol, 1.0 equiv.) in storage buffer (50mM HEPES pH 7.8, 50mM NaCl, 5% glycerol, 0.5mM TCEP) was loaded onto a SpinTrap™ G-25 desalting column and eluted with p38 reaction buffer (50mM HEPES pH 7.8).

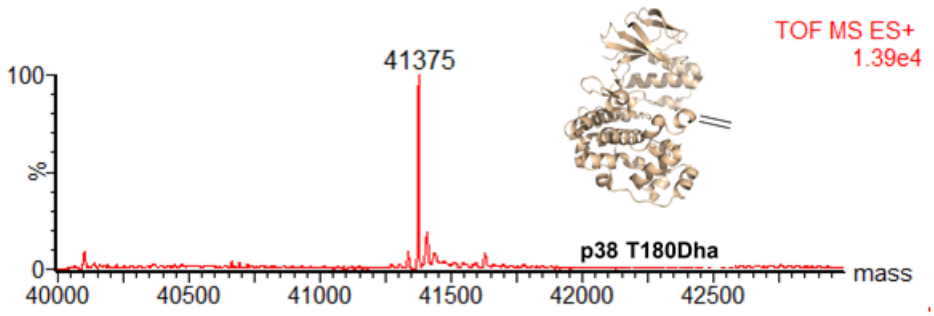
A suspension of DBHDA (50 µL of a 2.6 mg/mL suspension in reaction buffer, 0.42 µmol, 100 equiv.) was added and the reaction shaken at 37 °C, 600 rpm for 4 h. Analysis of the samples by LC-MS showed quantitative conversion to the desired Dha product after 4 h. The reaction mixture was loaded onto a SpinTrap™ G-25 desalting column and eluted with 50mM HEPES pH 7.8.

ESI-MS for the modified kinase is shown below:

Calculated mass: 41,373(p38α T180Dha)

Observed mass: 41,375





cAb-Lys3-A104Dha

Sequence:

```
         10          20          30          40          50
60
DVQLQASGGG SVQAGGSLRL SCAASGYTIG PYCMGWFRQA PGKEREGVAA
INMGGGITYY

         70          80          90         100         110
120
ADSVKGRFTI SQDNAKNTVY LLMNSLEPED TAIYYCAADS TIYCSYYECG
HGLSTGGYGY

        130         140
DSWGQGTQVT VSSGRYPYDV PDYGSGRA
```

Protein expression and purification[45]

Transformation and Plasmid Isolation

7.5 g lysogeny broth (LB) mix (Miller), 4.5 g agar and 300 mL H_2O were mixed and autoclaved. Upon cooling, Kanamycin (50x, 300 µL) was added and mixed well. Plates were poured in a sterile laminar flow hood, each to the volume of approximately 20 mL, covered and left to set.

β-mercaptoethanol (0.6 µL) was added to each of two aliquots of supercompetent XL10-Gold cells in an Eppendorf tube. Either Wild-type (WT) or A104C plasmid DNA (pDNA) was added to one of each prepared cell aliquot and both mixtures were incubated at 37 °C for 30 min. Both samples were then heat pulsed at 42 °C for 30s, cooled on ice for 2 min and sterilised Super Optimal Broth with catabolite suppression (SOC) media (75 µL) was added and the tubes were shaken at 37 °C for 1 hour. Finally, each sample was inoculated onto kanamycin-containing agar plates and cultured overnight at 37 °C. A colony was picked from each plate and

transferred with a sterile pipette to LB broth (10 mL) (with kanamycin) and incubated overnight at 37 °C with shaking. Plasmid DNA was harvested with a QIAprep Spin Miniprep Kit (QIAGEN), according the manufacturer's instructions. Final concentrations of plasmid (NanoDrop): 86 ng μL^{-1} WT; 79 ng μL^{-1}.

The plasmids were sequenced with proprietary UCB primers that accompany the proprietary UCB expression vector. BLAST was used to confirm the desired gene sequences were within the UCB mammalian expression vector.

Large scale preparation of plasmid DNA was performed using an EndoFree Plasmid Giga Kit (QIAGEN), from 4 x 1 L LB cultures, according to the manufacturer's instructions. Final yield of plasmid DNA: 3.8 mg A104C, 4.1 mg WT.

Protein Expression

CHOSXE cells (UCB, at 25th passage, and split twice weekly) were counted (300 µL) at 5.6 x 10^6 cell mL^{-1}, 99.8% viability, measured with Trypan Blue (Cedex). Cells were harvested by centrifugation (15 min, 1500 rpm, 1 °C). Media was prepared (ProCHO, Gibco, Life Technologies) by addition of antibiotic-antimycotic (2 mL, 100x, 2%, Gibco, Life Technologies) and GlutaMAX supplement (100x, 10 mL, Gibco, Life Technologies). Cells were washed with Earle's balanced solution (EBS, 25 mL, Sigma Life Science) and re-pelleted (15 min, 1500 rpm, 1 °C). Cells were once again re-suspended (EBS, 10 mL) and mixed with each pDNA (700 µL). Cells were held on ice and electroporated in batches (800 µL) at 370 V, 9.6 A, then transferred to the prepared media (1 L ProCHO + antibiotic-antimycotic + GlutaMAX). Post-electroporation cell count indicated approximately 10% survival rate. Cultures were incubated at 37 °C, 140 rpm, 8% CO_2 for 24 hours before the temperature was lowered to 32 °C. After 4 days, sodium butyrate to final 3 mM concentration, and Fit A 2.8%, Fit B 0.4 %, were added and incubation continued for 14 days.

Protein Purification

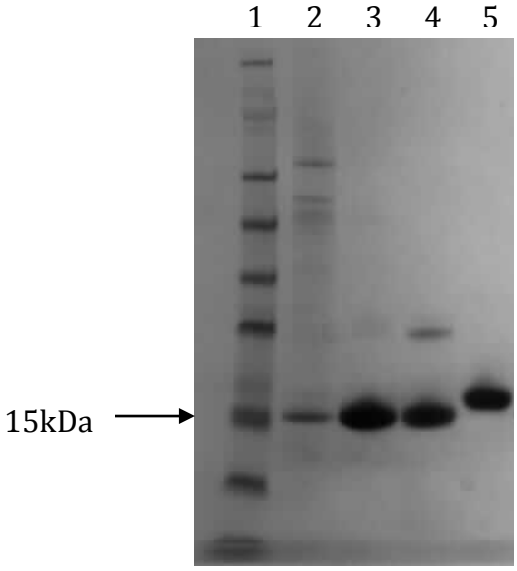
HisTrap binding buffer: PBS, 300 mM NaCl, 10 mM imidazole, pH 8.0
HisTrap elution buffer: PBS, 300 mM NaCl, 250 mM imidazole, pH 8.0

The FPLC was washed with 0.1M NaOH for sterilisation and then washed with TBS until pH 7.5 achieved, prior to use for the first time. The system was then equilibrated with binding buffer (flow rate 5 mL/min) until pH 8.0 and the 5 mL HisTrap excel column was then equilibrated with 5 CVs of binding buffer. To the supernatant containing the desired protein was added concentrated NaCl and imidazole solutions such that the final concentration of both approximately matched that of the binding buffer. 1L of protein-containing supernatant was loaded onto the 5 mL HisTrap excel column overnight at 1 mL/min, with flow-through collected in a sterile container. The column was then washed with 7 CVs of binding buffer (3 mL/min) to remove non-specifically bound proteins until a steady, low, UV absorbance (280 nm) was achieved. A step elution to 10% elution buffer until a steady UV absorbance was then used (approximately 5 CVs), then elution was stepped to 100% elution buffer over 15 CVs (all at 3 mL/min) to ensure clean and complete elution of the bound protein, collected by automated fraction collector into a 96 well plate (2 mL fractions).

Purified protein containing fractions identified by SDS-PAGE gel analysis were pooled together and 5% w/w EDTA in PBS pH 8.0 was added (1 mL for every 25 mL of protein solution) to chelate any leached Ni from the column – without this step it was observed that cAb-Lys3-WT and cAb-Lys3-A104C precipitated from solution on buffer exchange removal of imidazole. The resulting solution was spin concentrated (Centricon Plus-70 Centrifugal Filters, 10 KDa, Merck Millipore, Cat.: 701008) to ~15 mL total volume, noting a yellow colour to the solution on concentration. The cAb-Lys3-WT solution was then passed (3 mL/min, not more than 10 mL at a time) through a desalting column (HiPrep 26/10 Desalting), equilibrated in PBS pH 7.4, to yield the final protein solution. The cAb-Lys3-A104C solution was passed (1 mL/min, not more than 10 mL at a time) through a gel filtration column (HiLoad 16/600 Superdex 75pg), equilibrated with

PBS pH 7.4, to purify, desalt and buffer exchange to yield the final protein solution. The concentration of each was measured by nanodrop and the stock was split into 1 mg/mL aliquots (0.5 and 1 mL) and snap frozen in liquid nitrogen for future use. The final protein solutions were analysed by SDS-PAGE gel. Total yield of cAb-Lys3-WT estimated at 60 mg and yield of cAb-Lys3-A104C at 140 mg.

Dehydroalanine formation from cAb-Lys3-A104C[37,45]



NuPAGE 4-12% Bis-Tris SDS-PAGE Gel of: 1 = Ladder, 2 = Supernatant, 3 = Purified cAbLys3-WT, 4 = Purified cAbLys3-A104C (non-reducing loading buffer, some disulfide bridged dimer is present), 5 = Purified cAbLys3-A104C (reducing loading buffer)

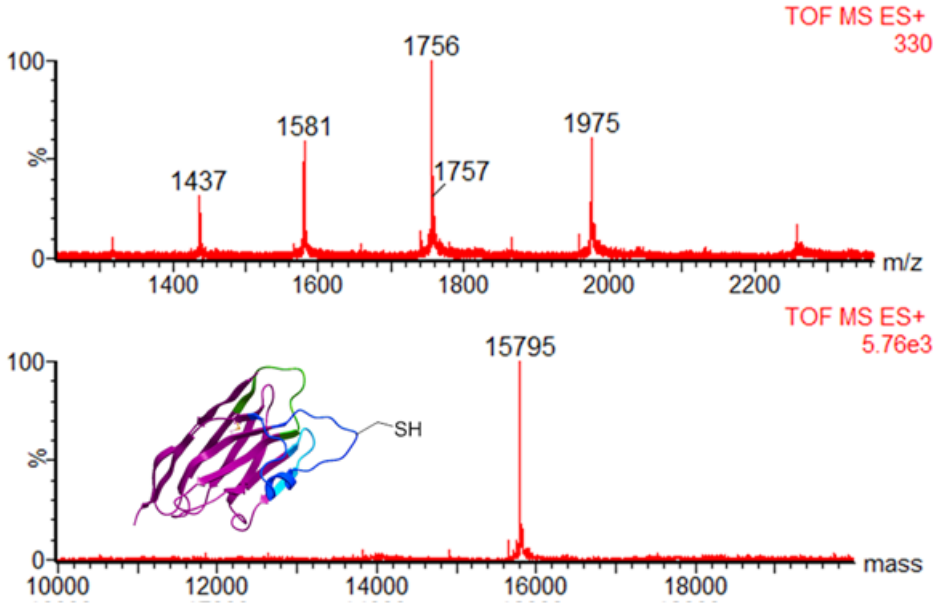
Reduction of cAb-Lys3-A104C

A 0.5 mL solution of cAb-Lys3-A104C (1 mg/mL, PBS pH 8.0) was treated with DTT (1000 equiv.) for 15 minutes at room temperature in an eppendorf tube, with rotation, then passed through a PD MiniTrap G-25 (GE Healthcare Life Sciences, previously equilibrated with 4 x 2 mL of PBS pH 8.0). The sample was eluted with 1 mL of PBS pH 8.0 and used immediately in the next step.

ESI-MS spectrum for the modified nanobody is shown below:

Calculated mass (cAb-Lys3 A104C): 15797

Observed mass: 15795



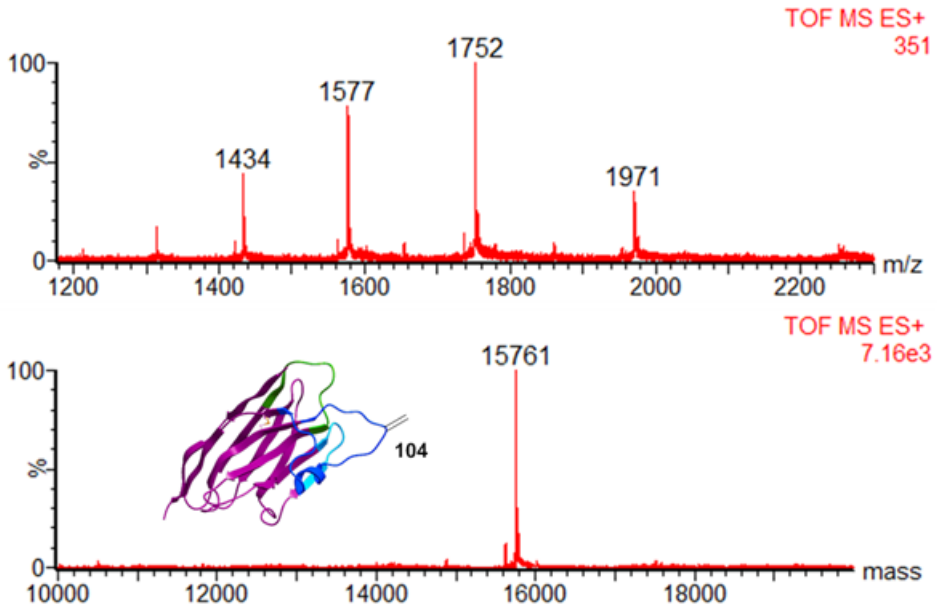
cAb-Lys3-A104C-Dha formation

The reduced cAb-Lys3-A104C sample was pipetted into an Eppendorf tube containing 1000 equivalents of DBHDA and shaken at 650 rpm at 37 °C for 5 hours. The solution was then centrifuged at 4000 rpm for 2 minutes to collect undissolved DBHDA and the supernatant passed through a PD MiniTrap before being moved into the glovebox at 4°C overnight for the solution to de-gas in an N_2 atmosphere.

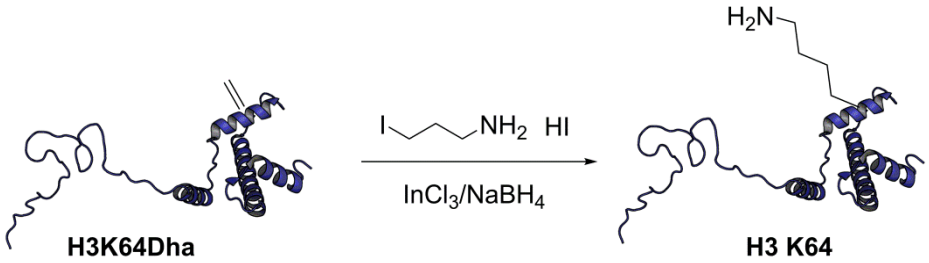
ESI-MS spectrum for the modified nanobody is shown below:

Calculated mass (cAb-Lys3 104Dha): 15763

Observed mass: 15761



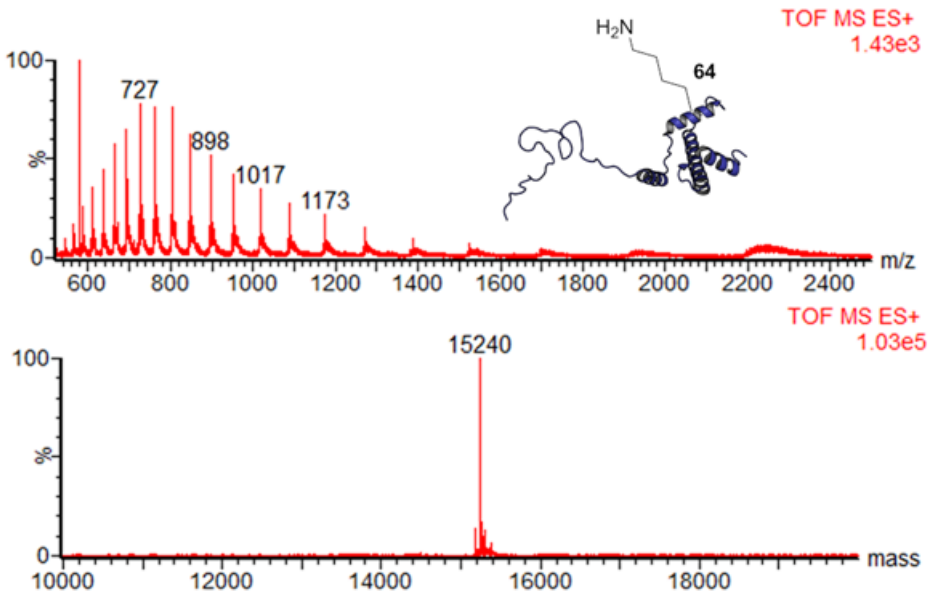
Installation of lysine from dehydroalanine H3 K64



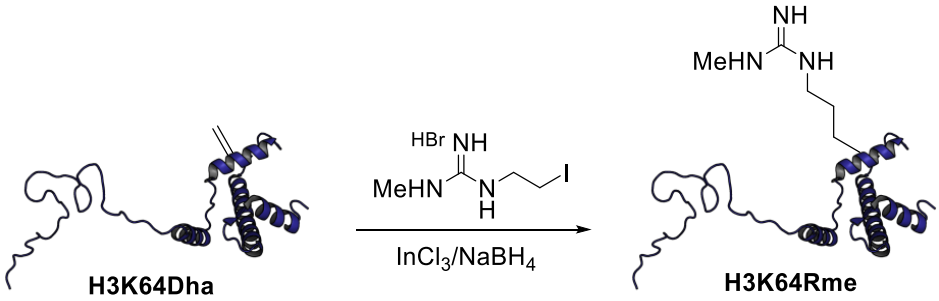
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K64Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was used to dissolve 7 mg (3-iodopropyl)amine hydroiodide **S1** and 2 mg of powdered $InCl_3$. The resultant pale yellow solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification and decolouration. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass: 15238.8 (H3 K64, C110A)
Observed mass: 15240



Installation of methylarginine H3 K64R$_{me}$



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K64Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

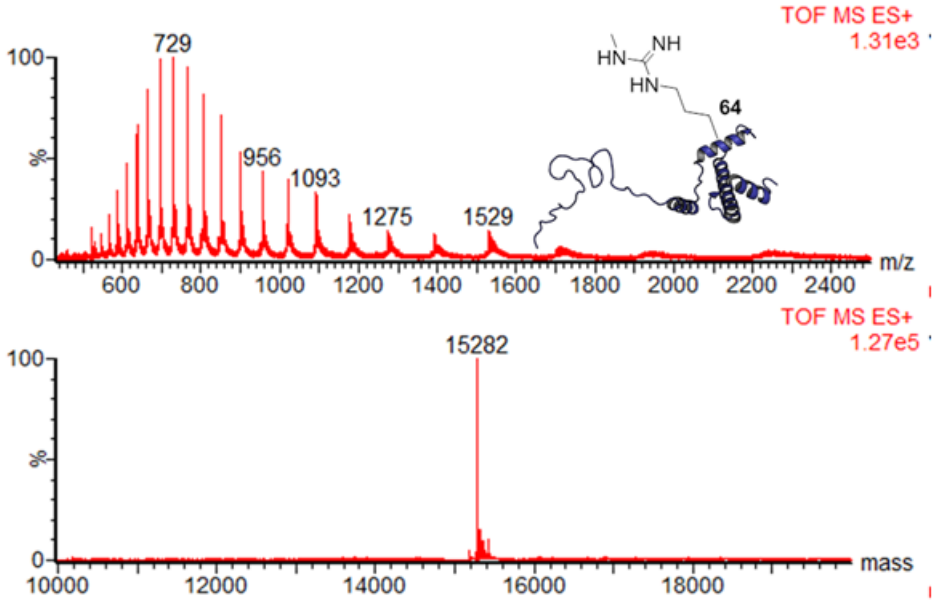
The following morning, within the glovebox, the degassed protein solution was used to dissolve 40 mg 3-(2-iodoethyl)-1-methylguanidine **S15** and 5.8 mg of powdered $InCl_3$. The resultant pale yellow solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 2 mg powdered $NaBH_4$. The tube was gently shaken by

hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification and decolouration. The reaction was shaken at room temperature for two hours before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

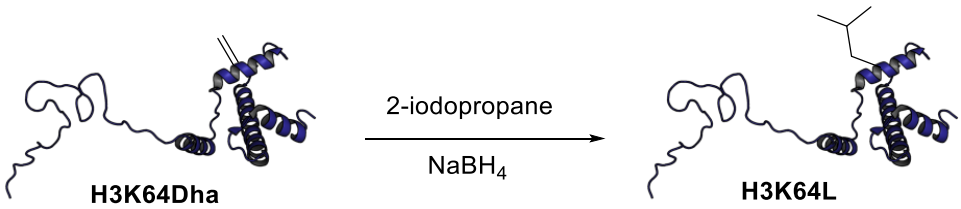
ESI-MS for the modified histone is shown below:

Calculated mass (H3K64Rme C110A): 15280.8

Observed mass: 15282



Installation of leucine from dehydroalanine (H3 K64L)

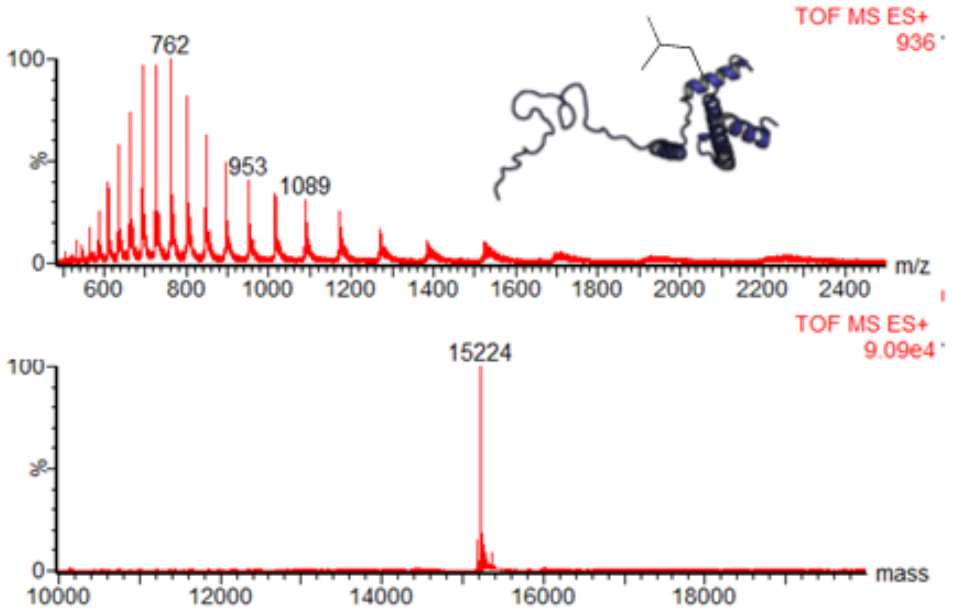


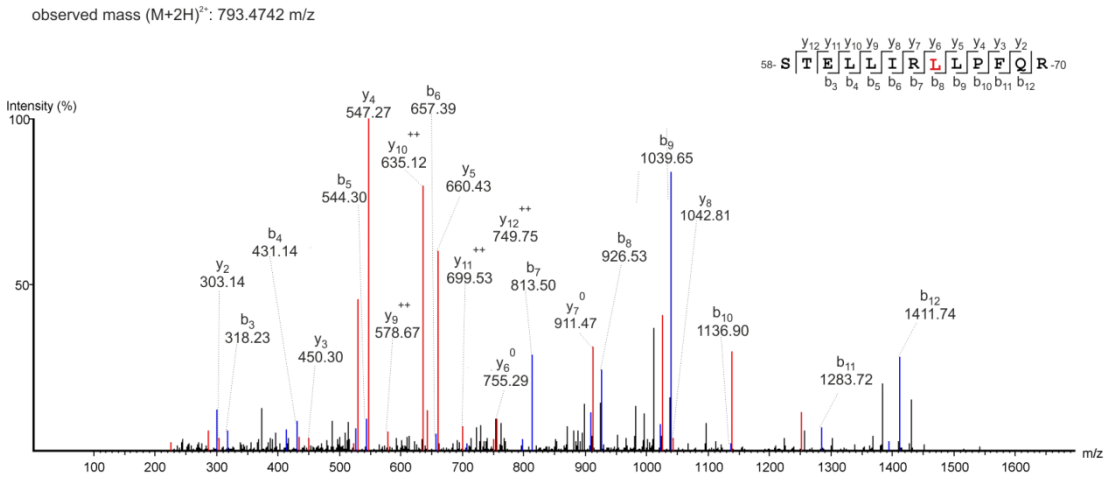
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K64Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant

protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was combined with 1.3 µL 2-iodopropane. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

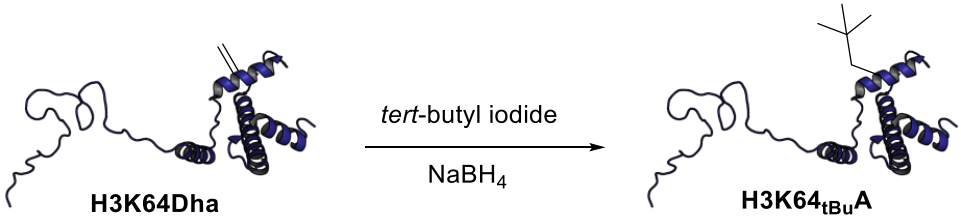
ESI-MS for the modified histone is shown below:
Calculated mass: 15223.7 (H3 K64L ,C110A)
Observed mass: 15224





MS/MS spectrum of the peptide $_{58}$**STELLIRLLPFQR**$_{70}$

Installation of *tert*-butylalanine from dehydroalanine H3 K64



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K64Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

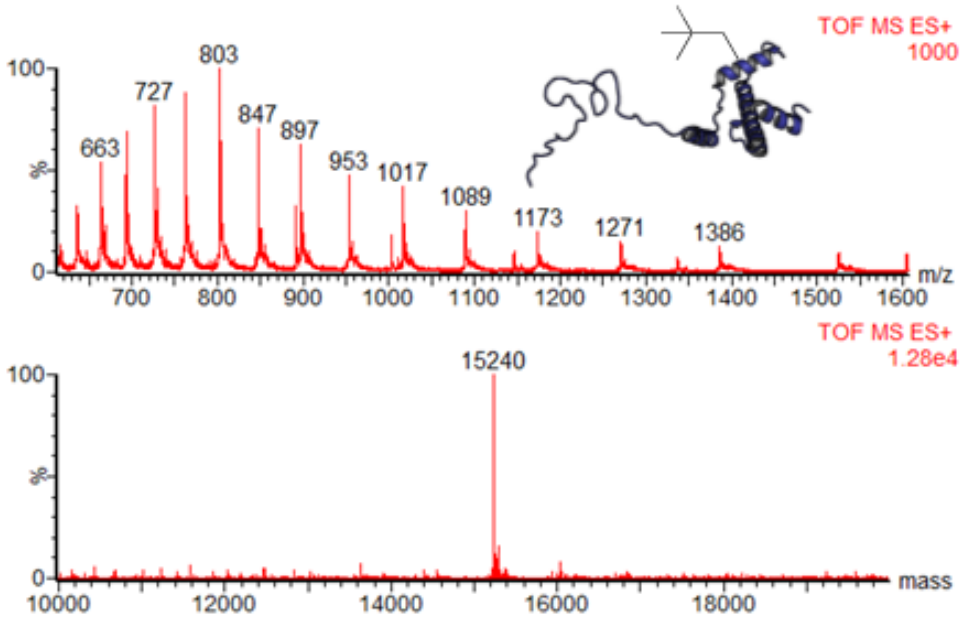
The following morning, within the glovebox, the degassed protein solution was combined with 3.2 µL *tert*-butyl iodide. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched

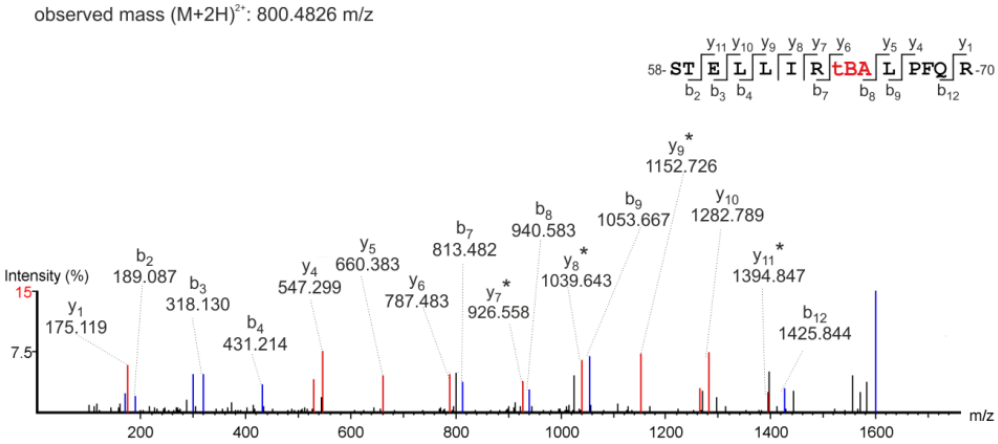
by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:

Calculated mass: 15238.82 (H3 K64$_{tBu}$A, C110A)

Observed mass: 15240

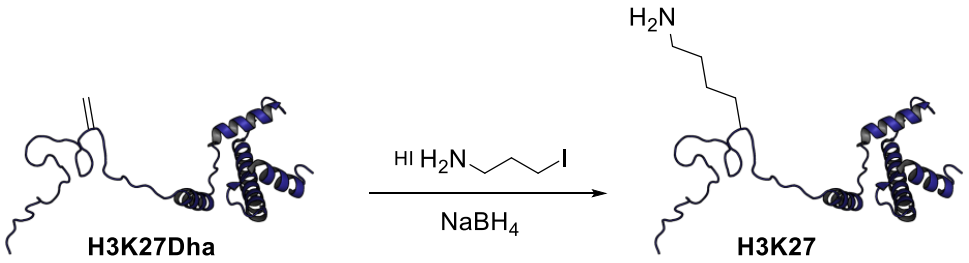




MS/MS spectrum of the peptide $_{58}$**STELLIR$_{tBu}$ALPFQR$_{70}$**

Chemical mutagenesis from Histone H3K27Dha

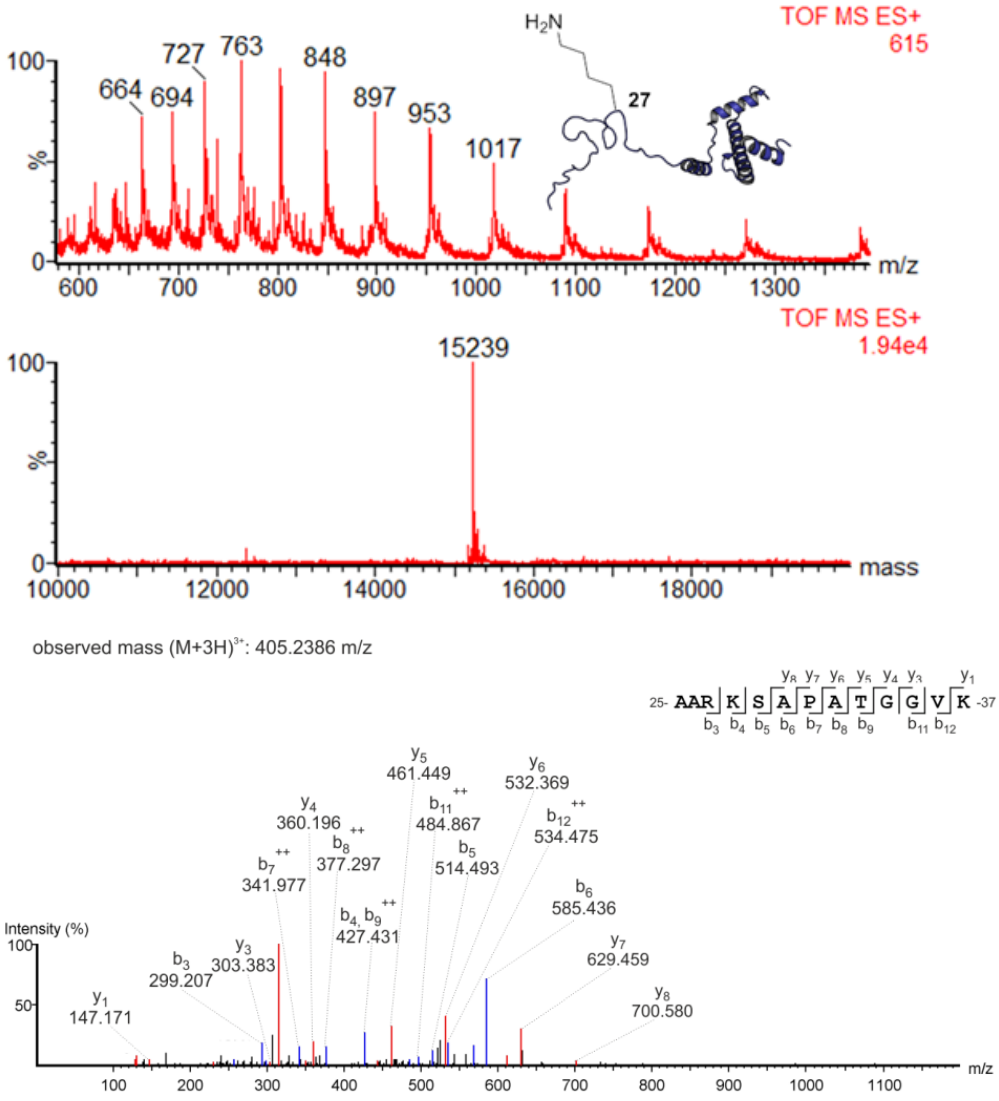
Installation of lysine from dehydroalanine H3 K27



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

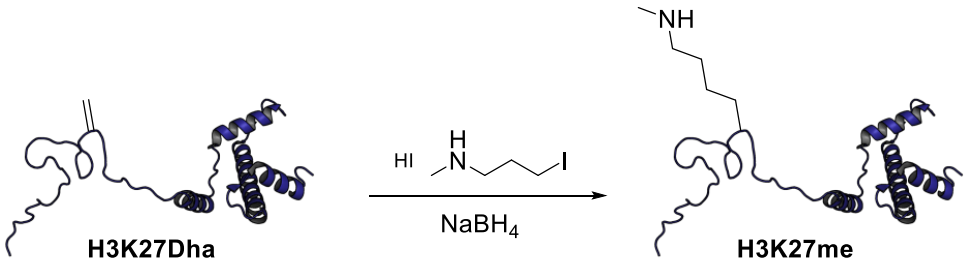
The following morning, within the glovebox, the degassed protein solution was used to dissolve 6 mg (3-iodopropyl)amine hydroiodide **S1**. The resultant pale yellow solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification and decolouration. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below.
Calculated mass: 15238.8 (H3K27 C110A)
Observed mass: 15239



MS/MS spectrum of the peptide $_{25}$**AARKSAPATGGVK**$_{37}$

Installation of monomethyllysine H3 K27$_{me}$

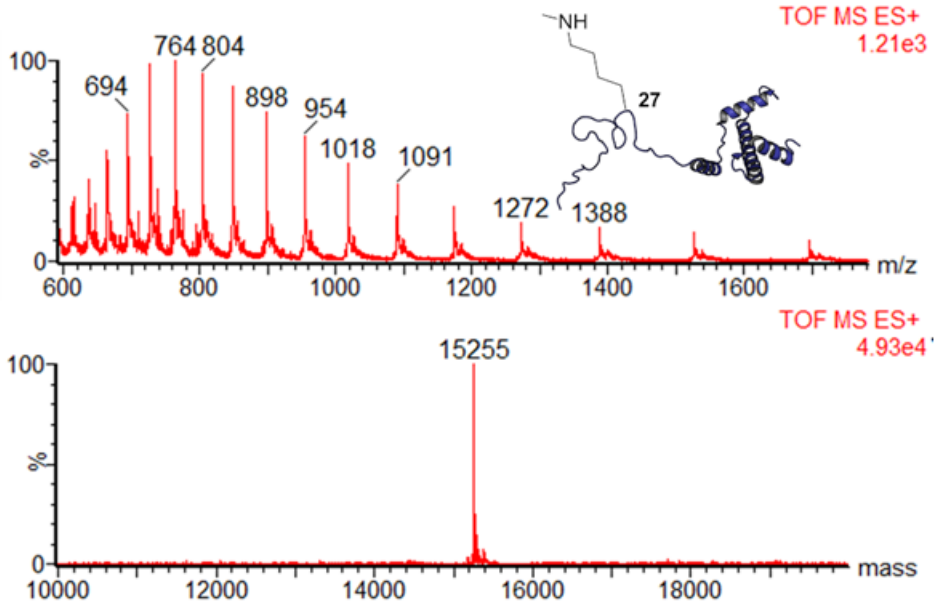


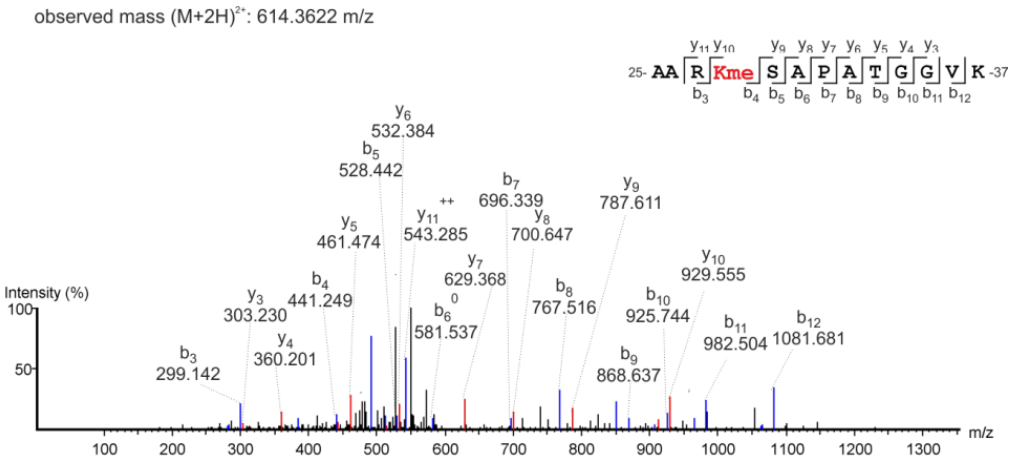
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer)

to give a final protein concentration of 1 mg/mL (~66 μM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was used to dissolve 7 mg (3-iodopropyl)methylamine hydroiodide **S4**. The resultant pale yellow solution was gently mixed by pipette. The entire volume (~200 μL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification and decolouration. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

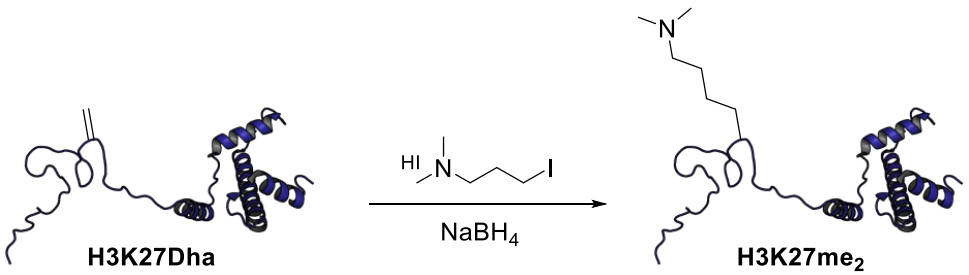
ESI-MS for the modified histone is shown below.
Calculated mass: 15253.8 (H3 K27me, C110A)
Observed mass: 15255





MS/MS spectrum of the peptide $_{25}$**AARKmeSAPATGGVK**$_{37}$

Installation of dimethyllysine H3 K27$_{me2}$



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

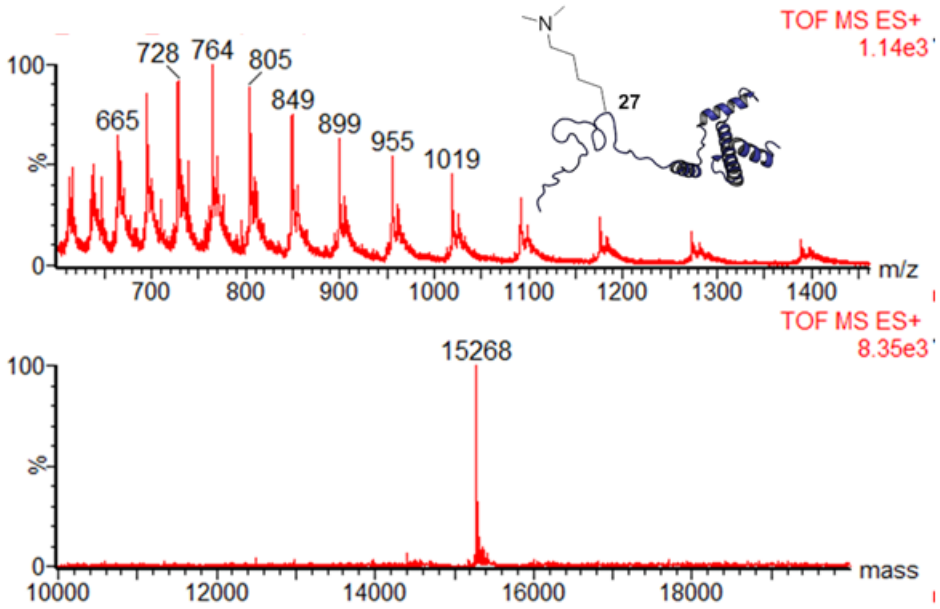
The following morning, within the glovebox, the degassed protein solution was used to dissolve 8 mg (3-iodopropyl)dimethylamine hydroiodide S5. The resultant pale yellow solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification and decolouration. The reaction was shaken at 4°C for 30

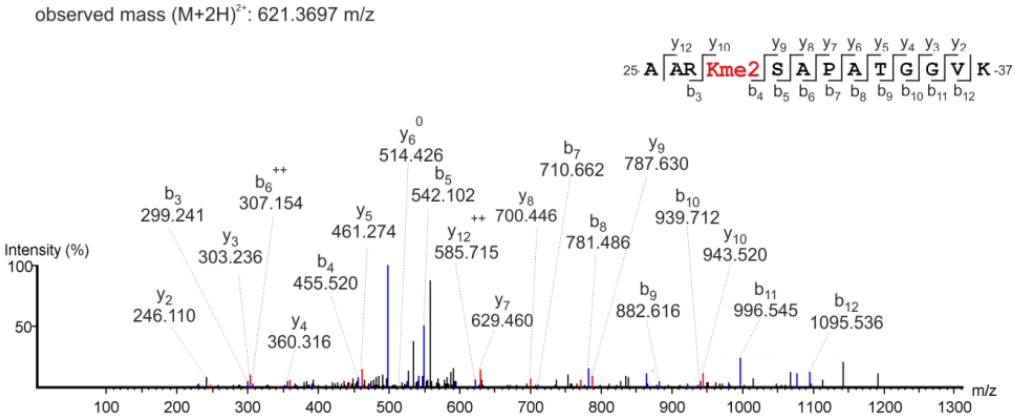
minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below.

Calculated mass: 15267.8 (H3 K27me$_2$, C110A)

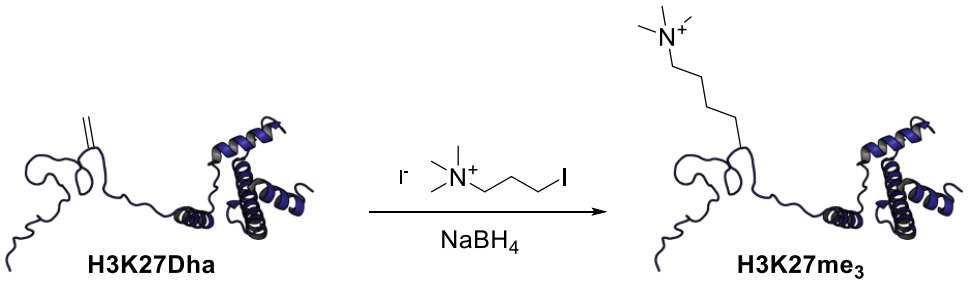
Observed mass: 15268





MS/MS spectrum of the peptide $_{25}$**AARKme2SAPATGGVK**$_{37}$

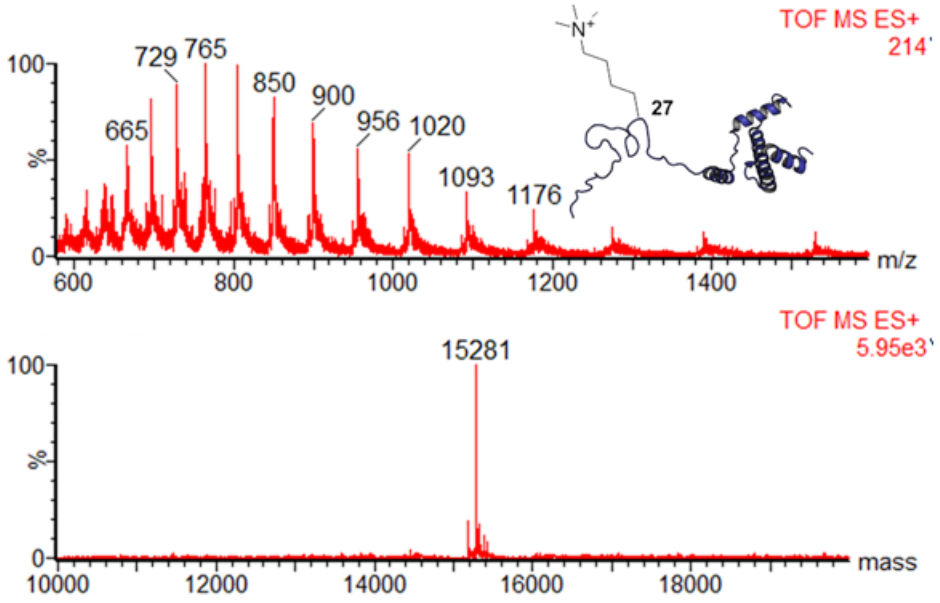
Installation of trimethyllysine H3 K27$_{me3}$



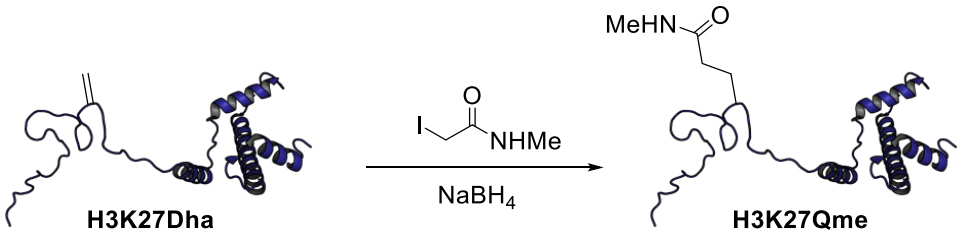
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was used to dissolve 8 mg (3-iodopropyl)trimethylamine hydroiodide **S6**. The resultant pale solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass: 15282.8 (H3 K27me$_3$, C110A)
Observed mass: 15281



Installation of methylglutamine H3 27Q$_{me}$



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

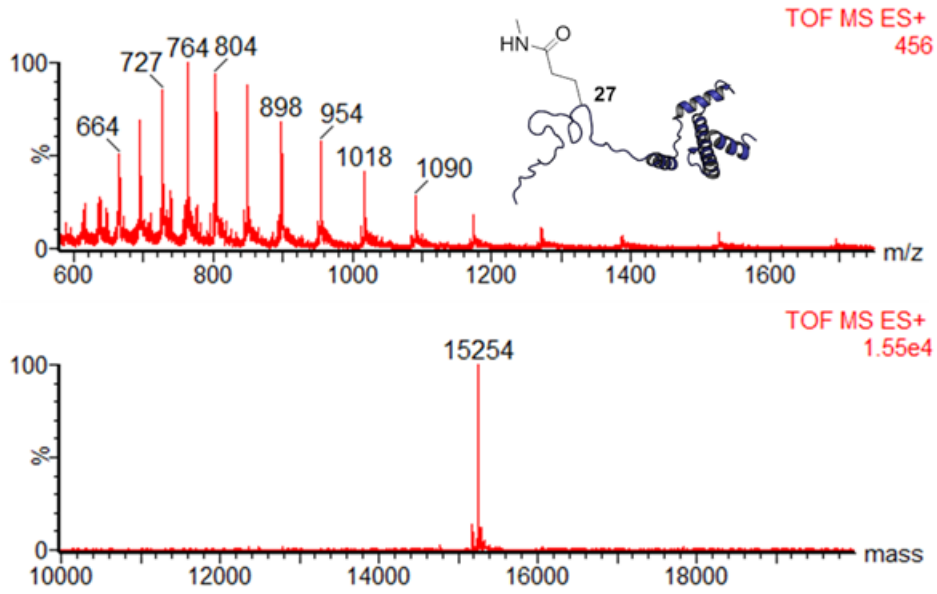
The following morning, within the glovebox, the degassed protein solution was used to dissolve 1.3 mg 2-iodo-*N*-methylacetamide **S10**. The resultant pale solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.25 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread

through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

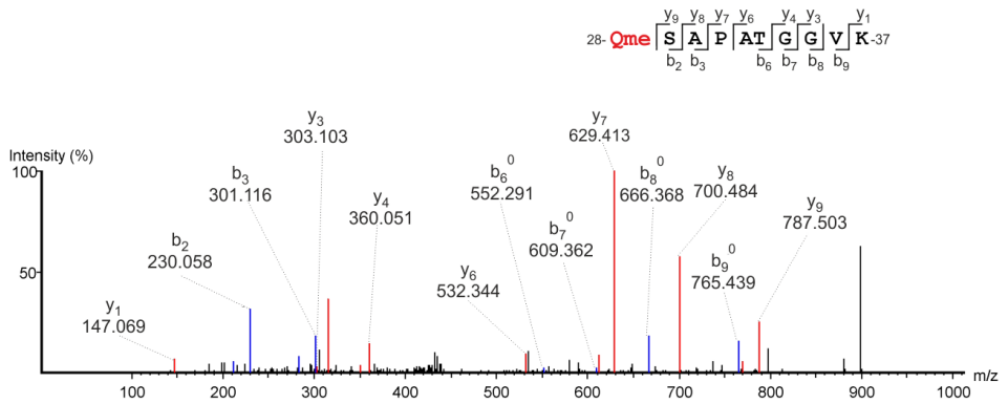
ESI-MS for the modified histone is shown below:

Calculated mass: 15253.7

Observed mass: 15254

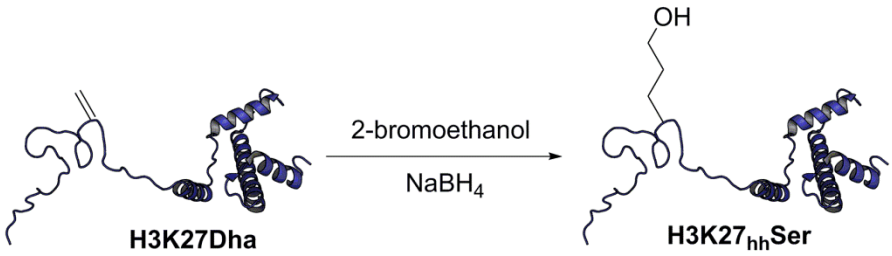


observed mass $(M+2H)^{2+}$: 465.2563 m/z



MS/MS spectrum of the peptide $_{28}$**QmeSAPATGGVK**$_{37}$

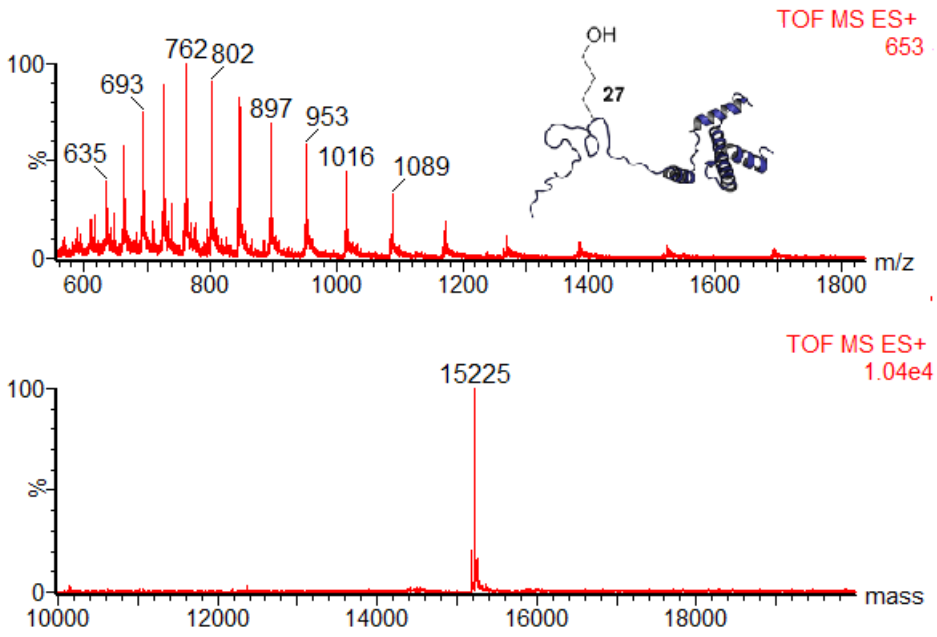
Installation of bis-homoserine H3 27_{hh}Ser



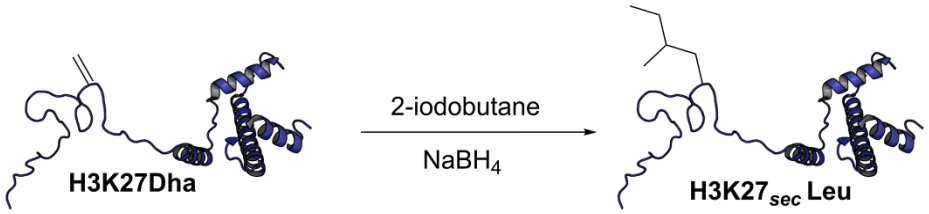
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was combined with 3 µL of 2-bromoethanol. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass: 152227 (H3K27_{hh}Ser C110A)
Observed mass: 15225



Installation of *sec*-Leucine at site H3 27



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

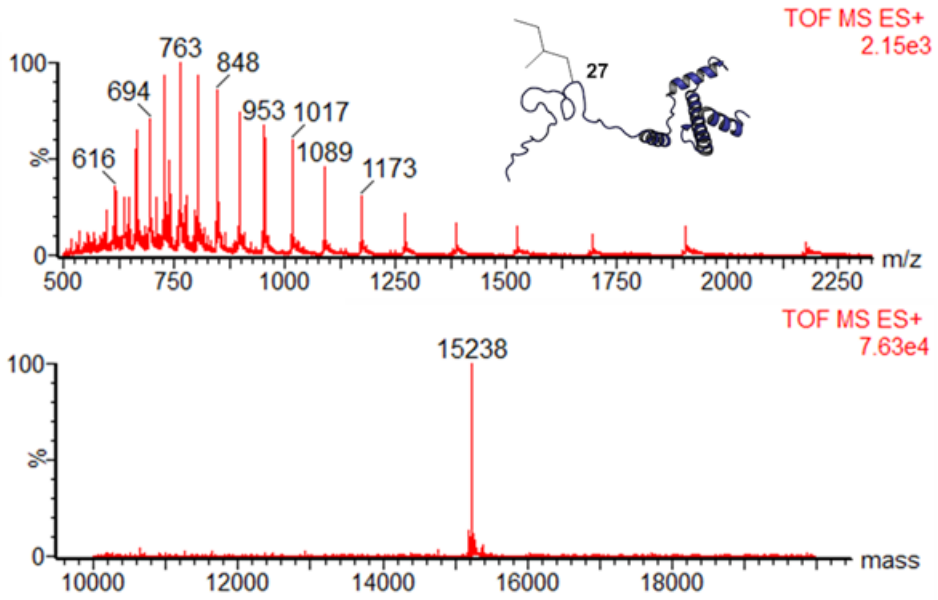
The following morning, within the glovebox, the degassed protein solution was combined with 4 µL of 2-iodobutane. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution,

followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

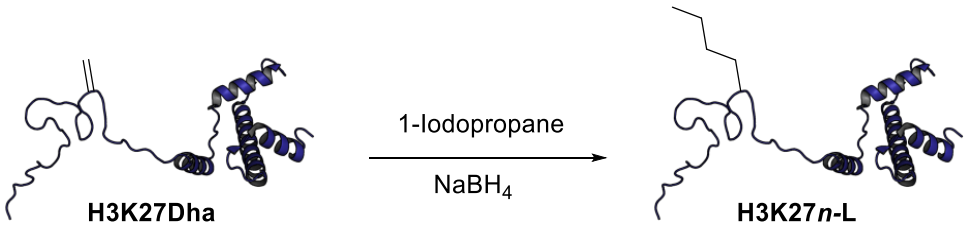
ESI-MS for the modified histone is shown below:

Calculated mass: 15238.8

Observed mass: 15238



Installation of norleucine at site H3 27



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant

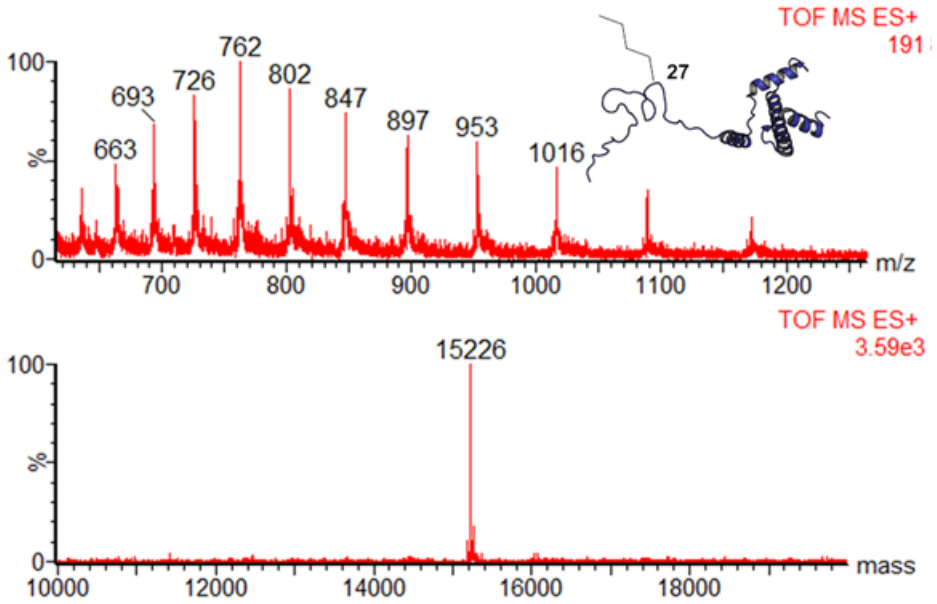
protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was combined with 2 µL of propyl iodide. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

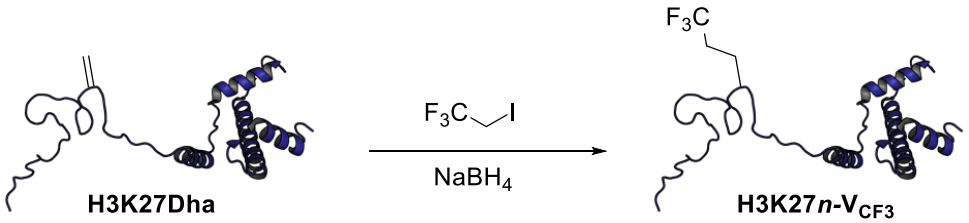
ESI-MS for the modified histone is shown below:

Calculated mass: 15223.8

Observed mass: 15226



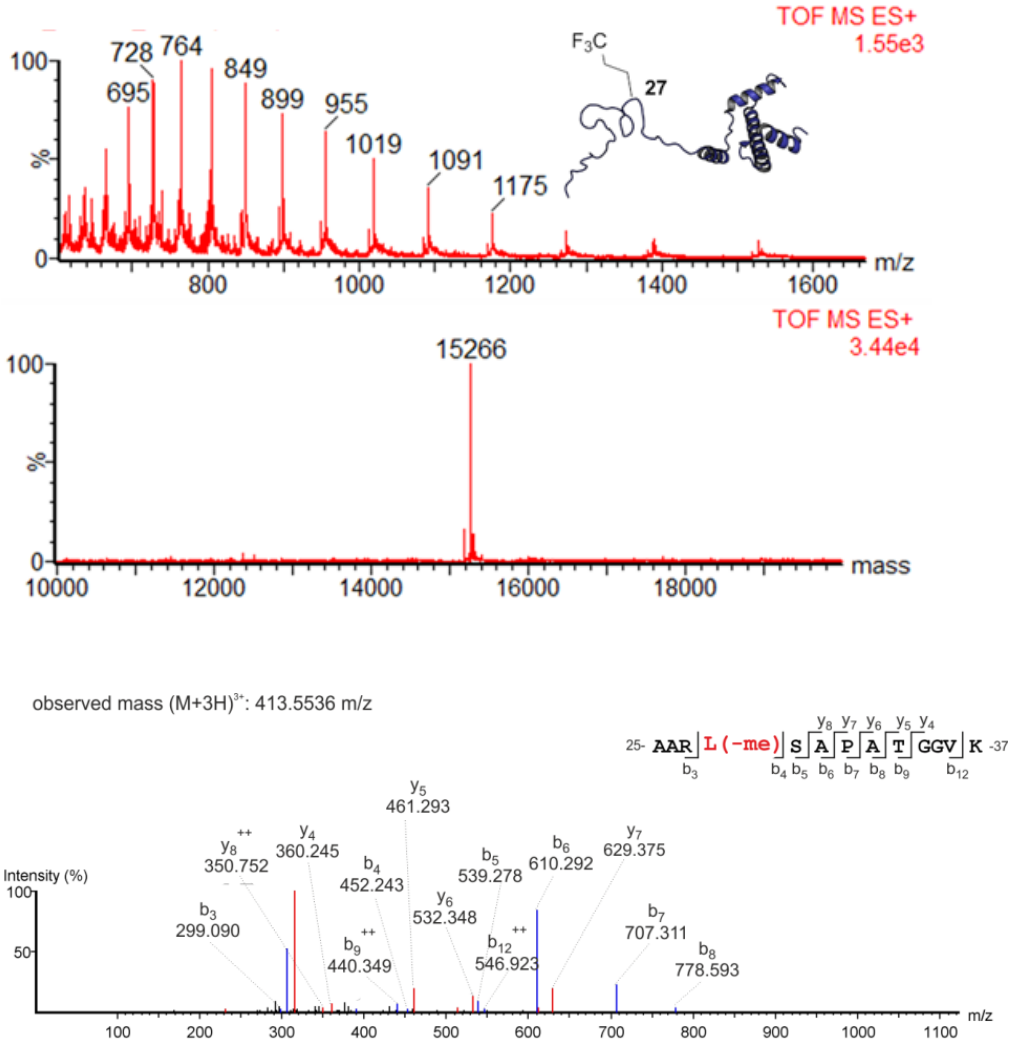
Installation of CF_3-norvaline at site H3 27



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

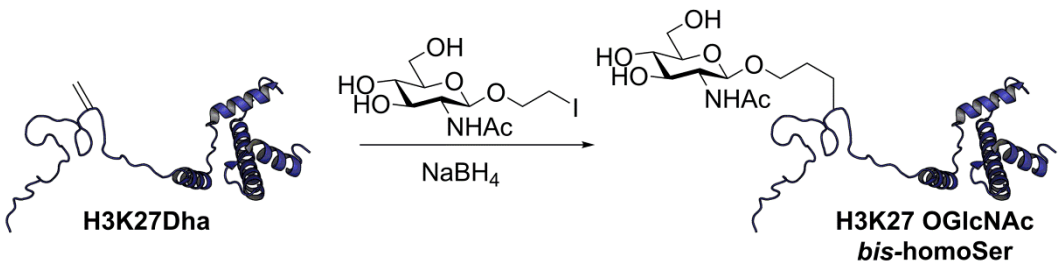
The following morning, within the glovebox, the degassed protein solution was combined with 6 µL trifluoro-2-iodoethane. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass: 15263.7
Observed mass: 15266



MS/MS spectrum of the peptide $_{25}$**AARL(-me)CF$_3$SAPATGGVK$_{37}$**

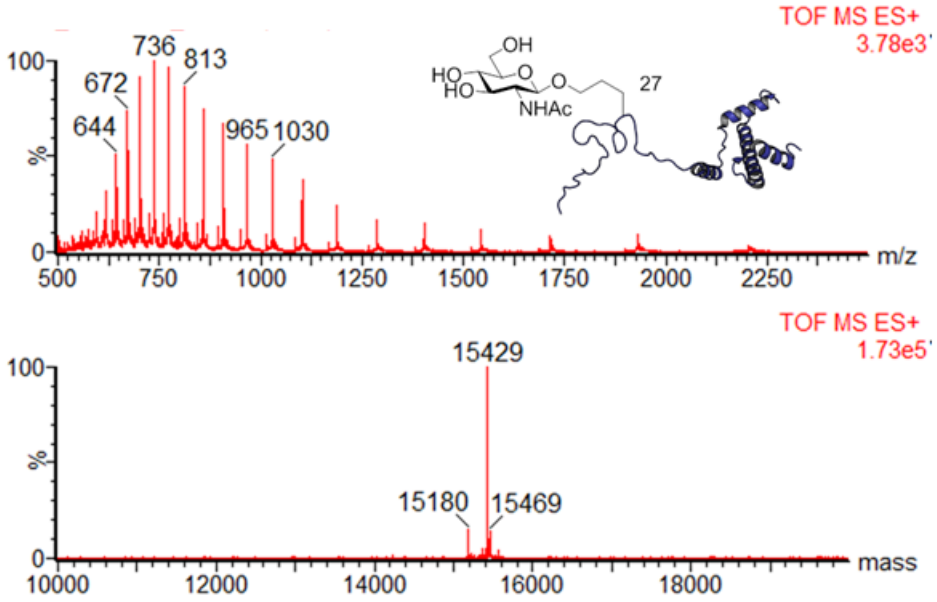
Installation of O-GlcNAc *bis*-homoserine at site H3 27



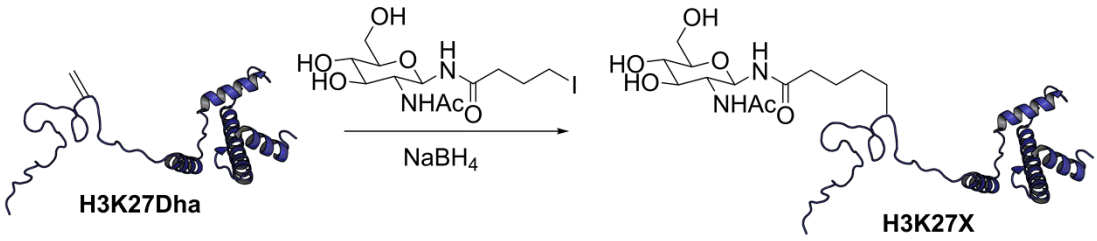
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was used to dissolve 2.5 mg 1-iodoethyl-β-D-GlcNAc **S19**. The resultant pale solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass: 15430
Observed mass: 15429



Installation of extended *N*-GlcNAc at site H3 27



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

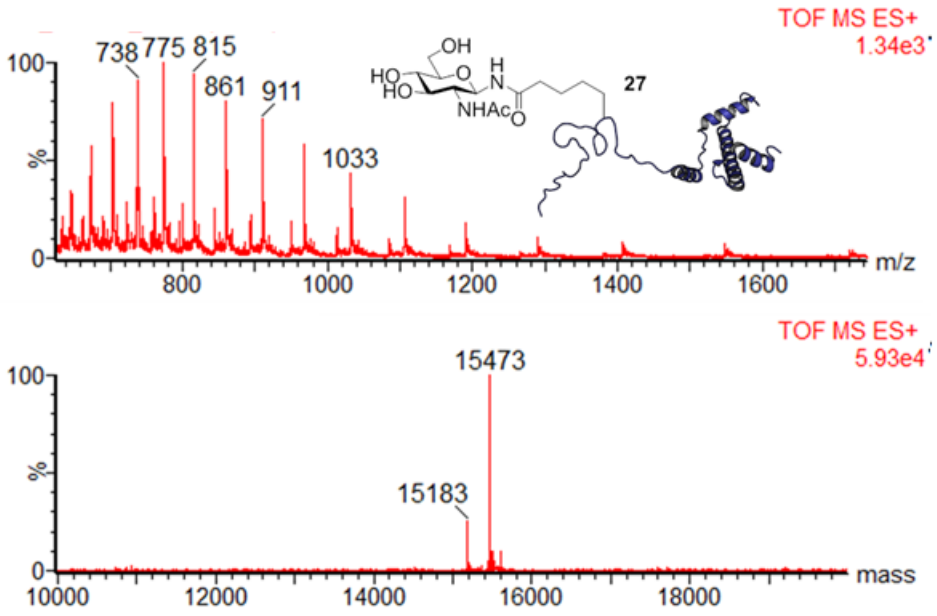
The following morning, within the glovebox, the degassed protein solution was used to form a suspension of 10mg 1-(4-iodobutyramido)-β-D-GlcNAc **S24**. The resultant cloudy solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing

effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

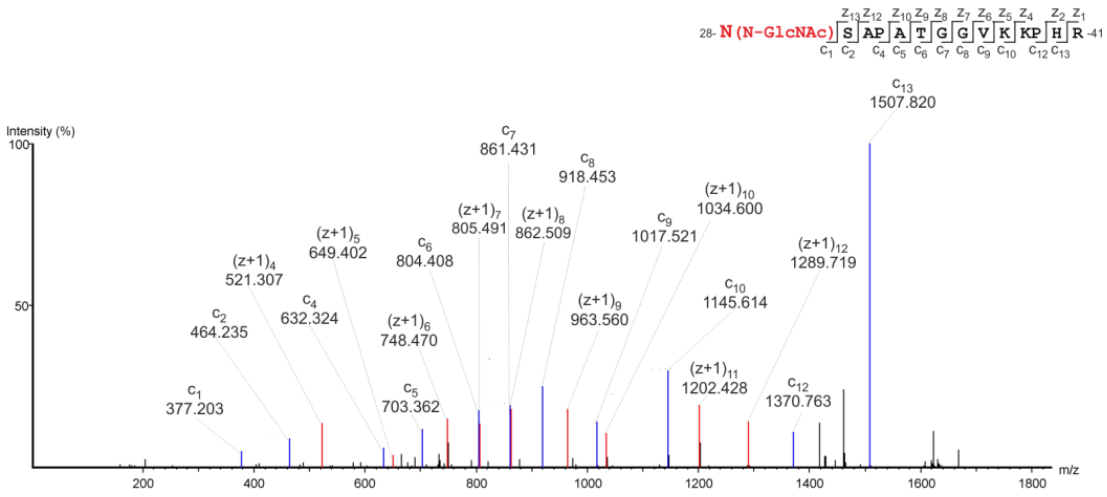
ESI-MS for the modified histone is shown below:

Calculated mass: 15471

Observed mass: 15473



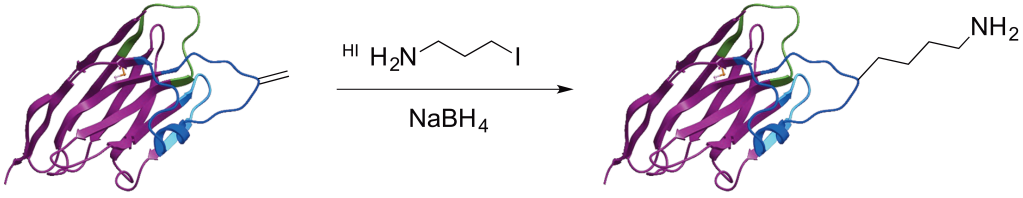
observed mass $(M+4H)^{4+}$: 416.9829 m/z



MS/MS spectrum of the peptide $_{28}$**N(GlcNAc)SAPATGGVKKPHR**$_{41}$

Chemical mutagenesis from cAb-Lys3-A104Dha

Installation of lysine from cAb-Lys3-A104Dha



cAbLys A104Dha **cAbLys A104K**

A 200 µL aliquot of cAb-Lys3-A104Dha (0.5 mg/mL solution in PBS pH 8.0) was thawed and transferred to a N_2 atmosphere glovebox and incubated for 12 hr at 4°C (overnight).

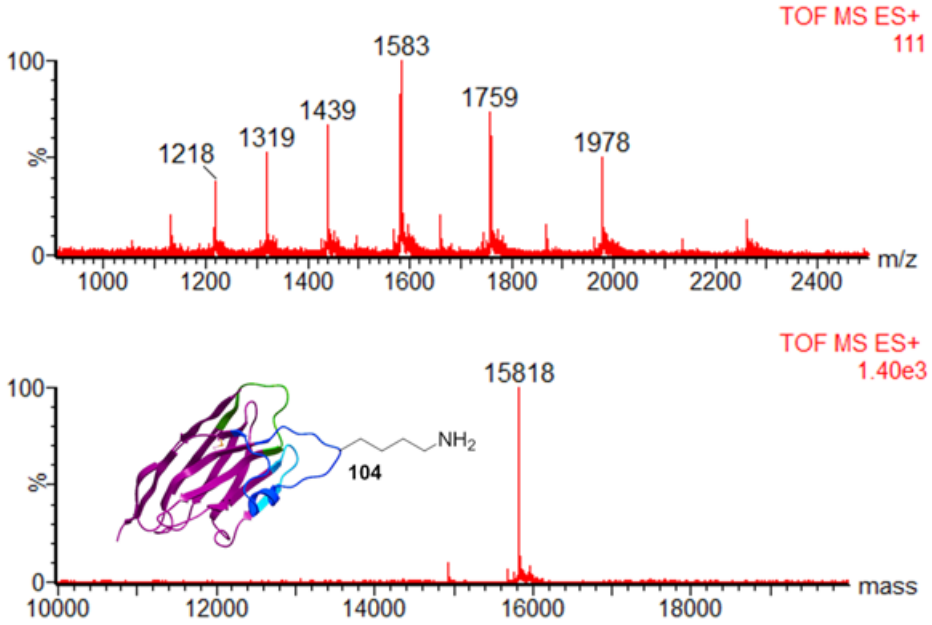
The following morning, within the glovebox, the degassed protein solution was used to dissolve 4 mg solid (3-iodopropyl)amine hydroiodide S1. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate eppendorf tube, kept at 4°C and containing approximately 0.25 mg powdered $NaBH_4$.

The reaction was gently shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

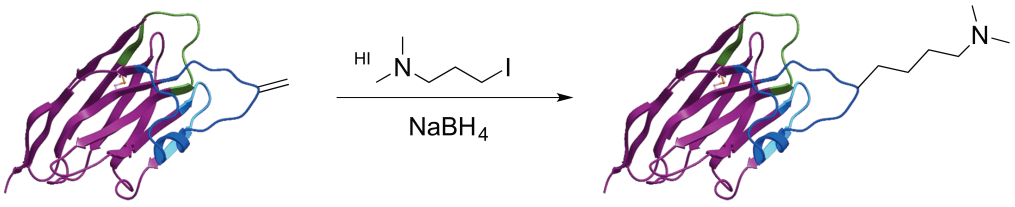
ESI-MS for the modified nanobody is shown below.

Calculated mass: 15820

Observed mass: 15818



Installation of dimethyllysine from cAb-Lys3-A104Dha



cAbLys A104Dha **cAbLys A104Kme$_2$**

A 200 µL aliquot of cAb-Lys3-A104Dha (0.5 mg/mL solution in PBS pH 8.0) was thawed and transferred to a N_2 atmosphere glovebox and incubated for 12 hr at 4°C (overnight).

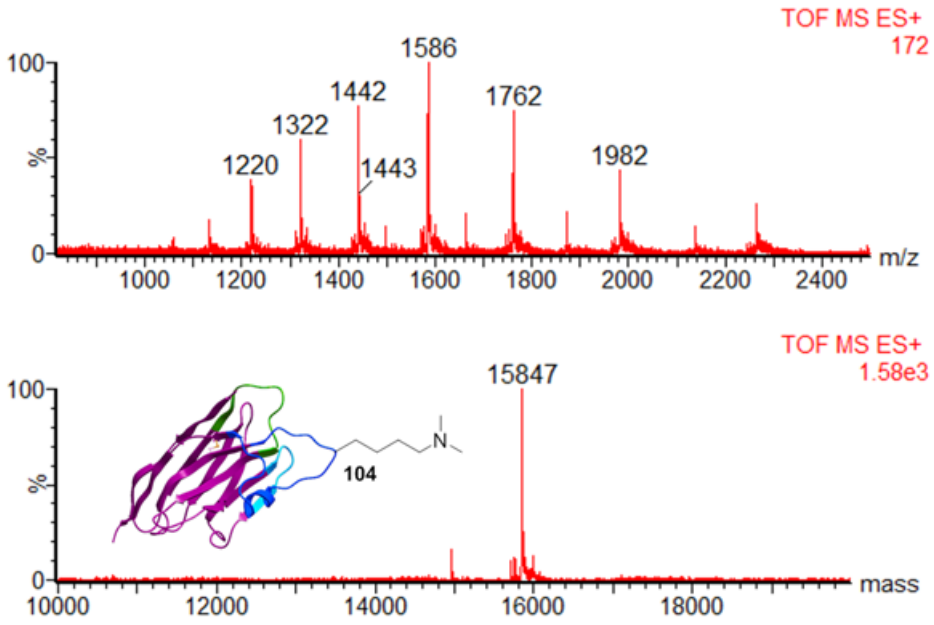
The following morning, within the glovebox, the degassed protein solution was used to dissolve 6 mg of (3-iodopropyl)dimethylamine hydroiodide **S5**. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate eppendorf tube, kept at 4°C and containing approximately 0.25 mg powdered $NaBH_4$.

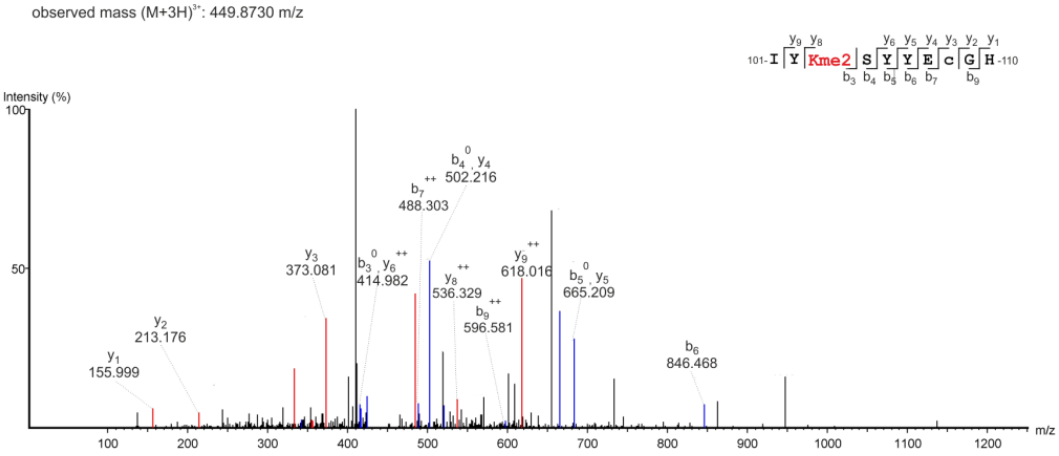
The reaction was gently shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified nanobody is shown below.

Calculated mass: 15849

Observed mass: 15847

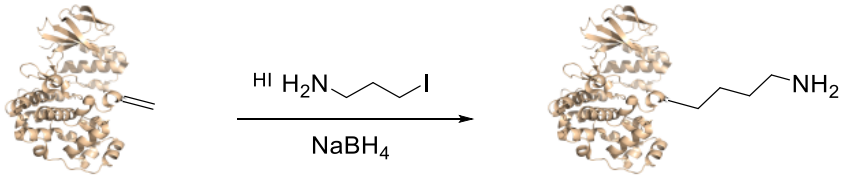




MS/MS spectrum of the peptide $_{101}$IYKme$_2$SYYECGH$_{110}$

Chemical mutagenesis from p38α kinase 172Dha

Installation of lysine from p38α kinase 172Dha



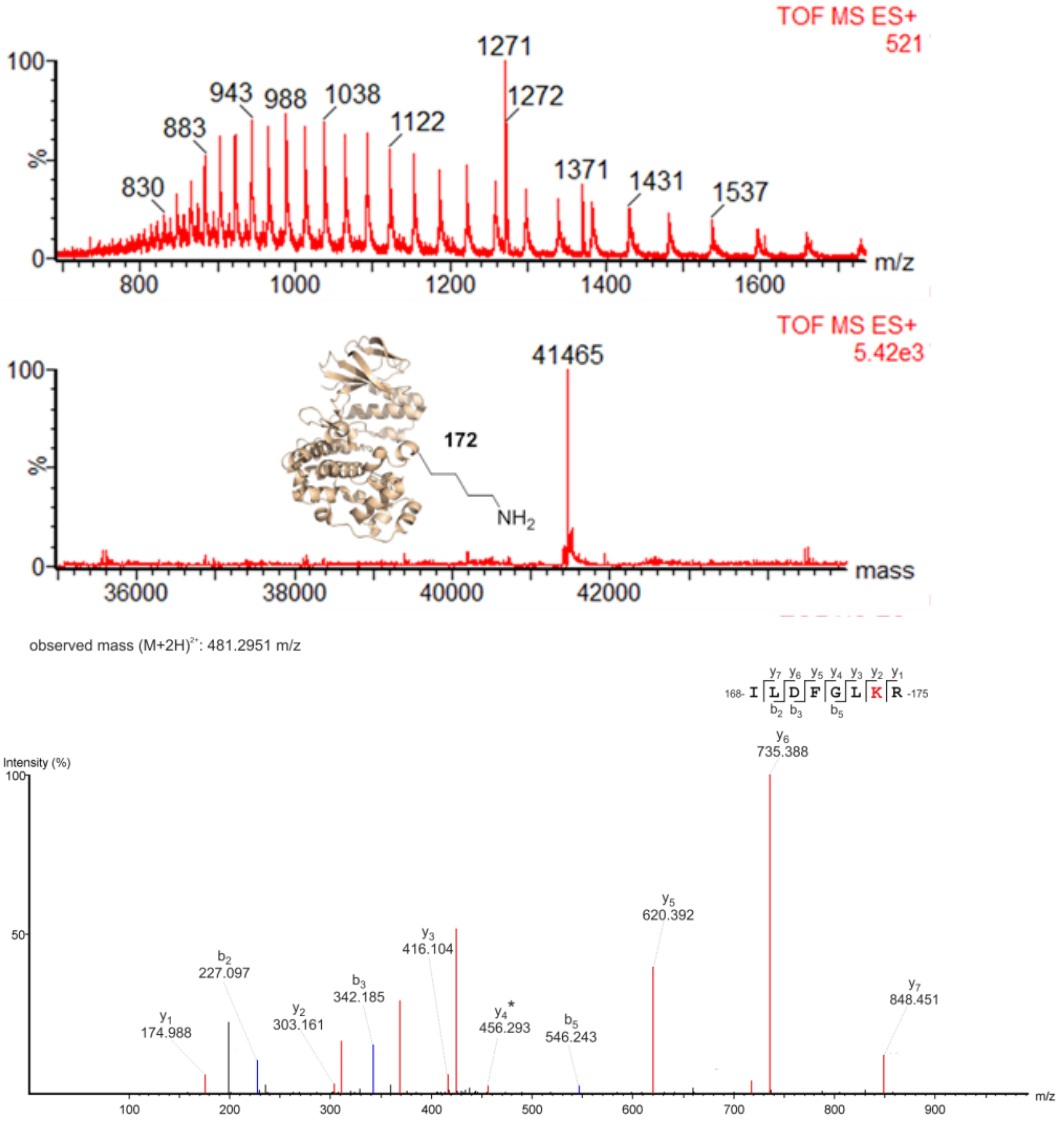
p38A172Dha **p38A172K**

A 75 µL aliquot of p38α-172Dha (0.67 mg/mL solution in 50mM HEPES pH 7.5) was thawed and transferred to a N_2 atmosphere glovebox and incubated for 12 hr at 4°C (overnight).

The following morning, within the glovebox, the degassed protein solution was used to dissolve 1 mg solid (3-iodopropyl)amine hydroiodide **S1**. The resultant solution was gently mixed by pipette. The entire volume (~75 µL) of the protein solution was then rapidly added via cautious pipetting to a separate 500 µL eppendorf tube, kept at 4°C and containing approximately 0.1 mg powdered $NaBH_4$.

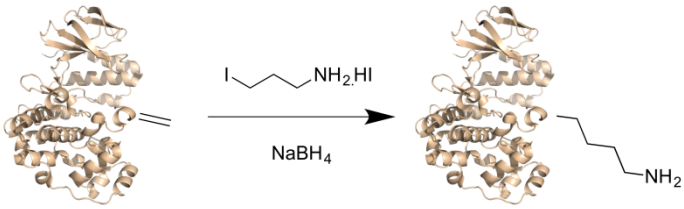
The reaction was gently shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified kinase is shown below:
Calculated mass: 41462 (p38α A172K)
Observed mass: 41465



MS/MS spectrum of the peptide $_{168}$ILDFGLKR$_{175}$

Chemical mutagenesis from p38α kinase 180Dha: Installation of lysine from p38α kinase 180Dha



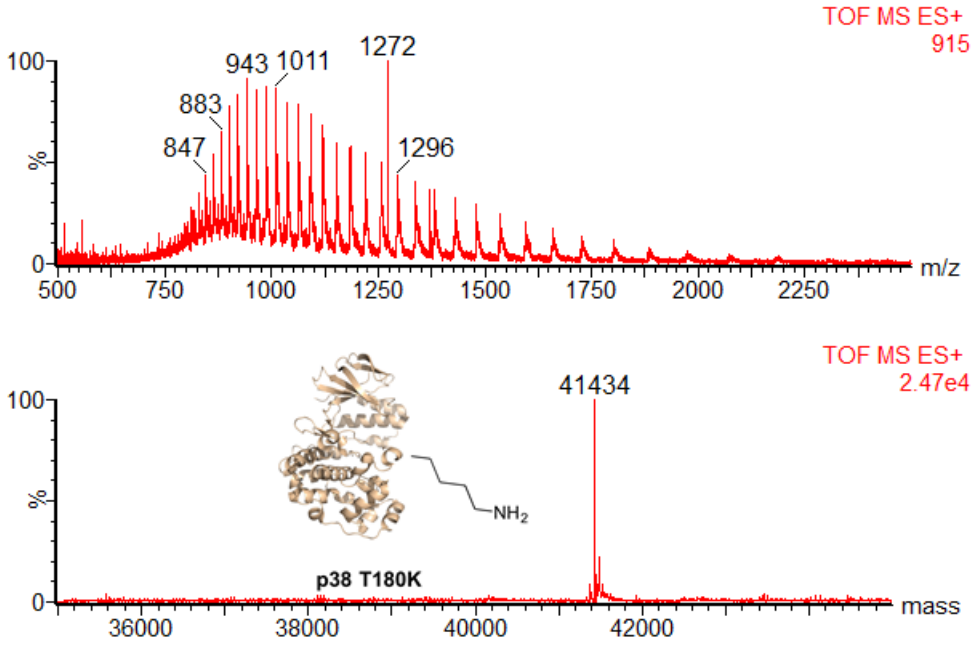
p38 T180Dha **p38 T180K**

A 100 µL aliquot of p38α-180Dha (0.85 mg/mL solution in 50mM HEPES pH 7.8) was thawed and transferred to a N_2 atmosphere glovebox and incubated for 12 hr at 4°C (overnight).

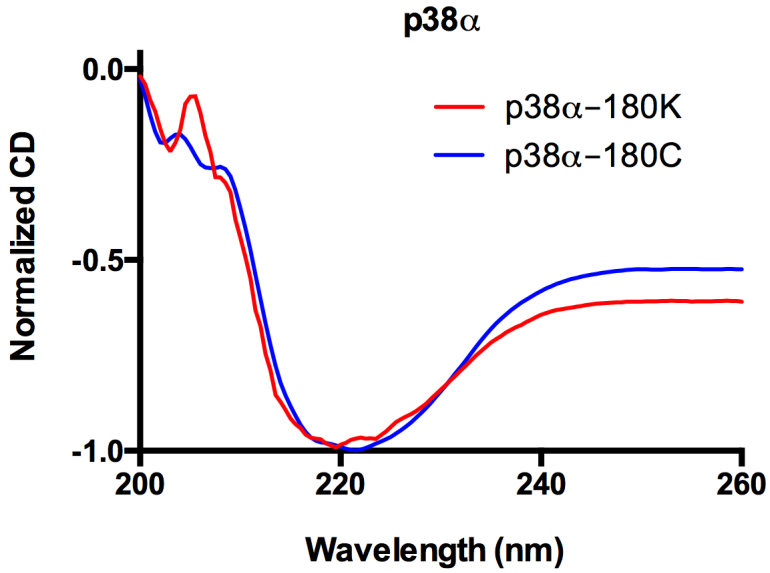
The following morning, within the glovebox, the degassed protein solution was used to dissolve 1 mg solid (3-iodopropyl)amine hydroiodide **S1**. The resultant solution was gently mixed by pipette. The entire volume (~100 µL) of the protein solution was then rapidly added via cautious pipetting to a separate 500 µL eppendorf tube, kept at 4°C and containing approximately 0.2 mg powdered $NaBH_4$.

The reaction was gently shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified kinase is shown below:
Calculated mass: 41,432 (p38α T180K)
Observed mass: 41,434



Circular dichroism spectroscopy reveals retention of overall secondary structure and folding during the course of chemical mutation p38α 180C →K



Chemoselectivity

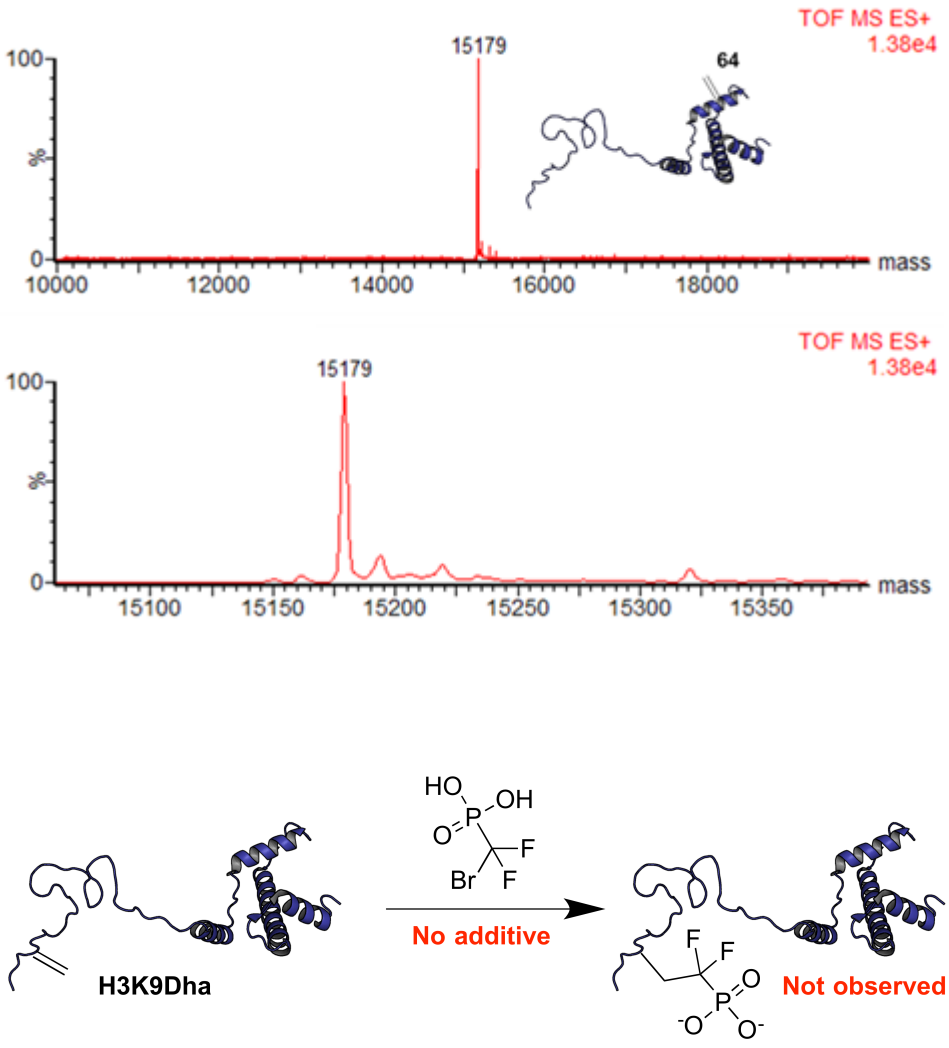
Chemical controls for reaction in absence of initiator/ radical mediator

Two differing halide substrates are shown here, all tested to date behave similarly.



A solution of histone H3 K64Dha (200 µL at 1 mg/mL, 500 mM NH_4OAc pH 6.0, 5M Gdn.HCl) was was incubated overnight at room temperature within a N_2 atmosphere glovebox. The following morning, within the glovebox, the protein solution was combined with 1.3 µL 2-iodopropane. The reaction was shaken at room temperature for 1 hr before being quenched by elution through a SpinTrap™ G-25 desalting column (same buffer) and analysed by LC-MS.

ESI-MS for the reaction mixture is shown below.
Calculated mass: 15180.7 (H3K64Dha, C110A)
Observed mass: 15179



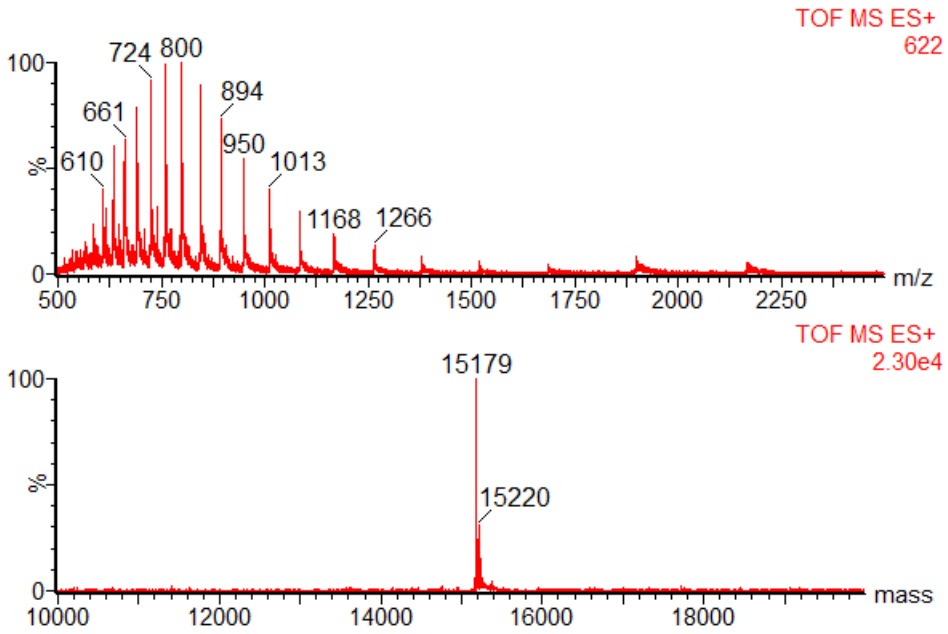
A solution of histone H3 K9Dha (200 µL at 1 mg/mL, 500 mM NH_4OAc pH 6.0, 5M Gdn.HCl) was incubated overnight at room temperature within a N_2 atmosphere glovebox. The following morning, within the glovebox, the protein solution was combined with 2 µL of a stock solution of (Bromodifluoromethyl)phosphonic acid S9 (prepared by dissolving 55 mg S9 in 275 µL water and adjusting the pH to 6.0). The reaction was incubated at room temperature for 30 minutes before being removed from the glovebox and analysed by LC-MS.

ESI-MS for the reaction mixture is shown below.

Calculated mass: 15180.7 (H3K64Dha, C110A)

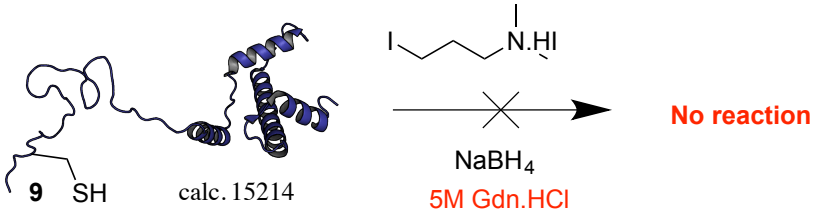
Observed mass: 15179

No peak is observed at 15179 +132Da, which is the expected alkylation product. The peak at +41 is likely a MeCN adduct observed on certain ESI instrument.



Chemical controls for reaction in absence of dehydroalanine plus presence of free cysteine

'Denaturing' protocol:



A reduced and desalted solution of H3K9C (200 µL, 1 mg/mL, ~66 µM), in 'high pH' buffer (as a stringent test for S_N2-type reactions, 100 mM sodium phosphate pH 8.0, 5M Gdn.HCl), was ported into a N_2 atmosphere glovebox and left overnight at room temperature.

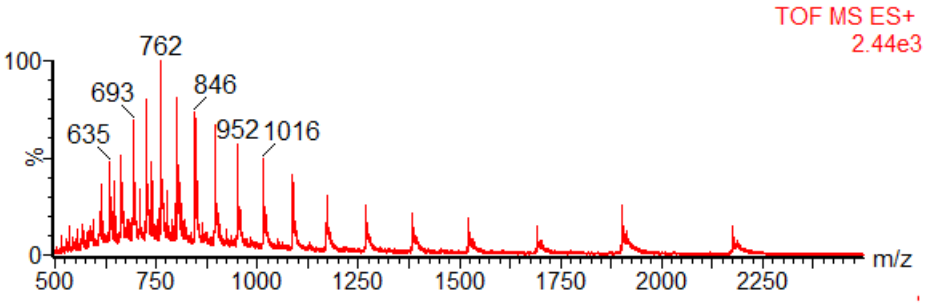
The following morning, within the glovebox, 4 mg powdered (3-iodopropyl)dimethylamine hydroiodide **S5** was dissolved in the degassed protein solution and the resultant pale yellow solution gently mixed by pipette.

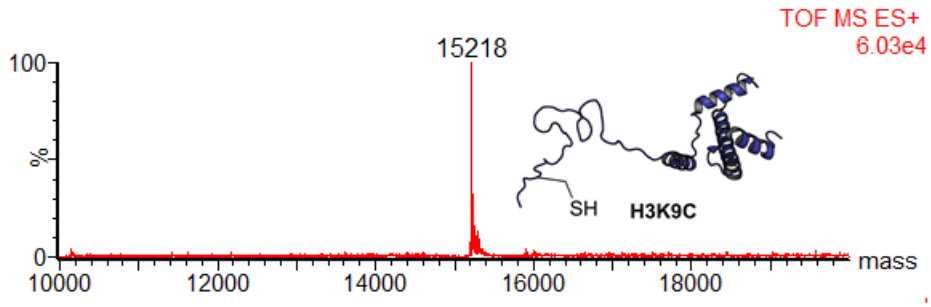
The entire volume (~200 µL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was incubated at 4°C for 30 minutes before being removed from the glovebox and directly analysed by LC-MS.

ESI-MS for the reaction mixture is shown below.

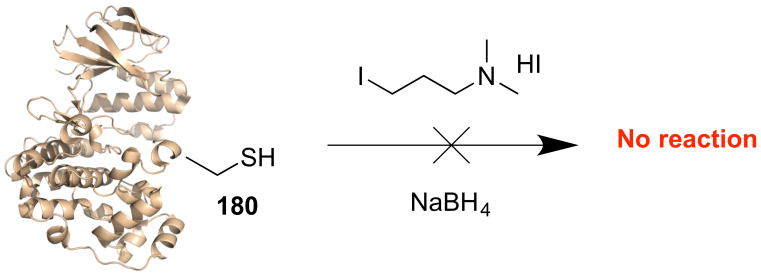
Calculated mass: 15214 (H3K9C C110A) 15300 (+1 alkylation)

Observed mass: 15218





'Native' protocol:



A 200 µL aliquot of reduced p38α-180C (1 mg/mL solution in 50mM HEPES pH 7.8) was thawed and transferred to a N_2 atmosphere glovebox and incubated for 12 hr at 4°C (overnight).

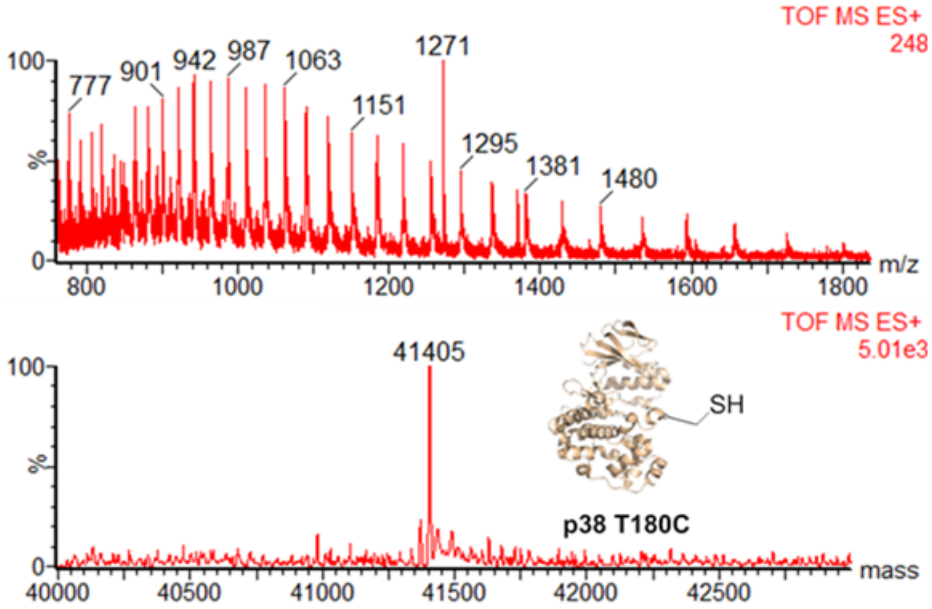
The following morning, within the glovebox, the degassed protein solution was used to dissolve 1.6 mg (2000 molar equiv. relative to protein) solid (3-iodopropyl)dimethylamine hydroiodide **S5**. The resultant solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via cautious pipetting to a separate 2 mL eppendorf tube, kept at 4°C and containing approximately 0.2 mg powdered $NaBH_4$ (1000 molar equiv. relative to protein).

The reaction was gently shaken at 4°C for 30 minutes before being removed from the glovebox and analysed directly by LC-MS.

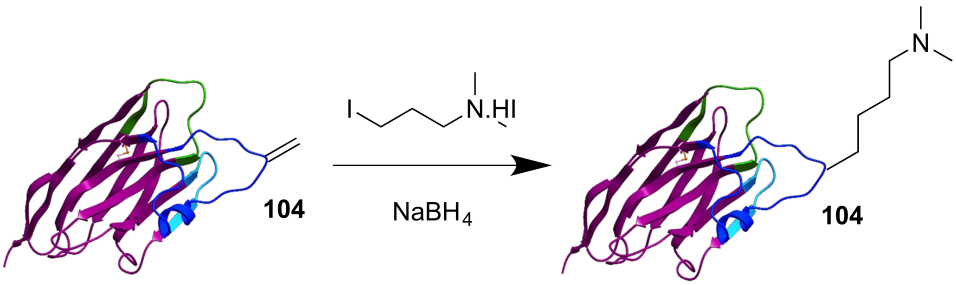
ESI-MS for the reaction mixture is shown below.

Calculated mass (p38α T180C): 41407

Observed mass: 41405

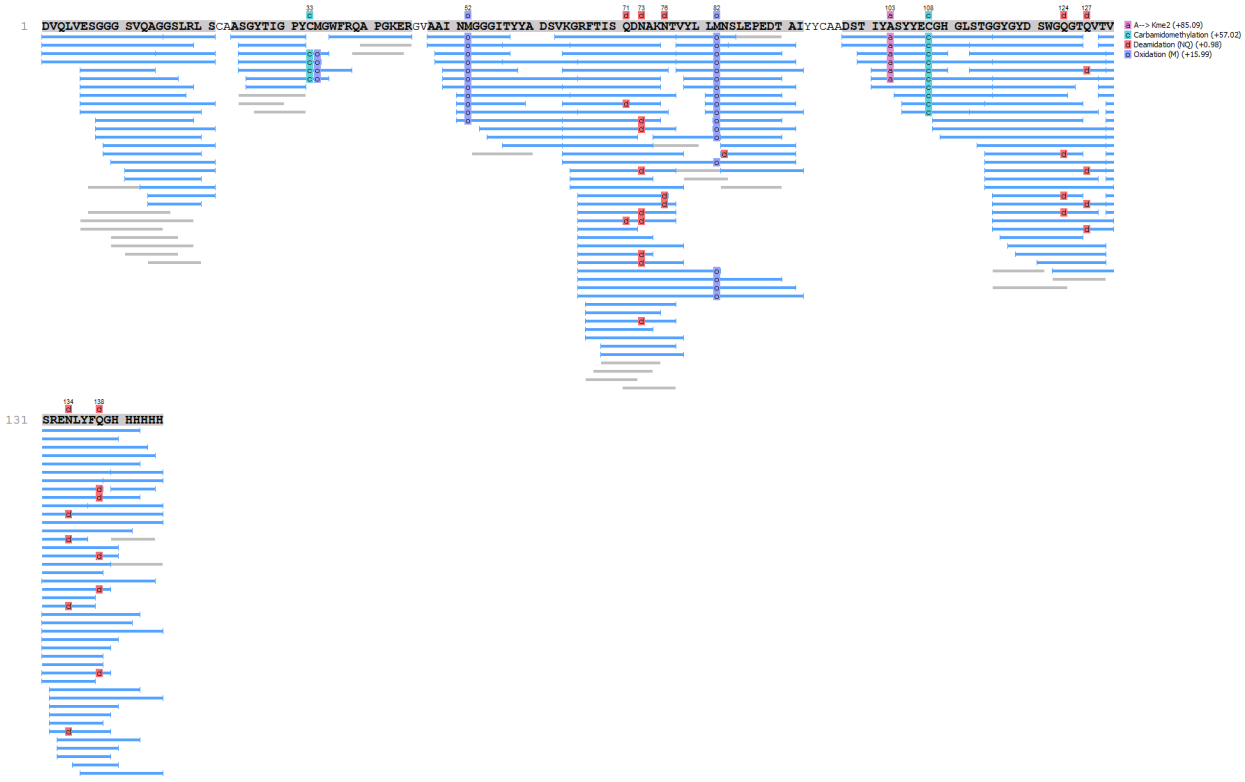

TOF MS ES+
248
100
%
0
777
901
942
987
1063
1151
1271
1295
1381
1480
800
1000
1200
1400
1600
1800
m/z
TOF MS ES+
5.01e3
100
%
0
41405
SH
p38 T180C
40000
40500
41000
41500
42000
42500
mass

cAbLys chemoselectivity

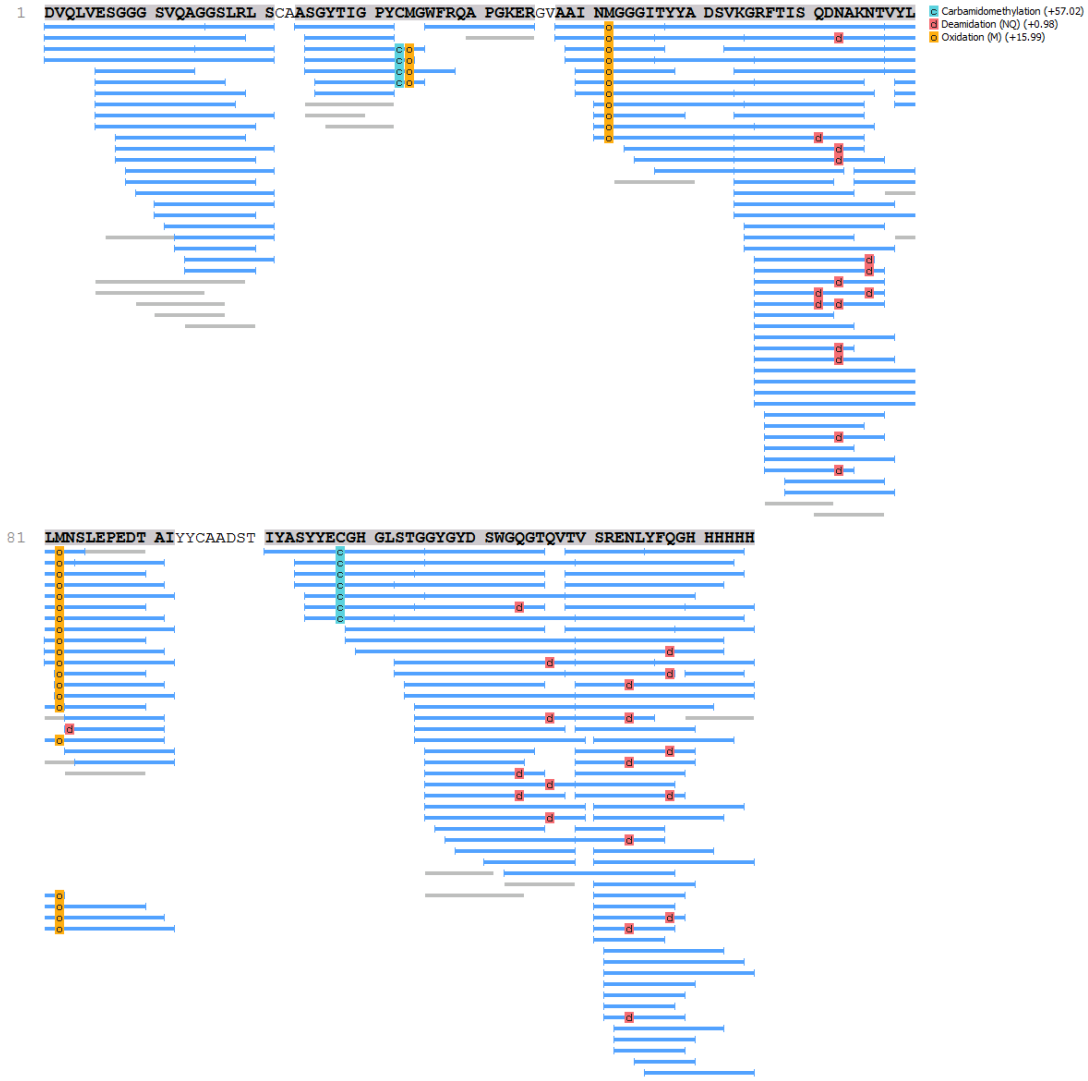


1 surface Dha (shown)
4 Cys 'buried' within disulfides

Complete selectivity observed by LC/ MS and MS^2

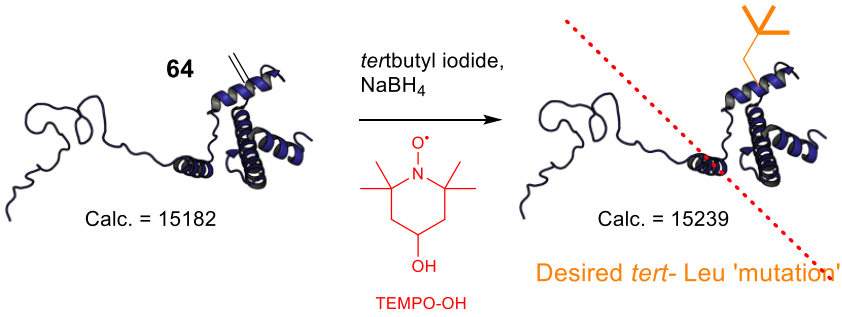


Searching for modification at nucleophilic residues (+86.16 Da):



Evidence for a radical mechanism

Reaction is inhibited by TEMPO-OH



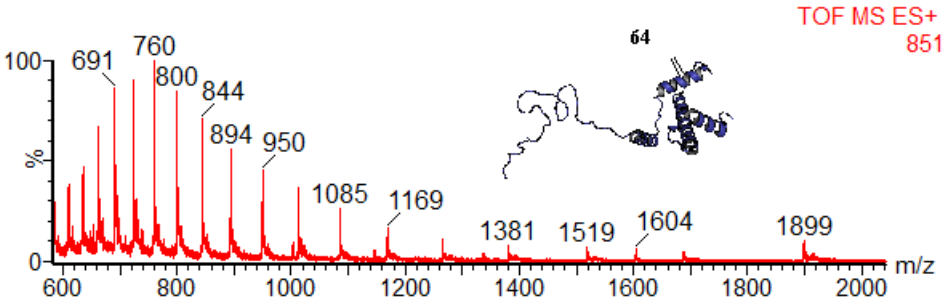
Following overnight incubation in an N_2 atmosphere glovebox to equilibrate to anaerobic conditions, a solution of histone H3 K64Dha (1 mg/mL protein, 500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0) was combined with 3.2 µL *tert*-butyl iodide (2000 molar equiv. relative to protein) and 1 mg of sodium borohydride (2000 molar equiv. relative to protein). In addition, 4-hydroxy TEMPO (20 µL of a stock prepared by dissolving 1.25 mg in 100 µL of reaction buffer, 0.25 mg total, 100 equiv.) was added. Notably, no effervescence or bubbling was observed.

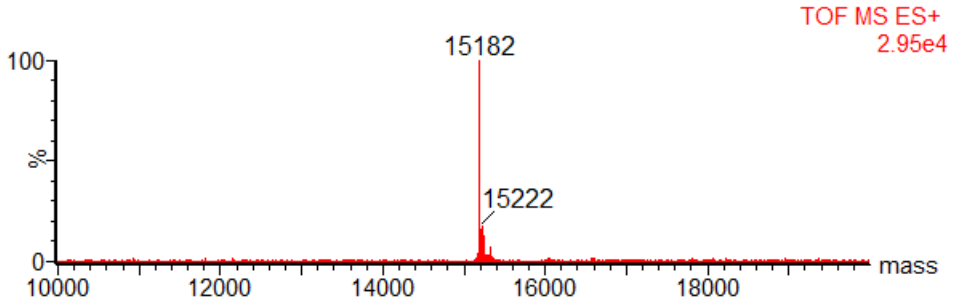
The reaction was incubated, still within the glovebox, at room temperature for two hours before being quenched by elution through SpinTrap™ G-25 desalting column (according to the manufacturers instructions) and analysed by LC-MS, revealing only starting material. *Identical results were observed for experiments with other initiators (2000 molar equiv. of Zn, In, or $InCl_3$ + $NaBH_4$).*

ESI-MS for the experiment is shown below:

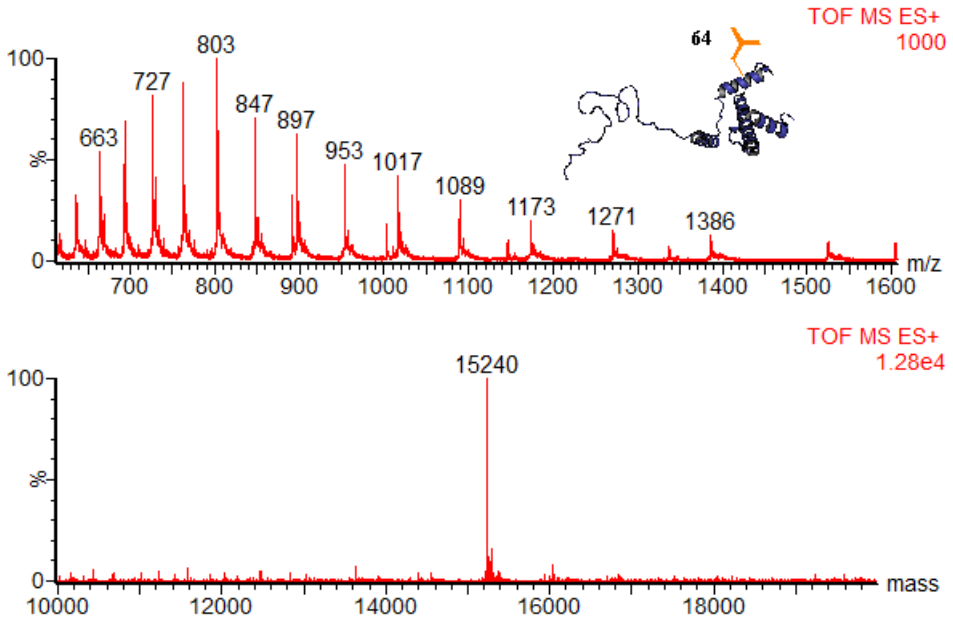
Calculated mass: 15238.82 (H3 $K64_{tBu}A$, C110A)

Observed mass: 15182 (H3 K64Dha, C110A)





Control reaction performed in parallel to that described above, identical except without addition of TEMPO-OH:



Calculated mass: 15238.82 (H3 K64$_{tBu}$A, C110A)

Observed mass: 15240 (H3 K64 $_{tBu}$A, C110A)

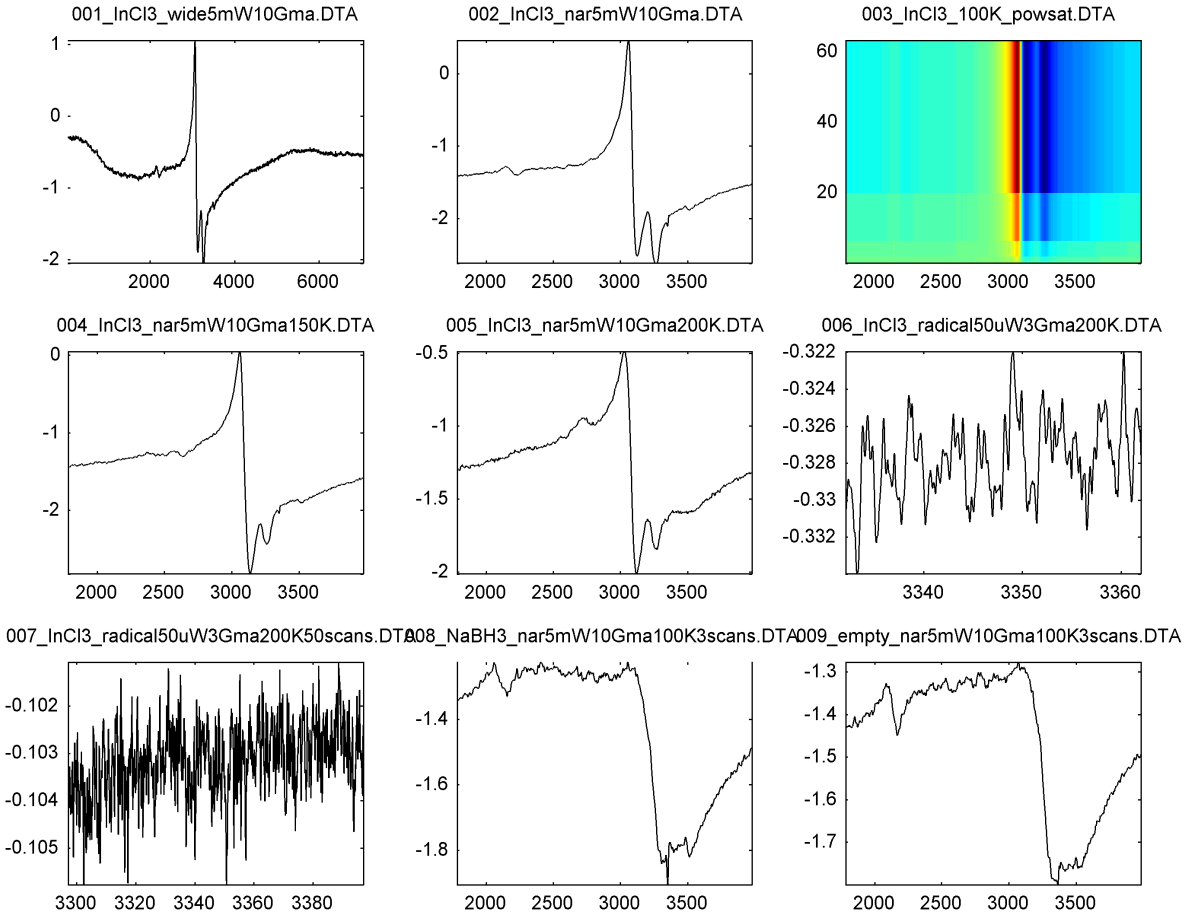
EPR Spectroscopy

Performed by Benjamin Bower and Will Myers

EPR spectra were recorded on a Bruker EMX spectrometer in the X-band (9.1-9.9 GHz), at the stated temperature (N_2 (l) cooled) with a field range between 0.0–0.6 T. EPR spectra were recorded with continuous wave spectroscopy. Data was processed using MATLAB (Mathworks) and spin simulations run using scripts from Easyspin[58]. The spectrometer was manually tuned before each spectrum was recorded. Organoiodides used for EPR spectroscopy were distilled prior to use. Samples were prepared in an N_2 purged glovebox (typically 2.0-3.0 ppm O_2) unless otherwise stated.

General Procedure for EPR spectroscopy Sample Preparation

A J. Youngs tube was charged with radical initiator and loaded into a glovebox. After a few hours of equilibration, distilled and degassed 2-iodopropane (100 μL) and NH_4OAc buffer (500 mM, pH 6.0, 500 μL) were carefully balanced at the top of the tube. The tubes were sealed and removed from the glovebox. With a flick of the wrist the liquids were sent to the bottom of the tube, the samples were immediately placed into liquid nitrogen and stored there until measurement.



001 Wide frequency scan of $InCl_3/NaBH_4$ system, 10 G 100K

002 Narrow frequency scan of $InCl_3/NaBH_4$ system 10 G 100K

003 Saturation cure of $InCl_3/NaBH_4$ system 10 G 100K

004 Narrow frequency scan of $InCl_3/NaBH_4$ system 10 G 150K

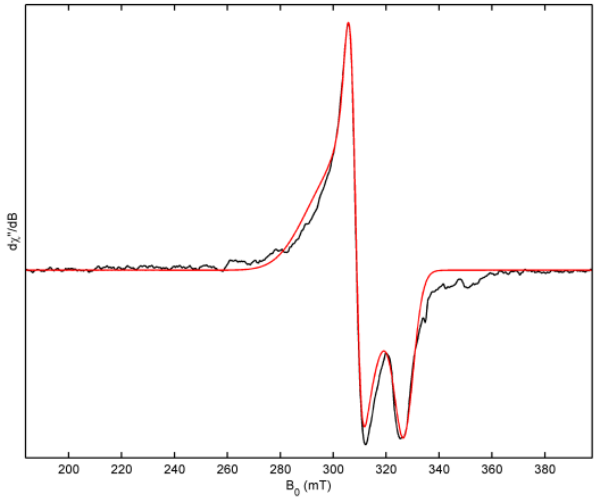
005 Narrow frequency scan of $InCl_3/NaBH_4$ system 10 G 200K

006 V. Narrow frequency scan of $InCl_3/NaBH_4$ system 10 G 200K

007 V. Narrow frequency scan of $InCl_3/NaBH_4$ system 10 G 200K 50 scan average

008 Narrow frequency scan of $NaBH_4$ system 10 G 100K

009 Blank Sample 10 G 100 K



Continuous wave electron paramagnetic resonance spectrum of $InCl_3/NaBH_4$ mixed with 2-iodopropane in 500 mM NH_4OAc buffer to generate an organoradical. The spectrum was recorded on a Bruker EMX at a temperature of 100 K, 5 mW microwave power and 10 G modulation amplitude under non saturated conditions. The red line indicates a spin simulation performed using the EasySpin software package for MatLab.

NaBH4-mediated reduction of dehydroalanine to alanine in the absence of an iodo-compound

Procedure

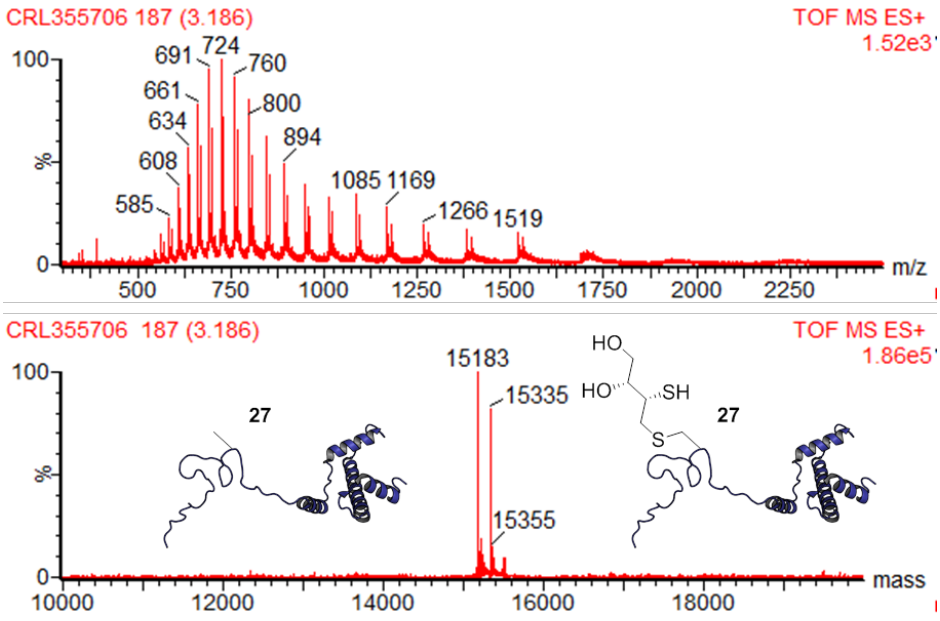
Reaction buffer (500 mM NH4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to three separate eppendorf tubes (representing three time points), followed by 20 µL of 10 mg/mL histone H3K27Dha protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solutions were then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions for 6 hours (such a period can be reduced if the protein solution is sparged with argon before entry).

The protein solution was then carefully pipetted into a separate eppendorf tube, kept at 4 °C and containing solid NaBH4 (1000 molar equiv. relative to protein). Effervescence is observed. The reactions were gently shaken at 4 °C until the indicated time points, at which point the reaction solution was loaded onto a SpinTrap™ G-25 desalting column, pre-equilibrated with 50 mM phosphate pH 8.0 buffer, to quench the reaction and exchange the protein into basic buffer for thia-michael addition. The solution was removed from the glovebox and eluted according to the manufacturers instructions. Addition of five mg solid dithiothreitol was followed by incubation at room temperature for five minutes before analysis by LC-MS at the given time points:

t = 30 min

Calculated mass (H3K27A): 15181.7 (DTT adduct = 15336)

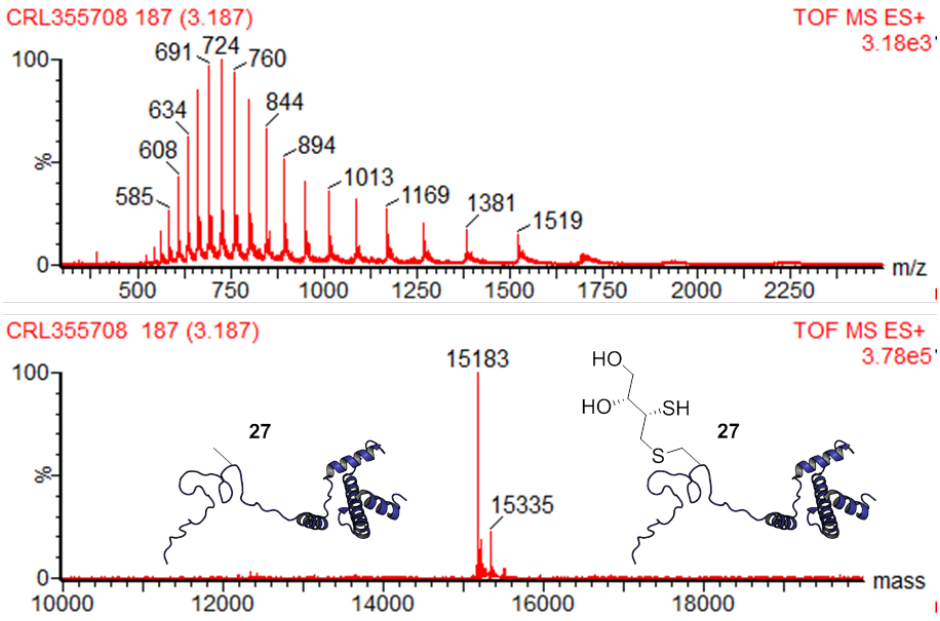
Observed mass: 15183



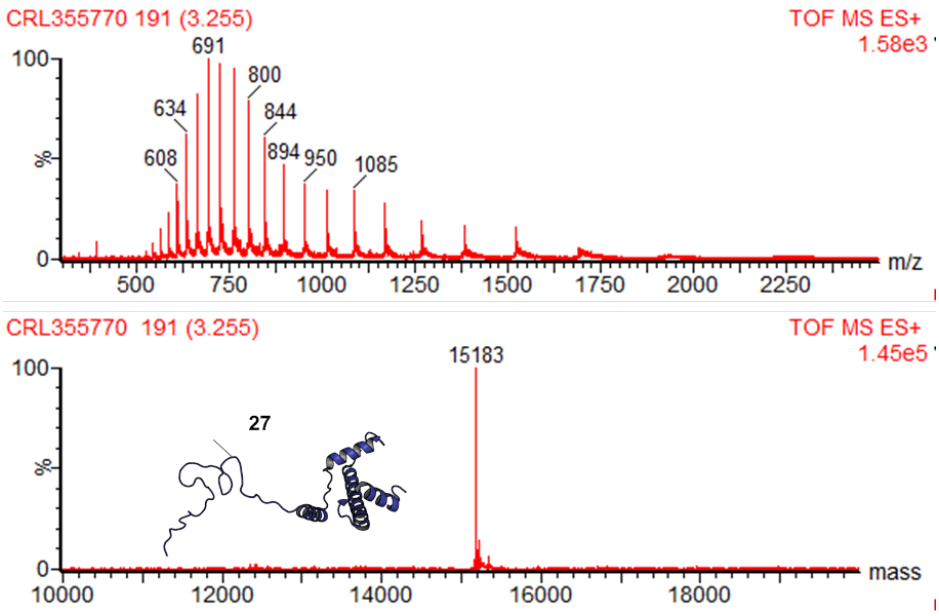
t = 1 hr

Calculated mass (H3K27A): 15181.7 (DTT adduct = 15336)

Observed mass: 15183



t= 4hr



Calculated mass (H3K27A): 15181.7 (DTT adduct = 15336)

Observed mass: 15183

References: Chapter 4

1 Giese, B. Syntheses with Radicals—C C Bond Formation via Organotin and Organomercury Compounds [New Synthetic Methods (52)]. *Angewandte Chemie International Edition in English* **24**, 553-565 (1985).

2 Studer, A. & Curran, D. P. Catalysis of radical reactions: A radical chemistry perspective. *Angewandte Chemie International Edition* **55**, 58-102 (2016).

3 Hart, D. J. Free-radical carbon-carbon bond formation in organic synthesis. *Science* **223**, 883-888 (1984).

4 Roberts, B. Polarity-reversal catalysis of hydrogen-atom abstraction reactions: concepts and applications in organic chemistry. *Chemical Society Reviews* **28**, 25-35 (1999).

5 Chatgilialoglu, C. & Griller, D. in *Organic Free Radicals* 25-26 (Springer, 1988).

6 Giese, B. & Kopping, B. Tris (trimethylsilyl) silane as mediator in organic synthesis via radicals. *Tetrahedron Letters* **30**, 681-684 (1989).

7 Ballestri, M. *et al.* Tris (trimethylsilyl) silane as a radical-based reducing agent in synthesis. *The Journal of Organic Chemistry* **56**, 678-683 (1991).

8 Graham, S. R., Murphy, J. A. & Kennedy, A. R. Hypophosphite mediated carbon–carbon bond formation: total synthesis of epialboatrin and structural revision of alboatrin. *Journal of the Chemical Society, Perkin Transactions 1*, 3071-3073 (1999).

9 Martin, C. G., Murphy, J. A. & Smith, C. R. Replacing tin in radical chemistry: N-ethylpiperidine hypophosphite in cyclisation reactions of aryl radicals. *Tetrahedron Letters* **41**, 1833-1836 (2000).

10 Jang, D. O. & Cho, D. H. N-Ethylpiperidine hypophosphite mediated intermolecular radical carbon-carbon bond forming reaction in water. *Synlett* **2002**, 1523-1525 (2002).

11 Birnbaum, E. R. & Javora, P. H. The reaction of orgatin chlorides with sodium borohydride: a new preparation of organotin hydrides. *Journal of Organometallic Chemistry* **9**, 379-382 (1967).

12 Corey, E. & Suggs, J. W. Method for catalytic dehalogenations via trialkyltin hydrides. *The Journal of Organic Chemistry* **40**, 2554-2555 (1975).

13 Hays, D. S. & Fu, G. C. Organotin Hydride Catalyzed Carbon− Carbon Bond Formation: Radical-Mediated Reductive Cyclization of Enals and Enones. *The Journal of Organic Chemistry* **61**, 4-5 (1996).

14 Terstiege, I. & Maleczka, R. E. A New approach for the generation and reaction of organotin hydrides: the development of reactions catalytic in tin. *The Journal of Organic Chemistry* **64**, 342-343 (1999).

15 Inoue, K., Sawada, A., Shibata, I. & Baba, A. Indium (III) Chloride− Sodium Borohydride System: A Convenient Radical Reagent for an Alternative to Tributyltin Hydride System. *Journal of the American Chemical Society* **124**, 906-907 (2002).

16 Zurita-Lopez, C. I., Sandberg, T., Kelly, R. & Clarke, S. G. Human protein arginine methyltransferase 7 (PRMT7) is a type III enzyme forming ω-NG-monomethylated arginine residues. *Journal of Biological Chemistry* **287**, 7859-7870 (2012).

17 Hodgins, D. S. Yeast phenylalanine ammonia-lyase purification, properties, and the identification of catalytically essential dehydroalanine. *Journal of Biological Chemistry* **246**, 2977-2985 (1971).

18 Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science* **2**, 1666-1676 (2011).

19 Groves, J. T. & Ma, K. W. Dehalogenations with sodium borohydride. Evidence for a free radical reaction. *Journal of the American Chemical Society* **96**, 6527-6529 (1974).

20 Ashby, E., Wenderoth, B., Pham, T. N. & Park, W. S. Evidence for single electron transfer in the reduction of organic halides by lithium triethylborohydride. *The Journal of Organic Chemistry* **49**, 4505-4509 (1984).

21 Abeywickrema, A. N. & Beckwith, A. L. Homolytic reductive dehalogenation of aryl halides by sodium borohydride. *Tetrahedron letters* **27**, 109-112 (1986).

22 Ryu, I., Uehara, S., Hirao, H. & Fukuyama, T. Tin-Free Giese Reaction and the Related Radical Carbonylation Using Alkyl Iodides and Cyanoborohydrides. *Organic Letters* **10**, 1005-1008, doi:10.1021/ol7031043 (2008).

23 Kawamoto, T. & Ryu, I. Radical reactions of borohydrides. *Organic & biomolecular chemistry* **12**, 9733-9742 (2014).

24 Kawamoto, T. *et al.* Borohydride-mediated radical addition reactions of organic iodides to electron-deficient alkenes. *The Journal of organic chemistry* **79**, 3999-4007 (2014).

25 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).

26 Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry* **17**, 2337-2342 (2003).

27 Chalker, J. *Reaction Engineering for Protein Modification: Tools for Chemistry and Biology.* DPhil Thesis thesis, University of Oxford, (2011).

28 Tessarz, P. *et al.* Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**, 564-568 (2014).

29 Duewel, H., Daub, E., Robinson, V. & Honek, J. F. Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein 19F NMR. *Biochemistry* **36**, 3404-3416 (1997).

30 Li, C. *et al.* Protein 19F NMR in Escherichia coli. *Journal of the American Chemical Society* **132**, 321 (2010).

31 Tkachenko, A. N. *et al.* A 19F NMR Label to Substitute Polar Amino Acids in Peptides: A CF3‐Substituted Analogue of Serine and Threonine. *Angewandte Chemie* **125**, 1526-1529 (2013).

32 Marsh, E. N. G. & Suzuki, Y. Using 19F NMR to probe biological interactions of proteins and peptides. *ACS chemical biology* **9**, 1242-1250 (2014).

33 Ye, L., Larda, S. T., Li, Y. F. F., Manglik, A. & Prosser, R. S. A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in 19F NMR studies of proteins. *Journal of biomolecular NMR* **62**, 97-103 (2015).

34 Rydzik, A. M. *et al.* Monitoring Conformational Changes in the NDM‐1 Metallo‐β‐lactamase by 19F NMR Spectroscopy. *Angewandte Chemie International Edition* **53**, 3129-3133 (2014).

35 Zheng, W. *et al.* Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *Journal of Biological Chemistry* **280**, 10462-10467 (2005).

36 Gamblin, D. P., Scanlan, E. M. & Davis, B. G. Glycoprotein synthesis: an update. *Chemical Reviews* **109**, 131-163 (2008).

37 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

38 Giese, B. Formation of CC bonds by addition of free radicals to alkenes. *Angewandte Chemie International Edition in English* **22**, 753-764 (1983).

39 Tedder, J. M. & Walton, J. C. The importance of polarity and steric effects in determining the rate and orientation of free radical addition to olefins: Rules for determining the rate and preferred orientation. *Tetrahedron* **36**, 701-707 (1980).

40 Clerici, A., Minisci, F. & Porta, O. Nucleophilic character of alkyl radicals—IX: The benzyl radical. *Tetrahedron* **29**, 2775-2779 (1973).

41 Walbiner, M., Wu, J. Q. & Fischer, H. Absolute Rate Constants for the Addition of Benzyl (PhĊH2) and Cumyl (PhĊMe2) Radicals to Alkenes in Solution. *Helvetica chimica acta* **78**, 910-924 (1995).

42 Zytowski, T. & Fischer, H. Absolute rate constants for the addition of methyl radicals to alkenes in solution: new evidence for polar interactions. *Journal of the American Chemical Society* **118**, 437-439 (1996).

43 Zytowski, T. & Fischer, H. Absolute Rate Constants and Arrhenius Parameters for the Addition of the Methyl Radical to Unsaturated Compounds: The Methyl Affinities Revisited. *Journal of the American Chemical Society* **119**, 12869-12878, doi:10.1021/ja973128y (1997).

44 Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. & Muyldermans, S. Selection and identification of single domain antibody

fragments from camel heavy ‐ chain antibodies. *FEBS letters* **414**, 521-526 (1997).

45 Gunnoo, S. B. *et al.* Creation of a gated antibody as a conditionally functional synthetic protein. *Nature communications* **5** (2014).

46 Chooi, K. P. *et al.* Synthetic phosphorylation of p38α recapitulates protein kinase activity. *Journal of the American Chemical Society* **136**, 1698 (2014).

47 Provencher, S. W. & Gloeckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37 (1981).

48 Yang, J. T., Wu, C.-S. C. & Martinez, H. M. [11] Calculation of protein conformation from circular dichroism. *Methods in enzymology* **130**, 208-269 (1986).

49 Dingley, A. J. *et al.* Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *Journal of biomolecular NMR* **6**, 321-328 (1995).

50 Wilkins, D. K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424-16431 (1999).

51 Kay, L. E. Field gradient techniques in NMR spectroscopy. *Current opinion in structural biology* **5**, 674-681 (1995).

52 Piotto, M., Saudek, V. & Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR* **2**, 661-665 (1992).

53 Wishart, D., Sykes, B. & Richards, F. Simple techniques for the quantification of protein secondary structure by 1H NMR spectroscopy. *FEBS letters* **293**, 72-80 (1991).

54 Cotts, R., Hoch, M., Sun, T. & Markert, J. Pulsed field gradient stimulated echo methods for improved NMR diffusion measurements in heterogeneous systems. *Journal of Magnetic Resonance (1969)* **83**, 252-266 (1989).

55 Stejskal, E. O. & Tanner, J. E. Spin diffusion measurements: spin echoes in the presence of a time‐dependent field gradient. *The journal of chemical physics* **42**, 288-292 (1965).

56 Johnson, G. L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298**, 1911-1912 (2002).

57 Chooi, K. P. *et al.* Synthetic phosphorylation of p38α recapitulates protein kinase activity. *Journal of the American Chemical Society* **136**, 1698-1701 (2014).

58 Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *Journal of Magnetic Resonance* **178**, 42-55 (2006).

Chapter 5: Post-translational mutagenesis for the investigation of histone methylation

Introduction

In the previous chapter, optimized conditions for 'chemical mutation' from a dehydroalanine 'tag' were reported. A range of halides could be successfully utilized as 'side-chain donor' reagents for chemical installation of natural residues, their PTMs and unnatural variants. The optimized borohydride-mediated conditions were shown to possess exquisite chemoselectivity for dehydroalanine and, furthermore, to have minimal effect on protein secondary structure and folding. Building on these results, it was of interest to investigate the biological *function* of proteins modified by such post-translational mutagenesis.

Protein methylation

The identification of ε-*N*-methyllysine in hydrolytic extracts of the bacterial flagellar protein in 1959[1] was the first evidence for the post-translational methylation of amino acid residues within whole proteins. Today, post-translational methylation of proteins is known to be a key regulator of cellular processes including transcription[2], RNA processing[3], translation and signal transduction[4]. In addition to lysine, other residues including arginine, glutamine, glutamate, aspartate, histidine, tyrosine, cysteine, serine and threonine, as well as the polypeptide N-terminus, are also known to undergo enzymatic methylation. The regulatory functions of protein methylation are underscored by the dynamic nature of the process[4,5]; both 'writer' enzymes that methylate proteins and 'eraser' enzymes with demethylase activity are known. Further, many 'reader' domains that can recognize and bind methylated residues within specific proteins have also been identified, often implicating a given methylation event in complex networks of protein interactions[6-8].

Histone methylation

One area in which protein methylation plays a key role is in the regulation of chromatin structure and function[8]. Chromatin acts as the physical template for DNA in eukaryotes, enabling the dynamic spatial organisation of the genome - a fact of key importance in the regulation of gene expression. The nucleosome is the fundamental building block of chromatin, consisting of DNA wrapped around histone protein octamers. Four different canonical histones are known (H2A, H2B, H3, H4), each consisting of a globular domain and an N-terminal tail. The N-termini of the histone proteins extend from the nucleosome core as highly dynamic, disordered 'tail' structures, which are known to play important roles in higher-level chromatin folding[9,10]. The packing of nucleosomal units into higher-level arrays has important consequences for transcription and other processes that involve protein-DNA interactions[11,12]

The histone proteins in chromatin are also subject to diverse post-translational modifications (PTM) that can modulate their dynamics and molecular interactions[13]. 'Writer', 'reader' and 'eraser' proteins interact with these histone 'marks' and establish chromatin as a dynamic signaling platform. There are at least two (interconnected) mechanisms by which histone PTMs exert their effects on down-stream signaling from chromatin[13]. Firstly, PTMs can directly modulate the biophysical properties of the nucleosome core or histone tails, resulting in localized or more global rearrangements. For example, acetylation of histone N-terminal 'tails' decreases the overall charge of the tail, which can result in a weakening of interactions between the tail and the nucleosomal DNA, leading to a more open chromatin structure. Secondly, PTMs can facilitate the recruitment of effector proteins that recognize the PTM through 'reader' domains. Many such PTM-reader domains are present in chromatin-interacting protein complexes, enabling the coupling of PTM-recognition to downstream effects such as transcriptional activation through effector recruitment.

Histone lysine methylation illuminates the complexities involved in the regulation of chromatin states. Many 'writer' enzymes, termed lysine methyltransferases (KMTs)[14], have been identified that catalyse the methylation of lysine residues in histone H3 and H4 proteins. KMTs occur in two classes – SET domain and non-SET domain containing methyltransferases[15], with the SET-containing KMTs being the larger class. Both classes of methyltransferase utilize S-adenosyl-L-methionine (SAM) as the methyl-donating cofactor. Critically, KMTs have been demonstrated to possess specificity for both the particular substrate lysine residue targeted and for the degree of methylation transferred. Thus, a given lysine residue may occur in an unmodified or mono-, di- or tri-methylated state. For example, lysine 9 of histone H3 can be trimethylated by KMT1 family members A, B, E and F[16]. KMT1A/B install trimethyllysine from a pre-existing monomethylated state (H3K9me$_1$). In contrast, KMT1C (G9a)[17], can only catalyse H3K9 dimethylation, with an enzymatic preference for the mono- to di-methylation event.

Histone lysine demethylases (KDMs, 'erasers')[5,18-20] and 'readers'[7,21-23] also possess specificity for particular lysines and methylation states. For example, H3K9me$_3$ has been demonstrated to recruit heterochromatin protein 1 (HP1)[21], which can enforce a repressive transcriptional environment by bridging adjacent nucleosomes, leading to chromatin condensation[24]. H3K9 trimethylation can be removed by lysine demethylase 4 (KDM4a, also known as JMJD2a), which is a 2-oxoglutarate (2OG) dependent lysine demethylase that catalyses the removal of two methyl groups, leaving H3K9me$_1$[25]. As a consequence of such enzymatic specificity, the functional effects of modification at alternate lysines, and the degree of modification of each lysine, can be divergent and even opposed. Thus, while H3K4 methylation is typically observed at active gene promoters, H3K9 and H3K27 methylation states are traditionally associated with gene repression[16].

To further add to the complexity, most histones isolated directly from cells bear not just one but multiple PTMs, implying extensive 'cross-talk' between modifications. In 2001 Jenuwein and Allis proposed the 'histone code' hypothesis[26], predicting that *combinations* of histone 'marks' program distinct biological responses. In recent years, mass spectrometry has revealed the co-localisation of numerous marks within the same nucleosome[27] while genomic-level studies have shown *correlation* between patterns of modifications and distinct transcriptional and regulatory states[28], providing circumstantial evidence for the histone code. Despite this progress, the underlying biochemical mechanisms of the 'histone code' are still largely unknown[29]. Much work in the field has to this point relied on the use of peptide model systems for assay development. In peptide-based approaches, the 'context' of the histone 'language', in terms of both the mere presence of DNA and nucleosomes, and the specific pattern of co-occurring marks, is often lost[29,30]. Mechanistic understanding of chromatin-related processes such as transcription, replication and repair requires biochemical studies at the level of histones, nucleosomes and chromatin fibres in order to unravel the complexities involved. This goal would be aided greatly by new techniques for the rapid (semi)-synthesis of histones bearing defined patterns of post-translational modification.

Chemical approaches to the synthesis of histones bearing defined methylation states

As described in detail in the introduction to this thesis, there are three main approaches for the production of proteins bearing distinct PTMs – protein total synthesis, expression via stop-codon suppression techniques and hybrid 'tag-and-modify' strategies. All of these alternate approaches have been used to produce histones bearing PTMs, and their specific application to the study of methylated histones will be briefly outlined here[30].

As relatively small proteins (<150 a.a), histones are an attractive target for chemical protein synthesis. The first total synthesis of a modified histone protein was reported in 2011[31] and was soon followed by rapid advances in SPPS (Fmoc vs. Boc approaches), ligation chemistry (peptide hydrazides[32] as thioester precursors) and synthetic strategy (convergent[33,34], one-pot[32] approaches) to allow almost-routine access to modified histone proteins. Notably, a recent report from the Liu laboratory[32] combined a number of these advances, enabling efficient multi-milligram syntheses of five analogues of H3 and H4, including trimethylated H3-K4me$_3$ (**Figure 5.1**). It can now be envisaged that in the future, these advances, perhaps combined with recently reported flow-based SPPS[35], will render total chemical synthesis the preferred route to obtaining modified histones for biochemistry.

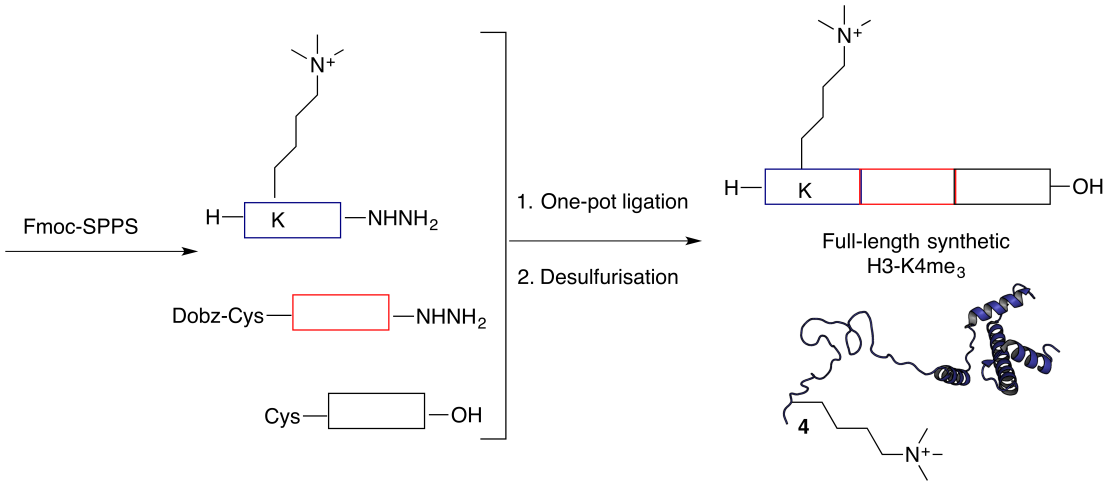


Figure 5.1: Convergent total synthesis of H3-K4me$_3$ via Fmoc-SPPS synthesis of *C*-terminal peptide hydrazides and one-pot ligation[32]

Expressed protein ligation (EPL), involving the ligation of a synthetic peptide to a recombinant protein fragment, is an attractive hybrid approach that extends NCL by incorporating recombinant protein partners in the ligation step[36]. The Muir laboratory, in concert with others, have leveraged EPL to extend the study of histone PTMs to the context of whole nucleosomes and higher order chromatin structures, enabling dissection of PTM-crosstalk and other processes not recapitulated at the peptide level[29,30]. In this work, EPL is the foundational chemical technology

enabling access to homogenously modified proteins. A number of methylated histone proteins have been synthesized by EPL, in some cases proving key reagents for discovery of novel chromatin pathways and mechanisms. Moving beyond synthesis, chemical approaches can also enable novel experiments; the recent combination of EPL with DNA barcoding[37] and sequencing of nucleosomes has been proposed as an example of the 'high-throughput' biochemistry[29] necessary to untangle the combinatorial 'histone code'.

Genetic incorporation of methylated residues into proteins

The methylated derivatives of the canonical amino acids pose a challenge for codon-suppression techniques due to the difficulty in engineering an aminoacyl-tRNA synthetase capable of discriminating methyl groups to charge tRNA with the modified, but not the natural amino acid. Indeed, natural mono-, di- and trimethyllysine residues are not recognized by any reported synthetase. However, introduction of a carbamate group enables recognition and incorporation of N^{ε}-Boc-lysine and N^{ε}-Boc-methyllysine by the wild-type pyrrolysyl-tRNA synthetase[38]. The Chin group has designed a route to mono-[39] and dimethyllysine[40] incorporation into histone proteins that utilizes Boc-protected amino acids, containing a carbamate linkage, in conjunction with the wild-type *M. barkeri* pyrrolysyl-tRNA/ pyrrolysyl-tRNA synthetase pair (**Figure 5.2**). To access monomethyllysine-containing proteins, N^{ε}-Boc-methyllysine is incorporated into a protein, the protein is purified and the Boc-group is removed using a dilute trifluoroacetic acid (TFA) solution.

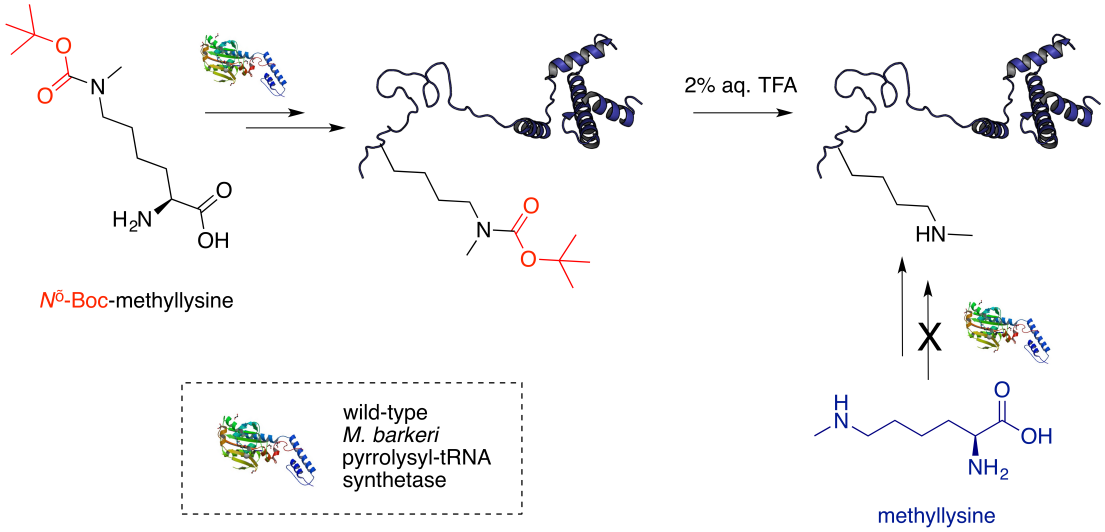


Figure 5.2: Incorporation of *N*-Boc-protected methyllysine by the *M. barkeri* pyrrolysyl-tRNA synthetase enables a route to the genetic incorporation of methyllysine[39]

The route to dimethyllysine is more circuitous[40], involving first the expression of the target protein encoding N^{ε}-Boc-lysine followed by global protection of the remaining lysines using Cbz-succinimide. Orthogonal deprotection of the Boc group followed by reductive dimethylation and subsequent global Cbz deprotection gives the target dimethyllysine-containing protein. While ingenious, the numerous chemical steps involving a cocktail of less-than-mild reagents (such as the strong acid TFA) render the method limited to highly stable proteins, such as the histones originally reported.

Directed evolution of the wild-type *M. barkeri* pyrrolysyl-tRNA synthetase has been used to expand the scope of non-canonical amino acids incorporated into proteins using this system. Photocaged methyllysine[41,42] (caged by a *o*-nitrobenzyl carbamate) derivatives have been incorporated into proteins using evolved pyrrolysyl-tRNA synthetases, in order to overcome the use of strong acid in the route towards methyllysine incorporation described above. The use of UV light for deprotection also enables the liberation of the target methyllysine in living cells, including in mammalian systems[42], enabling the study of this modification in a native context.

Tag and modify

The 'tag and modify'[43] approach to site-selective protein modification involves the chemoselective modification of a uniquely reactive 'tag' residue incorporated during expression. Cysteine is the preferred 'tag' amongst canonical amino acids, and has been used to generate 'methyl-lysine analogues'[44], MLAs, by alkylation with aminoethyl halide reagents (**Figure 5.3**). As the core histones contain only a single cysteine (H3-110C, mutated to alanine with no gross effect on nucleosome function), this strategy is attractive for the installation of MLAs into histone proteins. Notably, unlike EPL methods, the MLA method enables PTM installation at core sites (distant from terminii accessible for ligation). As with most cysteine-alkylation strategies, careful control of pH and reagent loading is required for optimal conversion and control of side-product formation. Histone H3-K9me$_3$ MLAs were recognised by antibodies generated against the native modification and could be reconstituted into nucleosomes for use in assays of both 'reader' binding and enzymatic activity.

The conversion of cysteine to dehydroalanine has also been used for the installation of thioether-linked mimics of lysine methylation[45] (**Figure 5.3**). The resultant mimics were shown to be specifically recognised only by antibodies raised towards the cognate PTM.

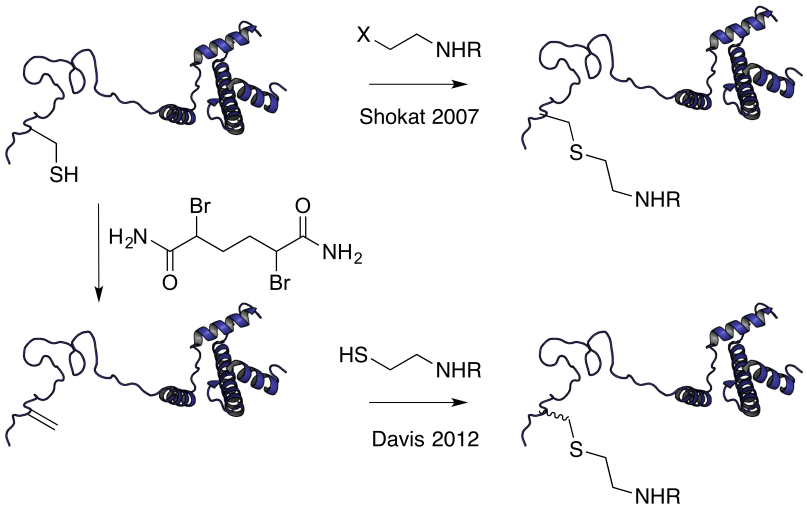


Figure 5.3: Cysteine modification to install thia-analogues of lysine PTMs. *Top*: Alkylation with aminoethyl halides[44] *Bottom*: Conversion to dehydroalanine followed by thia-Michael addition[45].

Goals of this chapter

In order for the developed 'chemical mutation' protocol to be of broad utility to biochemists, the PTMs thus generated must act in a manner (ideally) identical with that of the natural PTM installed enzymatically. In the next two chapters, biochemical assessment of the *functional* relevance of chemically introduced methylated and phosphorylated residues is described. Further, the potential of radical C-C bond formation at protein-Dha to provide *novel* probes for biological studies is also explored.

Results

Installation of defined Lys methylation states to H3-Dha

Utilising H3K27-Dha as a model protein, we previously demonstrated the chemical installation of lysine and its mono- and di-methylated congeners via treatment with the requisite iodide in the presence of $NaBH_4$ (**Figures 4.7 and 4.8, Chapter 4**). This result could be extended to trimethyllysine; treatment of H3-K27Dha with (3-iodopropyl)-trimethylammonium iodide **1** and $NaBH_4$ afforded the desired H3-K27me$_3$. Thus, all three lysine methylation states are readily accessible via the chemical mutation method (**Figure 5.4**).

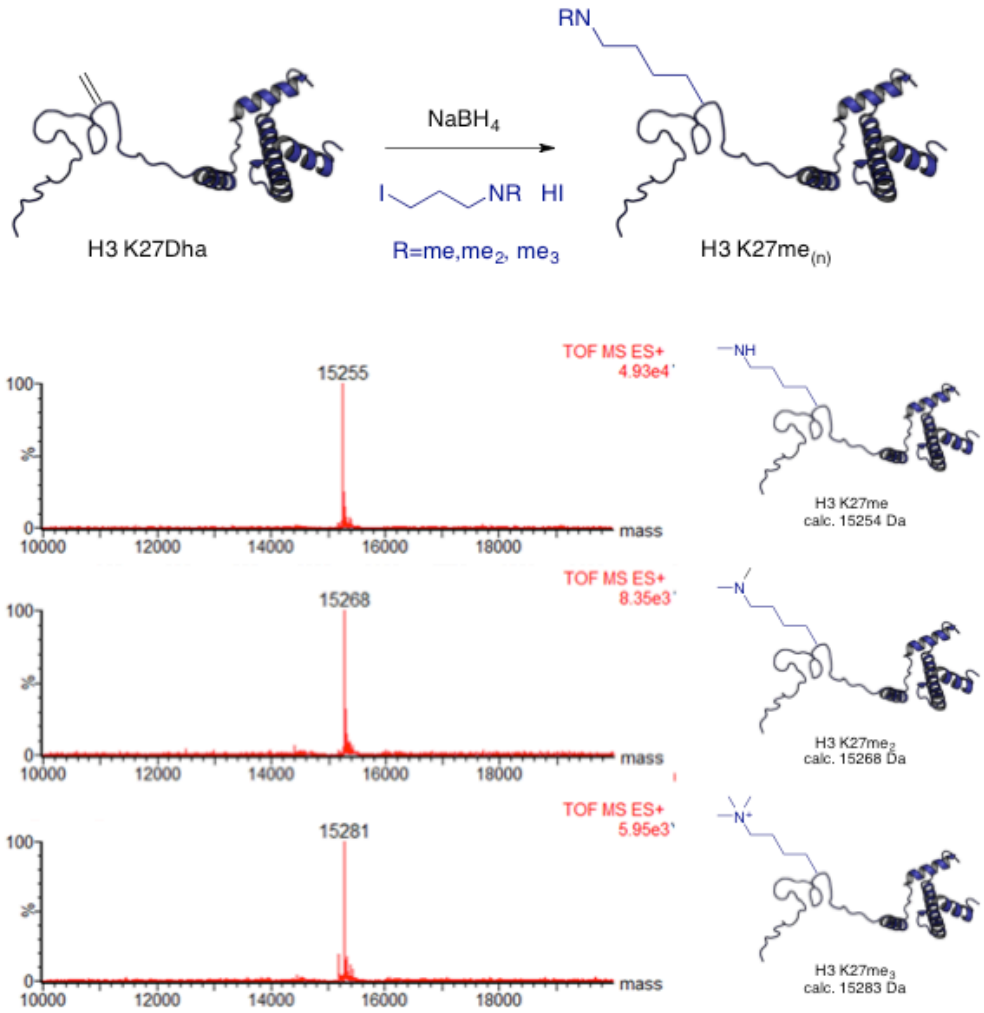


Figure 5.4: All three Lys methylation states are readily accessible *via* borohydride-mediated chemical mutation. See **Experimental Section** for details.

Installation of ^{13}C-methyl labelled H3K9me$_3$ as an 'isotope PTM'

Using 'post-translational mutagenesis', the atomic composition of the introduced amino acid side chains can be readily modulated by synthetic tailoring of the halide small molecule side-chain 'donor', allowing installation of precise variants not found in nature. One application that can be envisioned is isotope labeling[46-48]. We hypothesized that by simply using readily available ^{13}C labeled materials in the synthesis of the iodide side-chain 'donors' for mutagenesis, facile protein labeling with this NMR-active isotope[49] could be achieved (**Figure 5.5**). Current methods for isotopic labeling of natural residues are based on "feeding" labeled precursors during biosynthesis[46], requiring costly excesses of labeled amino acids. Further, such methods result in incorporation at every codon-determined site; using the proposed post-expression modification, isotopic labels can be installed at a single site. Such site-specific isotopic labels could find varied applications in NMR spectroscopy[50] and MS-based proteomics[51].

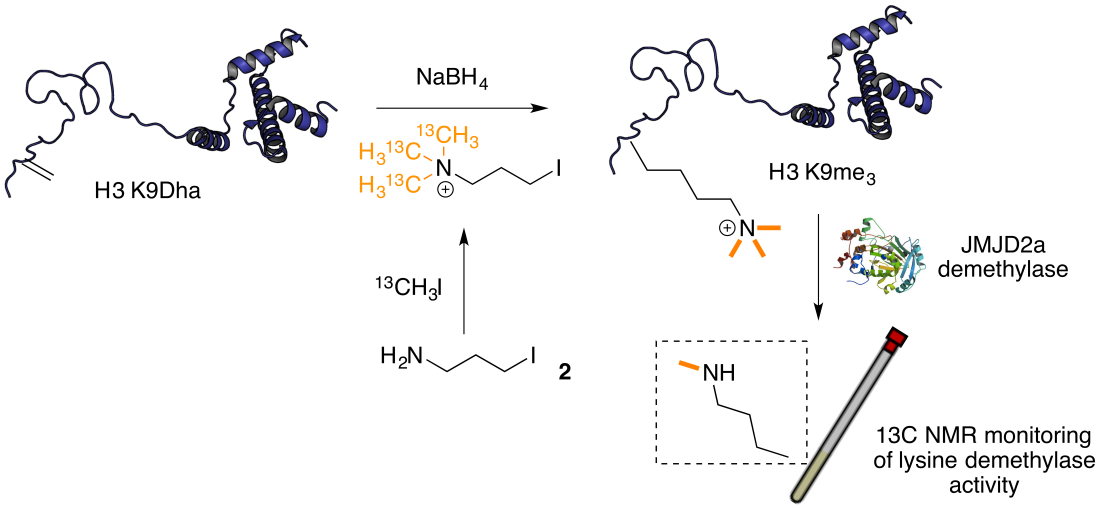


Figure 5.5: Installation of ^{13}C-trimethyllysine as an 'isotope PTM'

To illustrate this concept, we sought to install a variant of Nε-trimethyllysine in which the methyl groups contained ^{13}C-labels. ^{13}C labeled trimethylation should be readily observable using NMR spectroscopy[52] (**Figure 5.5**), potentially enabling observation of the enzymatic processing and release of this PTM. Such an NMR-based demethylase assay has been reported utilizing labeled peptides as

substrates[52]; our proposed method for isotope incorporation could allow direct use of proteins. Given the facile nature of nucleosome reconstitution from recombinant histones, the labels should also be readily incorporated into nucleosomes, potentially enabling NMR-based demethylase assays to be performed on whole nucleosomes[53-55]. Recent NMR-based approaches to enzymatic profiling have also demonstrated the feasibility of using cellular lysates[56,57], and even living cells[58], in these assays, highlighting opportunities to extend this approach to the identification of enzymatic processing of PTMs in more native-like environments.

H3-K9me$_3$ was chosen as a model system to assay the biological function of chemically installed lysine methylation. As outlined in the introduction to this chapter, a number of interacting partners, including a demethylase, KDM4a/JMJD2a[25], are known for this PTM, rendering it a good target for validation.

The required H3-K9C mutant was expressed on a large scale (4 L of culture providing ~200 mg of Cys-protein) and converted to H3K9Dha as previously described. After brief optimization of the iodide loading, H3-K9me$_3$ (^{12}C) could be readily installed as the sole product of the reaction between H3-K9Dha and **1** in the presence of $NaBH_4$ (**Figure 5.6**, top panel). In order to enable stable isotope labeling, ^{13}C-iodomethane was employed as a ready source of ^{13}C-methyl groups. Alkylation of (3-iodopropyl)amine hydroiodide **2** with this reagent provided expedient access to the ^{13}C-trimethylated iodide ^{13}**1** necessary for installation of ^{13}C-trimethyllysine to proteins (**Figure 5.5**, previous page). Next, the iodide was tested in the chemical mutation reaction of H3-K9Dha to afford ^{13}C-labeled H3K9me$_3$. Using conditions identical to those previously optimized for the ^{12}C-trimethyllysine variant, the reaction afforded full conversion to the desired ^{13}C-labeled H3K9me$_3$ (**Figure 5.6**, bottom panel).

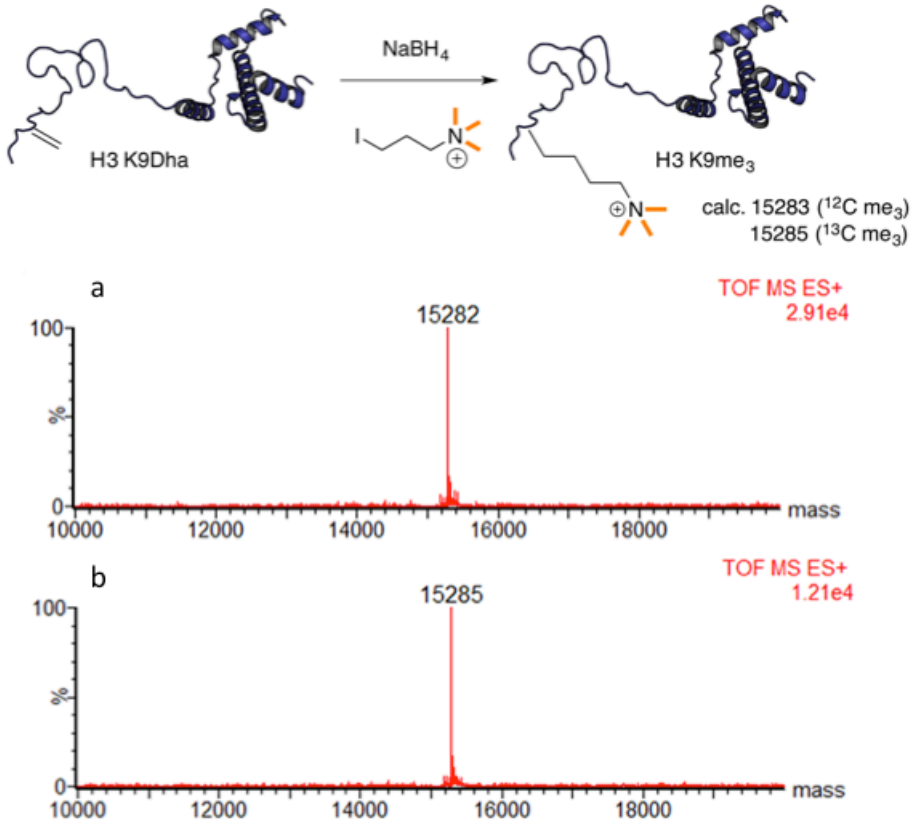


Figure 5.6: Installation of H3-K9me$_3$ and a ^{13}C-methyl variant *via* chemical mutation. *Conditions:* 1 mg/mL (66 µM, pH 6.0 500 mM NH_4OAc, 5M Gdn.HCl) H3-K9Dha, 1000 eq. NaBH4, 1000 eq. (3-iodopropyl)trimethylamine iodide **1** or ^{13}**1**, 30 minutes, 4 °C)

To test whether Nε-trimethyllysine installed by chemical mutation would be recognized and processed by demethylase 'eraser' enzymes, both ^{12}C and ^{13}C variants of H3-K9me$_3$ were treated with KDM4a, a known 'eraser' for this PTM (**Figure 5.7**). The His-tagged catalytic domain of KDM4a (kindly provided in plasmid form by A. Kawamura) was expressed in *E. coli* and purified by Ni-affinity and gel-filtration chromatography[59]. Chemically modified histones were dialysed into non-denaturing buffer (50 mM HEPES, pH 7.5) for the enzyme assay. Enzymatic reactions were conducted under conditions similar to those previously reported for demethylation of H3-K9me$_3$ peptides[52,59]. Initial reactions were monitored by LC-MS, which indicated robust demethylation of both ^{12}C and ^{13}C-methylated H3-K9me$_3$ proteins prepared by chemical mutation (**Figure 5.7**). Demethylation

proceeded, as expected, to the monomethyl state, and was observed to proceed through a dimethyl intermediate (**Figure 5.8**).

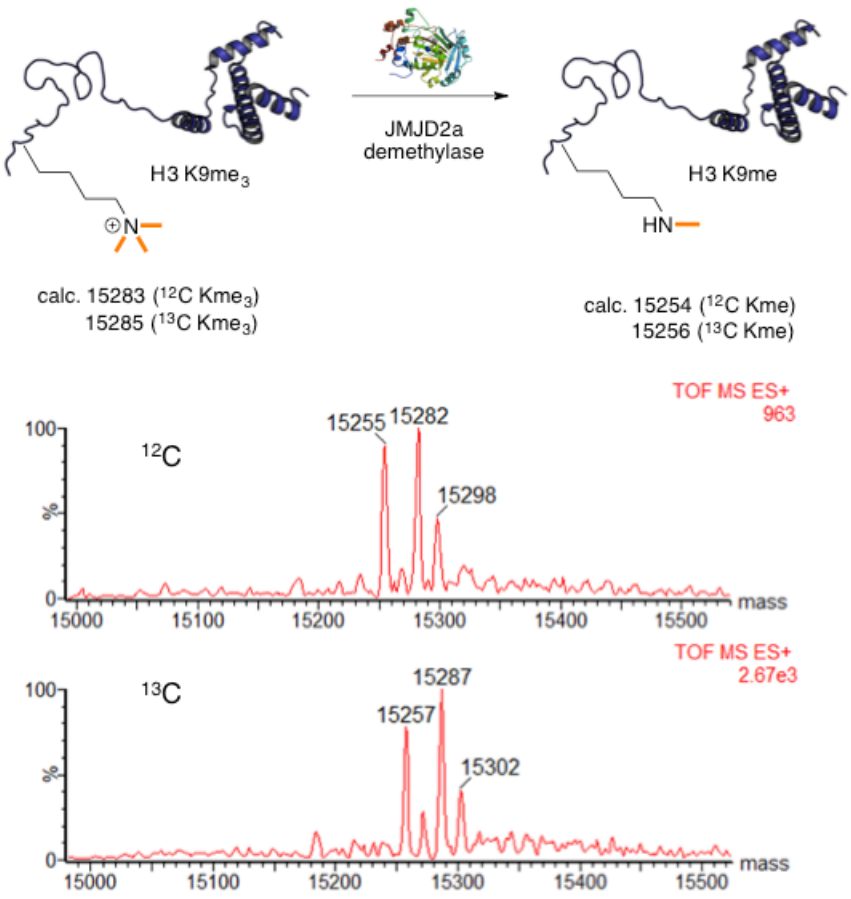


Figure 5.7: Demethylation of ^{12}C and ^{13}C H3-K9me$_3$ using KDM4a, t=10m monitoring by LC/MS. See **Experimental Section** for details.

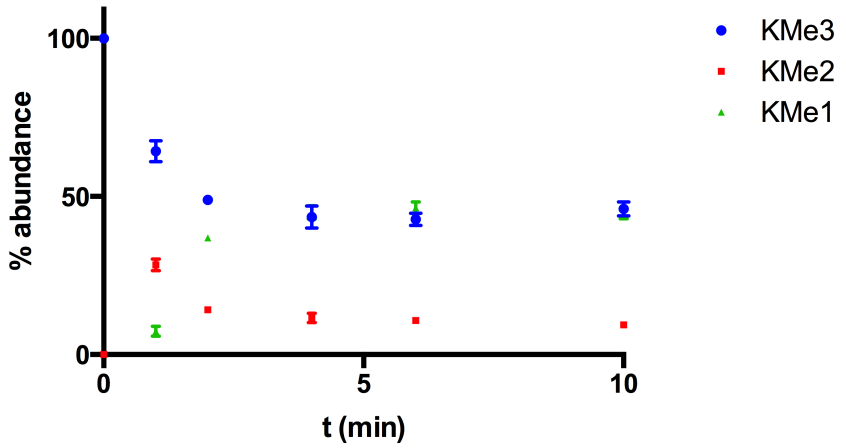


Figure 5.8: Time-course for KDM4a-catalysed demethylation of H3-K9me$_3$(^{12}C) prepared by chemical mutation. *Conditions:* 20 μM H3-K9me$_3$, 2 μM KDM4a

Notably, even with application of 'forcing' conditions (additional enzyme, substrate, heating), conversion of the chemically installed H3-K9me$_3$ does not proceed beyond ~50%. We take this result to indicate the inability of KDM4a to either recognize or to catalyse the demethylation reaction of the $_D$-epimer of the H3-K9me$_3$ protein resultant from chemical mutation. The radical mechanism proposed for the key C-C bond forming step of the chemical mutation, and the lack of a stereoselective hydrogen atom transfer to quench the resultant intermediate Cα radical, implies approximately equal formation of both $_D$- and $_L$-configurations at the mutation site. Such epimer selective processing of PTMs has also been reported for the deacetylation of thioether mimics of acetyllysine (installed via conjugate addition of thiols to Dha) by histone deacetylases (HDACs)[45].

Having confirmed enzymatic processing of the chemically installed modifications, it was of interest to determine whether the ^{13}C trimethylated 'isotope PTM' would enable NMR monitoring of the demethylation reaction. NMR techniques have the advantage of directly reporting on the atomic connectivity of the observed nucleus, adding 'chemical' context to the simple mass shifts observed by LC-MS. In order to explore this idea, a fresh sample of H3-K9me$_3$(^{13}C) was prepared by chemical mutation, as previously. For NMR experiments, the modified histone was dialysed into ammonium formate buffer to avoid any ^{13}C resonances in the region

expected for the trimethyllysine peak. HSQC experiments on the prepared sample demonstrated the expected single N- methyl peak (3.1, 55.0) for the trimethyllysine group (**Figure 5.9**), consistent with literature reports for this resonance in H3-K9me$_3$ containing peptides[52].

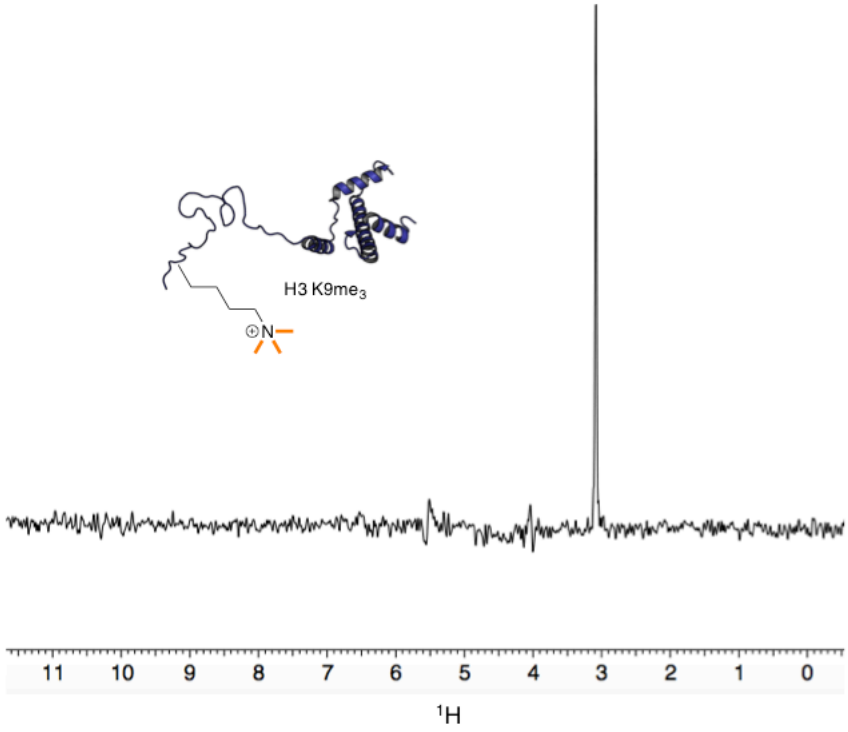


Figure 5.9: 1D ^{13}C-edited HSQC of H3-K9me$_3$(^{13}C)

Next, a sample of the labeled protein was prepared containing all of the cofactors necessary for KDM4a activity. Additional peaks were detected in the HSQC that can be assigned to natural abundance small-molecules in the buffer mixture. However, the trimethyllysine peak remained clearly visible in both 1D and HSQC experiments, enabling monitoring of this peak during demethylation. By simply arraying 1D ^{1}H ^{13}C-edited HSQC type experiments, it was straightforward to monitor demethylation of H3-K9me$_3$(^{13}C) following addition of KDM4a, over the course of an hour (**Figure 5.10**). Although not further investigated in this proof-of-principle study, quantitative kinetic data should be possible from monitoring the decrease in the main trimethyllysine peak. The loss of label and hence signal as methylation state is decreased complicates quantitative analysis

of the dimethyl and methyl peaks. Notably, only 125 μM (initial concentration) of the labelled histone was necessary to observe all methylation states. Finally, when the reaction appeared to slow/cease, a final 2D HSQC was performed (**Figure 5.11**). Pleasingly, there are clear peaks for all three expected species – H3K9me_3, me_2 and me. The chemical shift values correlate well with those previously published[52]. Thus it may be feasible to rapidly use protein NMR to 'fingerprint' methylation states after enzymatic treatment, with relatively low protein concentrations required.

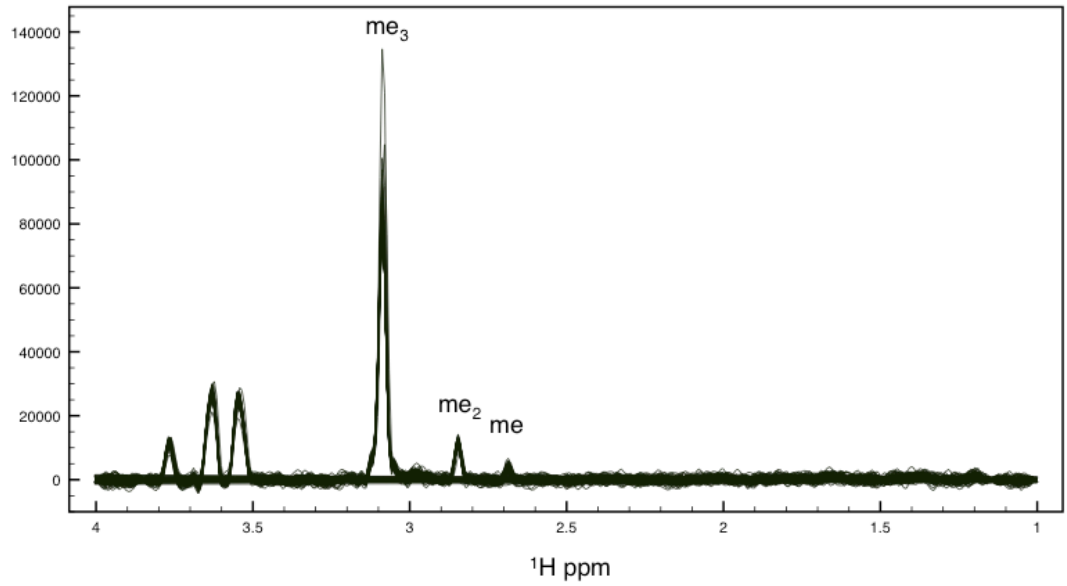


Figure 5.10: Monitoring of KDM4a-catalysed demethylation via 1D ^{13}C-edited HSQC. The trimethyl peak can be seen to decrease in intensity by ~50% over the course of an hour.

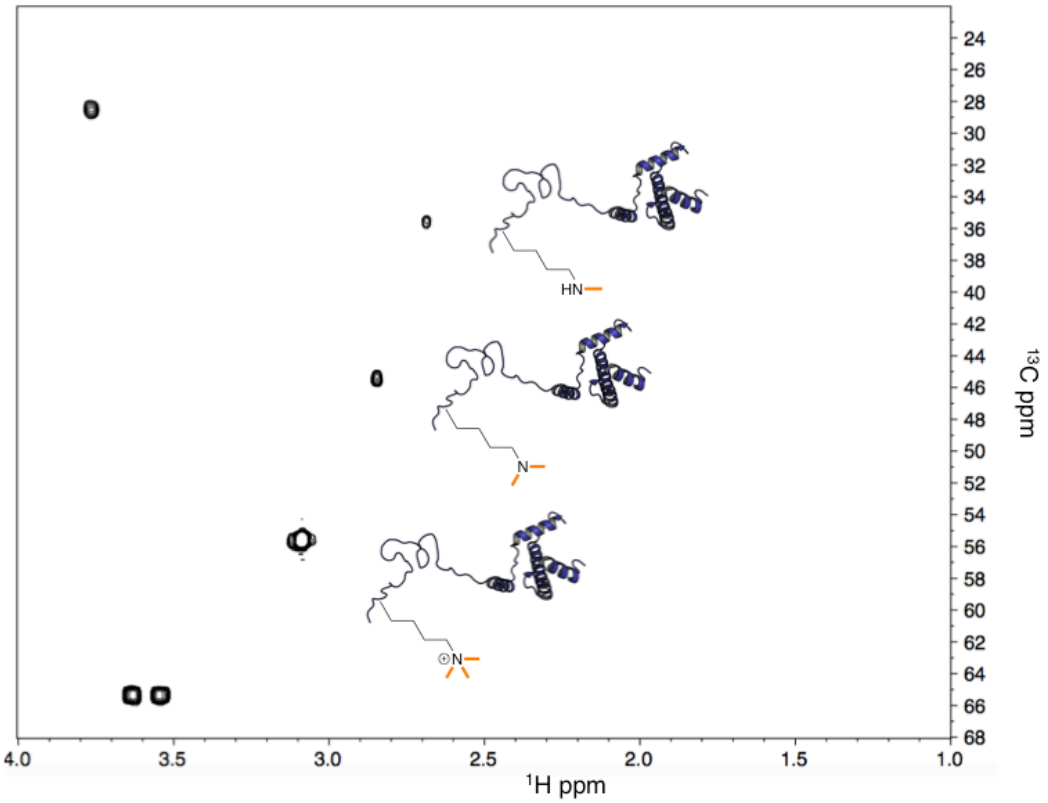


Figure 5.11: 2D HSQC analysis of the demethylation reaction.

Towards the application of 'isotope PTMs' in complex environments

In a native cellular environment, the biochemical and biophysical context for enzymatic activity is very different to the conditions usually employed for *in vitro* experiments. This contrast is particularly marked in investigations of histone PTM processing enzymes, where small peptides are used as surrogates for the complicated nucleosome and chromatin fibre structures on which the enzymes work in the cellular environment[29]. Elegant studies have demonstrated differences in affinity or activity for a number of PTM-processing enzymes dependent on the complexity of the substrate employed for their assay[29]. Hence, it is logical to attempt to conduct *in vitro* enzymatic assays in as 'native-like' an environment as is feasible.

Towards this end, it was sought to extend NMR studies of H3-K9me$_3$ to a nucleosomal context. Nucleosomes are readily assembled from recombinant histones[60], and many studies have demonstrated the

incorporation of fully synthetic or semi-synthetic histones into so-called 'designer' nucleosomes[30]. Further, NMR studies of enzymatic reactions on nucleosome substrates have been reported[53-55]. In these studies, fully labelled histones have been used; our site-specific 'isotope PTM' strategy may offer a cheaper and more accessible alternative to the use of fully labelled proteins.

A two-step procedure for nucleosome refolding was employed[61,62]. First, the chemically modified H3-K9me$_3$ (^{13}C) variant was combined with recombinant H4, H2A and H2B (in a 1:1:1.1:1.1 ratio) and dialysed into a high-salt 'refolding buffer' (2 M NaCl, 10 mM Tris pH 7.5, 1 mM DTT, 5 mM β-mercaptoethanol) to form octamers that could be separated from excess H2A/H2B dimer by size-exclusion chromatography (**Figure 5.12a**). Separately, a large-scale preparation of a 145 bp variant of the '601' strong nucleosome positioning DNA sequence was undertaken. Combining the histone octamer with the isolated DNA (in a 1: 0.85 ratio) and subjecting the mixture to step-gradient dialysis to low-salt buffer 250 mM KCl, 10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT) resulted in efficient reconstitution of nucleosomes, as evidenced by Native PAGE (**Figure 5.12b**). In tandem to the above procedures, wild-type nucleosomes containing unmodified H3 were also prepared to act as a control in subsequent experiments.

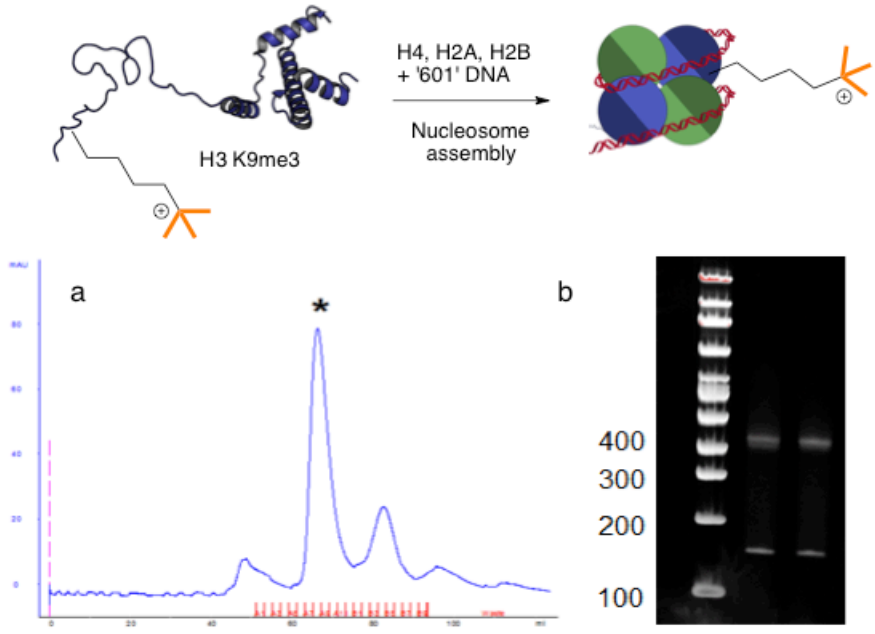


Figure 5.12: Assembly of H3-K9me$_3$ (^{13}C) histone octamers (a) and nucleosomes (b).

Next, we attempted to observe demethylation by KDM4a directly on nucleosomal substrates (**Figure 5.13**). Due to the small amount of nucleosomes initially prepared, we sought a sensitive method to qualitatively assess demethylation. Western blotting using appropriate modification-specific antibodies is a common method for detection of histone PTMs and can be used with small quantities of material. The Fujimori laboratory has previously demonstrated the demethylation of trimethyl-thia-lysine (H3-K9$_c$me$_3$) containing nucleosomes, prepared via the Shokat method, by KDM4a/JMJD2a[63]. Using conditions adapted from that work, we treated two aliquots of H3-K9me$_3$(^{13}C) containing nucleosomes with all requisite cofactors for KDM4a activity, with one aliquot also being treated with KDM4a. Following overnight incubation, the reaction mixtures were analyzed by SDS-PAGE (staining all present protein by Coomassie staining) and by western blotting with an antibody raised against H3-K9me$_3$ (**Figure 5.13**). The 'control' reaction mixture lacking the demethylase demonstrated robust recognition of the chemically installed H3-K9me$_3$ modification on a nucleosome substrate (**Lane 1, Figure 5.13b)**. In contrast, the western blot demonstrates quantitative loss of signal for

the KDM4a-treated sample, suggesting demethylation, at least to the dimethyl state. Based on the previous results gained from LC-MS and NMR profiling of demethylation of isolated histones, it is likely that the '$_D$' epimer remains and is not detected by the antibody used, which was raised against the native '$_L$' epimer of the modified residue.

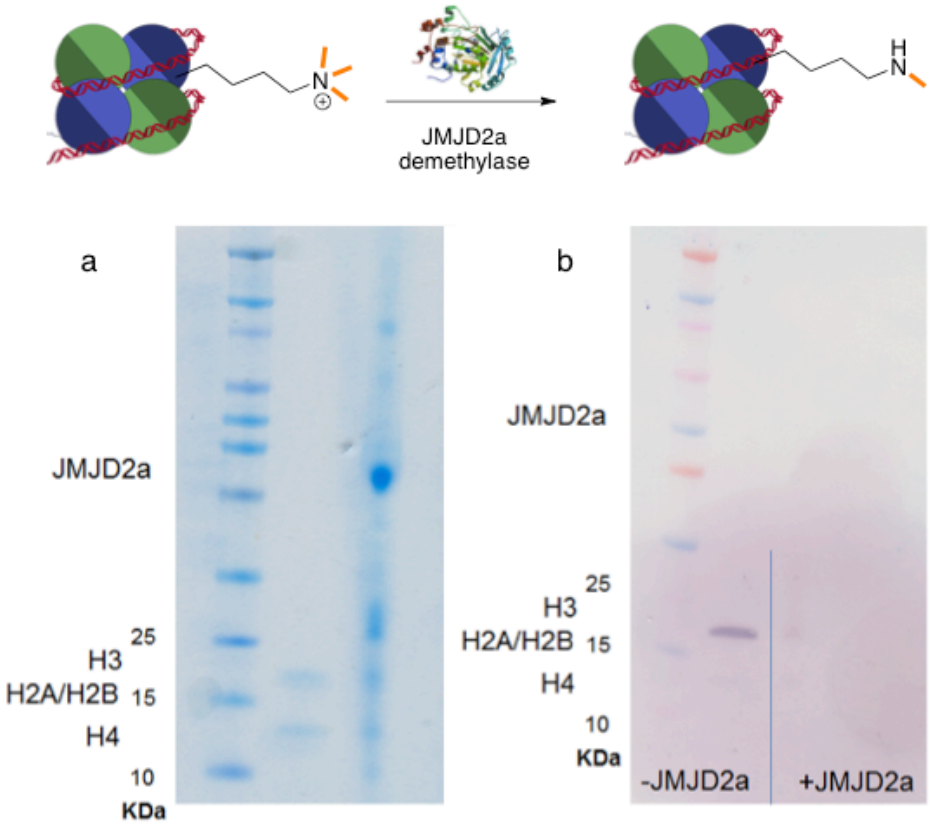


Figure 5.13: KDM4a/JMJD2a catalyzed demethylation of H3-K9me$_3$(^{13}C) nucleosomes. a. Coomassie staining. b. Western blot against H3-K9me$_3$. See **Experimental Section** for details.

Collectively, the above results demonstrate a key proof-of-principle for the application of 'isotope PTMs' to the study of histone PTMs and their processing by relevant enzymes. Extension of these results, to the direct monitoring of enzymatic demethylation on nucleosome substrates using ^{13}C-protein NMR, is ongoing and will be reported in due course.

Installation of defined Arg methylation states to H3-Dha

Protein arginine methylation was first described almost 50 years ago[64] yet today remains a relatively enigmatic PTM. Arginine methylation is known in three forms: methylArg (MMA), symmetric dimethylArg (SDMA) and asymmetric dimethylArg (ADMA), which together constitute one of the most abundant PTMs in mammalian cells[65] (**Figure 5.14**). Protein arginine methylation is an enzymatic process, with a number of 'protein arginine methyltransferases'[66], or PRMTs, having being identified. To date, PRMTs all fall in a single family related by sequence, which has been divided into three classes on the basis of methylation activity[66].

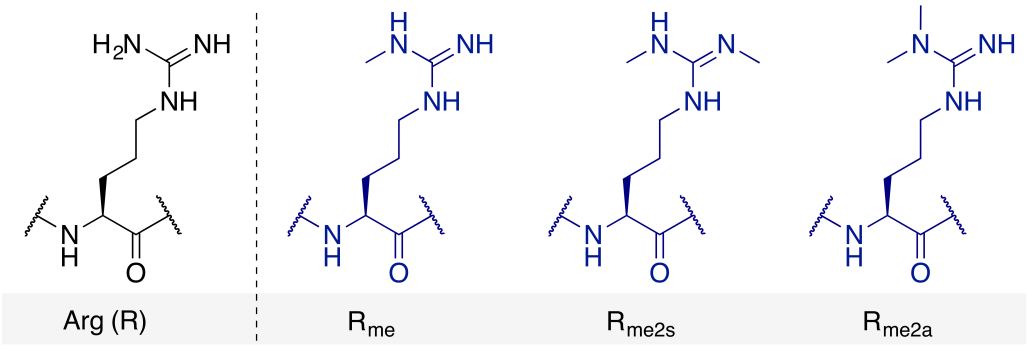


Figure 5.14: Arginine and it's methylated derivatives

The current poor understanding of the biology of arginine methylation is augmented by the paucity of methods for production of proteins bearing defined arginine methylation states. Few studies have reported chemical methods for installation of defined Arg PTMs or their mimics. The Fujimori[67] group has reported a cysteine alkylation strategy to access thioether mimics of Arg methylation. The resultant constructs were shown to be recognized by antibodies raised against the natural modification, and could at least partially recapitulate the binding specificity of the known Tudor-domain containing 'reader' of H4R3me$_2$a, TDRD3**.** 10mer Peptides derived from the histone H4 N-terminal tail containing either the native ADMA modification, or the thia- mimetic, at position R3 were tested as TDRD3 binders by means of a competitive fluorescence polarisation assay. Under the conditions of the assay, the cysteine derived ADMA mimetic

was~1.5-fold less effective as a binder of TDRD3 than the native modification. This result thus highlights the need for alternative methods to enable direct chemical installation of *native* ADMA and other arginine PTMs.

In **Chapter 4**, the chemical introduction of citrulline, a naturally occurring PTM derived from arginine by enzymatic deimination, into histone proteins was described. In order to access methyl arginine derivatives, a robust method was sought to access the requisite iodo-precursor reagents for chemical mutation. Early forays indicated that strategies based on selective alkylation of the guanidine moiety were not practical, and a route via nucleophilic ring-opening of oxazolidin-2-imine **3** was developed (**Figure 5.15**). Ring opening using methylamine or dimethylamine provides the 2-hydroxyethyl guanidine derivatives that can be subjected to iodination without purification. As in the case of citrulline precursor **3.1**, acidic conditions are required in the iodination step to avoid cyclisation. Notably, this route requires no protecting groups and only a single, final chromatography step for purification. Unfortunately, the current route is not amenable to synthesis of the iodo-donor required for SDMA installation.

Initial attempts at chemical mutation using the hydroiodide salts **4** and **5** were notably unsuccessful. It was suspected that this poor reactivity was due to the poor aqueous solubility of the guanidine hydroiodides. To circumvent this, an ion exchange step was incorporated at the end of the synthesis to provide the iodinated guanidines as hydrobromide salts **6** and **7**. These compounds were fully soluble in the aqueous buffer required for protein modification and could be used in chemical mutation reactions. Using reagent **6**, monomethyl Arg (MMA) could be readily installed to the unnatural H3-K64 site using the $InCl_3/NaBH_4$ system as radical initiator, demonstrating the first post-expression chemical installation of any methylArg PTM (**Figure 4.3**, **Chapter 4**).

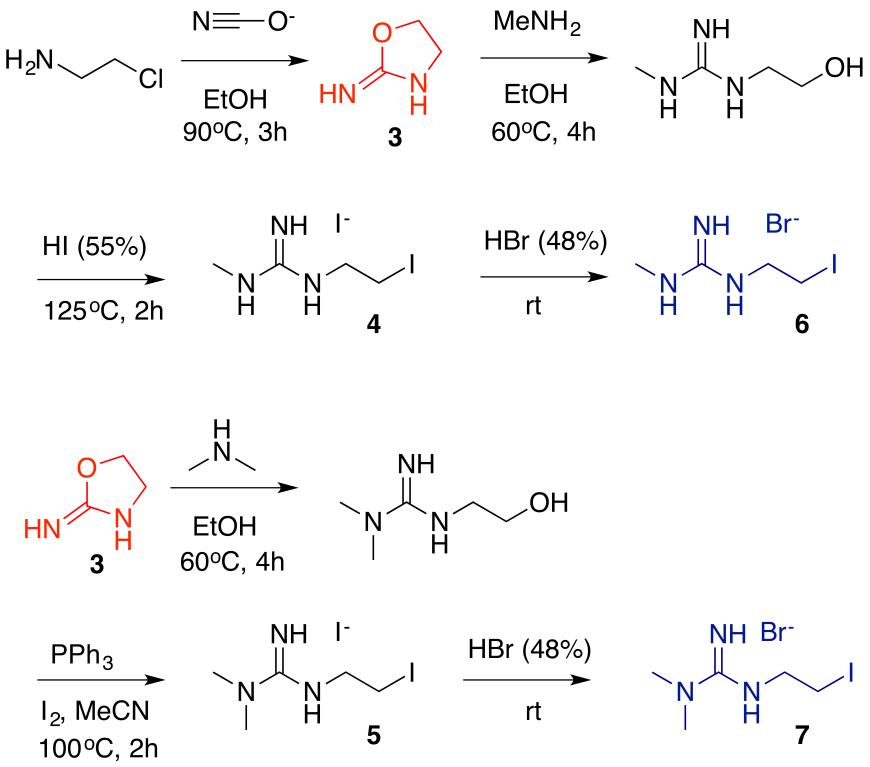


Figure 5.15: A facile route towards the requisite iodo-reagents for chemical mutation to install MMA (top) and ADMA (bottom)

Identification of binding partners for H3-R26me$_2$a from human cell lysates

We next sought to install a known modification of histone H3, H3-R26me$_2$a, in order to investigate further this little-understood PTM in a biologically relevant context.

In cells, H3-R26, along with H3-R2 and H3-R17, is methylated, forming ADMA, by coactivator-associated arginine methyltransferase 1 (CARM1, also known as PRMT 4)[68-70]. CARM1 methylates arginine to yield monomethylarginine (MMA), followed by a second methyl transfer (in a distributive fashion) to ultimately produce asymmetric dimethylarginine (ADMA) residues at the site of modification[71]. In addition to its activity on histones, CARM1 also methylates arginine residues in a range of transcriptional co-activators and splicing factors[72]. In a cellular context, CARM1 has been identified in complex with the SWI/SNF nuclesome remodelling complex[73] and may play a role in 'marking' chromatin for

enforcement of an 'open', transcriptionally active state. Indeed, CARM1 has been demonstrated to promote transcriptional activation[68], particularly in the context of estrogen signaling[69,74]. Stimulation of CARM1 histone methyltransferase activity with histone acetylation at H3-K18 and K-23 has also been reported[75], highlighting yet again the combinatorial nature of histone PTMs and suggesting a range of pathways via which CARM1 mediates an 'open' chromatin state.

In *in-vitro* reactions with CARM1 and short histone peptides, H3-R17 is the kinetically preferred site of methylation, raising the question of the functional role, if any, of the remaining methylation sites[71]. Notably, on the cellular level, H3-R26 methylation has been associated with cell fate determination in early mammalian embryos[76], suggesting a genuine functional role for this residue, at least in certain contexts. In order to fully untangle the functional implications of methylation at a particular arginine, access to the requisite singly- and multiply- modified histones is necessary, preferably in the context of nucleosomes or chromatin arrays. Chemical mutation may be a powerful approach to rapidly generate proteins bearing precisely modified arginine residues for use in such contexts.

In order to generate H3-R26me$_2$a via a chemical mutation strategy, the necessary histone mutant, H3-R26C, was expressed in *E. coli* and purified by size exclusion (SEC) and ion exchange chromatography. Treatment with DBHDA, using identical conditions to those previously described, afforded the requisite H3-R26Dha for chemical mutation. The reaction of H3-R26Dha with **7**, mediated by $NaBH_4$ was then investigated. In initial experiments, reduction to alanine was a competing reduction (~20-30% of the reaction mixture following a 30 minute incubation). It was discovered that this side-reaction could be minimized by careful control of reaction conditions, with a low temperature (<5 °C) during the course of reaction proving beneficial. Under optimized conditions, H3-R26me$_2$a could be installed in >85% purity, based on LC-MS analysis (**Figure 5.16**).

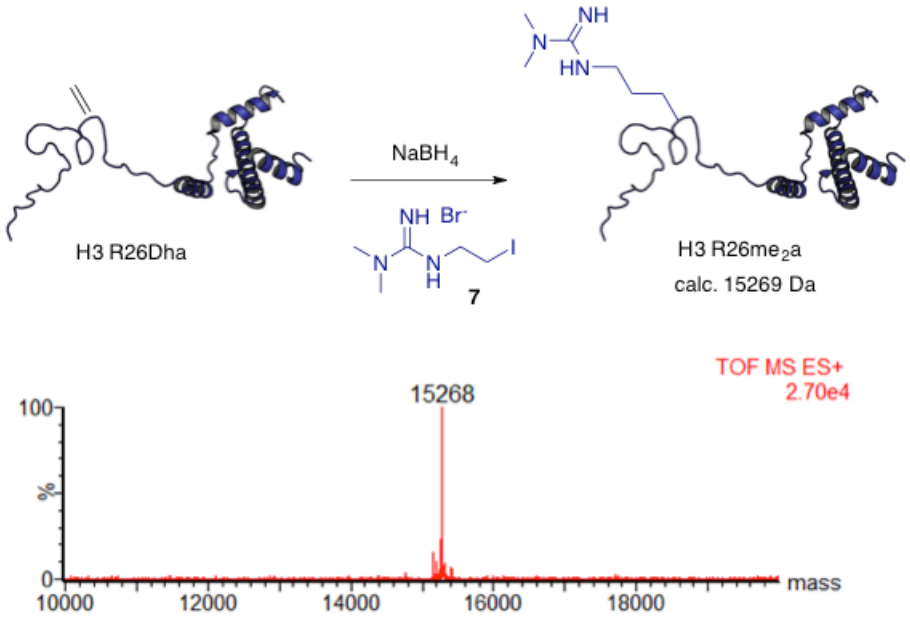


Figure 5.16: Installation of H3-R26$_{me2a}$ via chemical mutation with sodium borohydride. *Conditions:* 1 mg/mL (66 μM, pH 6.0 500 mM NH_4OAc, 5M Gdn.HCl) H3-R26Dha, 2000 eq. $NaBH_4$, 1000 eq. 3-(2-iodoethyl)-1,1-dimethylguanidine, 30 minutes, 4 °C.

With the modified histone H3 in hand, attention turned to generating nucleosomes for use in biochemical assays. The application of LC-MS/MS based proteomic workflows to the identification of potential binding partners ('interaction proteomics'[77]) and/or enzymatic processing[78-80] of the installed ADMA residue was envisaged as a first step in investigation of the functional roles of H3-R26me$_2$a and related arginine modifications (**Figure 5.17**).

For these experiments, we employed a 186 bp biotinylated variant[77] of the '601' nucleosome-binding DNA sequence to facilitate purification ('pull-down') of the treated samples. Both wild-type and H3-R26me$_2$a containing octamers were prepared by dialysis and isolated by size-exclusion chromatography (**Figure 5.18**). The isolated octamers were then combined with the 186 bp biotinylated DNA to reconstitute 'tagged' nucleosomes (**Figure 5.18**).

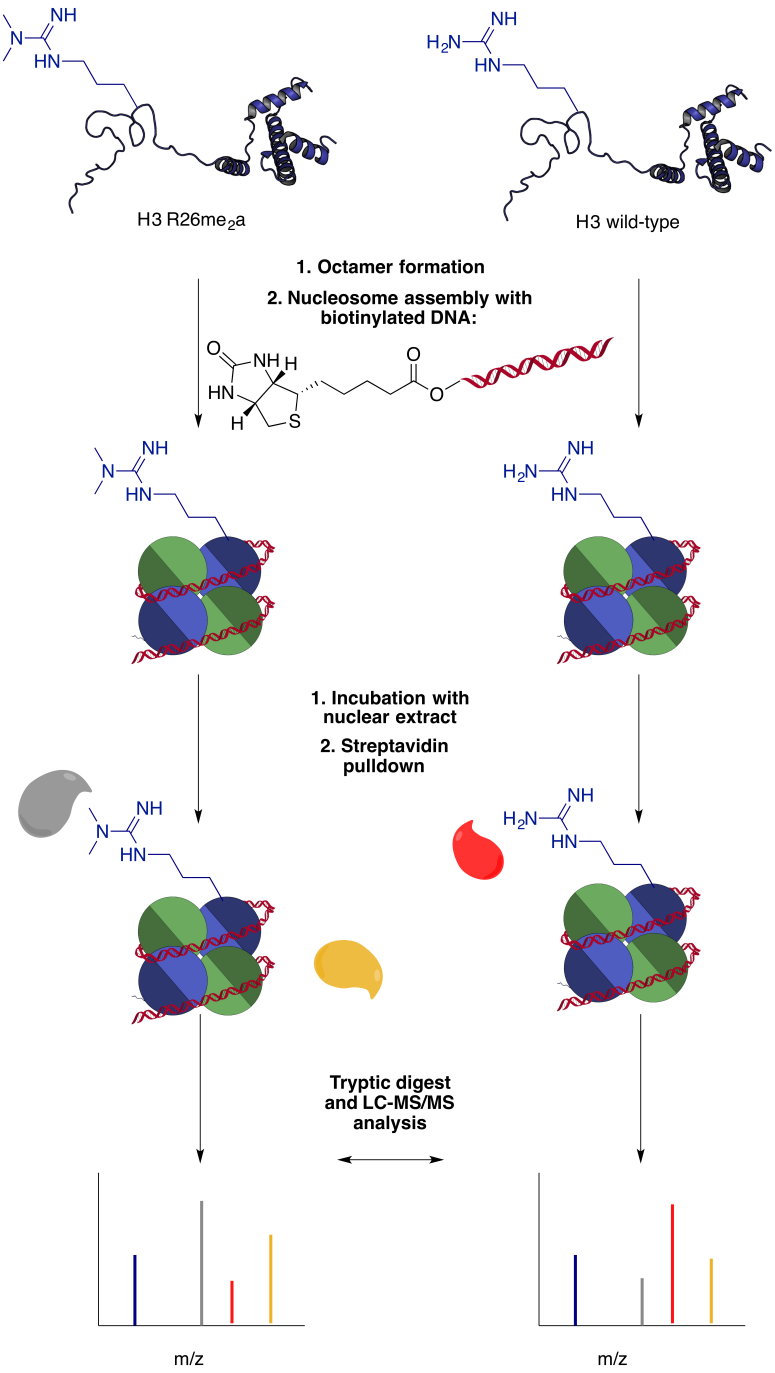


Figure 5.17: Schematic workflow for identification of H3-R26me$_2$a binding partners using nucleosome substrates

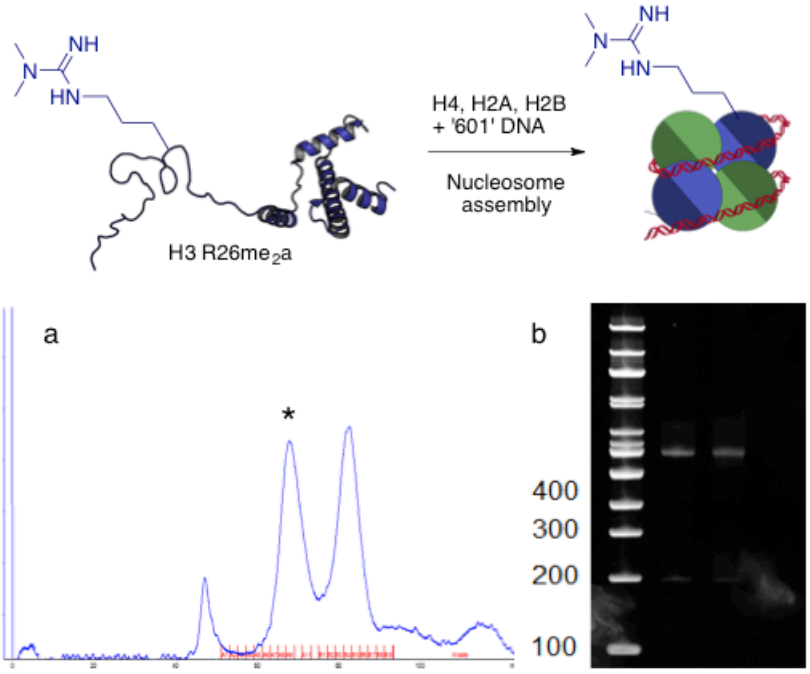


Figure 5.18: Assembly of H3-R26me$_2$a containing histone octamers (a) and nucleosomes (b).

Nucleosome probes containing semi-synthetic H3R26me$_2$a were next used for the identification of binding partners through affinity enrichment proteomics from human cells[62,81] (*experiments performed by Ritu Raj*). First, a small amount (16 µg) of each nucleosome (wild-type or H3-R26me$_2$a) was immobilized on magnetic streptavidin beads to act as the 'bait' for the pull-down experiment. The immobilized nucleosomes were then washed before incubation with HeLa nuclear extracts for four hours at 4 °C. The beads were recovered and rigorously washed with binding buffer to remove non-specifically bound proteins, followed by elution of interacting proteins under stringent, denaturing conditions (8 M urea, 2 M NaCl). Interacting proteins were subjected to proteolytic digest using trypsin and the resultant peptide mixtures separated and analyzed by ultra-high performance liquid chromatography (UHPLC) coupled to mass-spectrometry (Q-Exactive). Label free quantification (LFQ)[82] allowed identification of a number of interacting partners enriched upon asymmetric dimethylation of H3-R26. The results are displayed as **Figure 19.**

Analysis of the proteins enriched by asymmetric dimethylation suggests a number of directions for future research. Intriguingly, two of the strongest interacting partners, BEND3[83,84] and BANP, contain BEN domains[85], α-helical domains recently identified in chromatin-associated proteins. Future experiments could investigate the ability of BEN-domain containing proteins to bind methylarginine motifs *in-vitro*. BEND3, a repressor of rDNA transcription[84], is the single most enriched interacting protein partner. BEND3 can recruit the polycomb PRC2 complex[83] and thus mediate repression of transcription via H3-K27me$_3$. This suggests potential cross-talk between R26 and K27 modifications leading to generation of a repressive chromatin state. Notably, H3-R26me$_2$a has not previously been implicated in establishment of repressive chromatin states. As CARM1 methylation has previously been associated with transcriptional activation[69], this result suggests novel pathways modulated by methylation at R26, as opposed to R17; how these pathways interact in a cellular context will require further investigation.

The majority of the remaining significant interactors (SMARCAL1, RECQL, DDB1, DDB2, TOP3A) are annotated as being DNA binders and/or involved in DNA replication/repair. This suggests increased accessibility of nucleosomal DNA in the H3-R26me$_2$a modified sample relative to the wild-type control. As arginine residues in histones are involved in extensive hydrogen-bonding interactions with DNA, it could be anticipated that methylation, by removing these interactions, could indeed result in a 'loosening' of DNA around the histone octamer. Biophysical experiments, using chemically defined nucleosomes, are under way to explore this possibility[61].

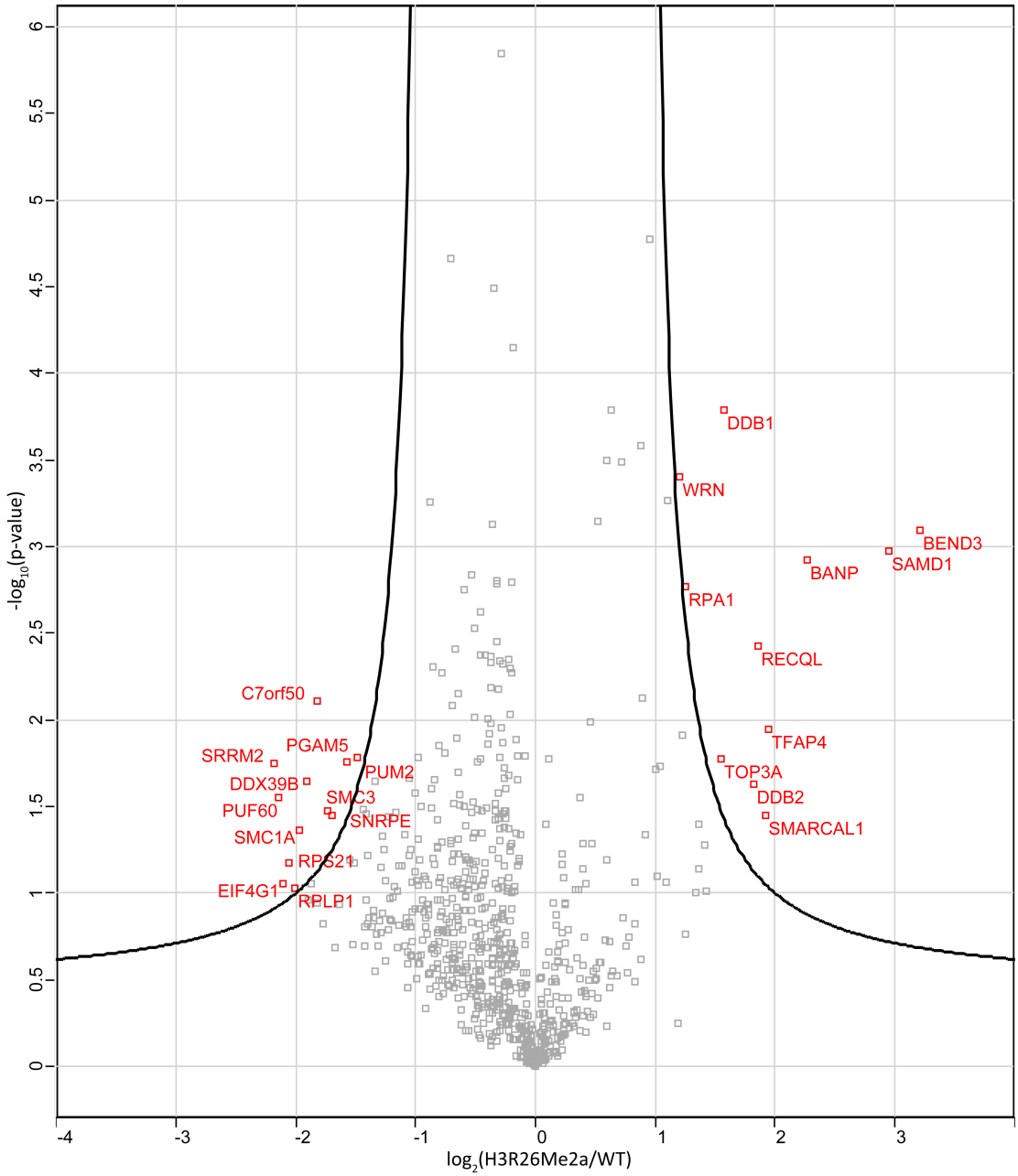


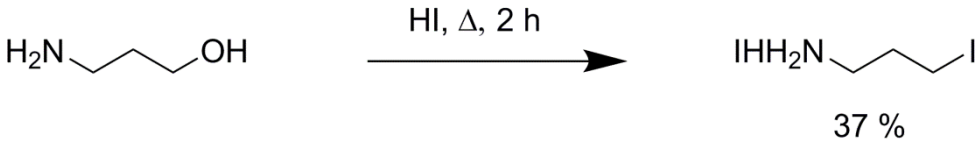
Figure 19: Significant interacting protein partners (denoted by their gene name) identified upon label-free quantification (LFQ) on Student's t test analysis are shown in red in the "volcano plot" inset (x axis, logarithmized ratio of LFQ intensity difference among two groups; y axis, logarithmized P value from the test statistics).

Experimental Section Chapter 5

Reagent synthesis

(3-iodopropyl)amine hydroiodide 2

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in[86]



To ice cold hydroiodic acid (55 %, 13.0 mL), 3-aminopropa-1-ol (2.0 mL, 26.6 mmol) was added. The reaction mixture was heated to 125 ºC and stirred for 2 h. The reaction mixture was cooled to room temperature to form a dark red residue, which was recrystallised twice from EtOH/$CHCl_3$ to give the title compound as colourless prisms (3.06 g, 9.77 mmol, 163-164 ºC, 37 %).

^{1}H NMR δH (400 MHz, MeOD) 3.33 (2H, t, J = 7.3, CH_2I), 3.09 (2H, t, J = 7.3, CH_2NH_2) and 2.20 (2H, quin, J = 7.3, CH_2CH_2I)

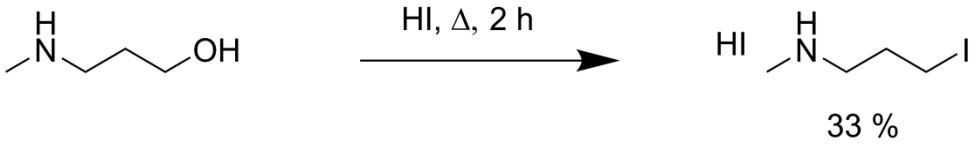
^{13}C NMR δC (100 MHz, MeOD) 40.6 (CH_2NH_2), 31.2 (CH_2CH_2I) and 0.1 (CH_2I)

IR (υ_{max}/cm^{-1}/KBr) 2991, 2288, 1860, 1582, 1588, 1309, 1194, 1120 and 1059

HRMS m/z (ESI$^+$): Found 185.9774 $[M+H]^+$. C_3H_9NI requires 185.9775.

(3-iodopropyl)methylamine hydroiodide

Reagent supplied by Justin Chalker. Procedure is included for reference, originally reported in [86]

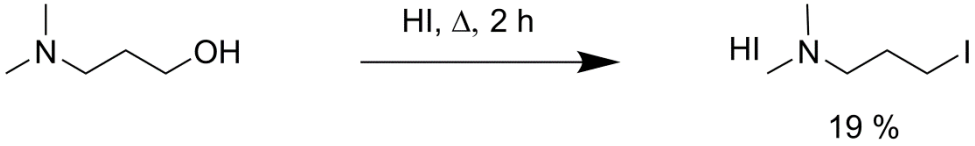


A solution of 3-(methylamino)propan-1-ol (2.00 g, 22.4 mmol) in hydroiodic acid (55 %, 12.0 mL) was heated to 125 ºC and stirred for 2h. The reaction mixture was cooled to room temperature then to -20 ºC

overnight, which allowed the dark red residue to crystallise. The crystals were filtered, washed with Et_2O (100 mL) and then recrystallised ($CHCl_3$/EtOH) and washed with Et_2O (100 mL) to afford the title product as yellow needles, which were dried under reduced pressure (2.41 g, 7.39 mmol, 53-55 ºC, 33 %).

^{1}H NMR *δH* (400 MHz, MeOD) 3.33 (2H, t, *J* = 6.8, CH_2I), 3.16 (2H, t, *J* = 7.8, CH_2NHMe), 2.76 (3H, s, $NHCH_3$) and 2.24 (2H, m, CH_2CH_2I)
^{13}C NMR *δC* (100 MHz, MeOD) 50.1 (CH_2NHCH_3), 33.1 ($NHCH_3$), 30.0 (CH_2CH_2I) and -0.40 (CH_2I)
IR (υ_{max}/cm^{-1}/KBr) 3480, 2935, 2760, 2398, 1457, 1416, 1268, 1203, 1038
HRMS m/z (ESI$^+$): Found 199.9928 $[M+H]^+$; $C_4H_{11}NI$ requires 199.9930

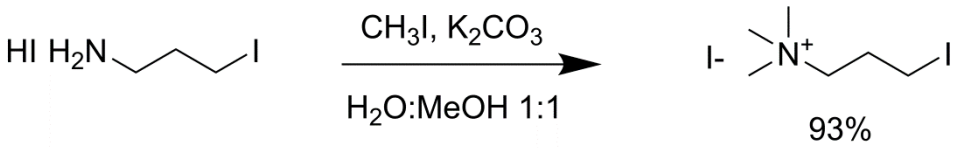
(3-iodopropyl)dimethylamine hydroiodide



19 %

3-dimethylamino-1-propanol (2.0 g, 19.39 mmol) was added to a 25 mL round bottom flask and stirred at room temperature. Hydroiodic acid (55%, unstabilised, 12 mL) was added dropwise to the stirred alcohol, giving a dark-red mixture. The reaction mixture was heated at 125 °C (open air) for 2 hours and then cooled to room temperature. The flask was placed on ice and the product crystallised from the reaction mixture, along with formation of a red solid mass. The solids were filtered and washed with Et_2O. The product was further recrystallised from $CHCl_3$. Remaining iodine was removed by extensive trituration with Et_2O to leave a white solid (1.4 g, 18.2%).
IR (υmax, KBr): 2918, 2686, 1459, 1313, 1187, 1101, 1036, 1004, 959, 843, 727 ^{1}H NMR (400 MHz, MeOD): δ= 2.29 (2H, m, $C\underline{H_2}CH_2I$), 2.95 (6H, s, $N\underline{Me_2}$), 3.28-3.33 (4H, m, contains $C\underline{H_2}NMe_2$ and $C\underline{H_2}I$). ^{13}C NMR (100 MHz, MeOD): δ = -1.69 ($\underline{C}H_2I$), 27.99 ($\underline{C}H_2CH_2I$), 42.39 ($N\underline{Me_2}$), 58.21 ($\underline{C}H_2NMe_2$).
LRMS m/z (ESI+): Found 214.0 [M+H]+; $C_5H_{13}IN$ requires 214.07

(3-iodopropyl)-trimethylammonium iodide **1**



To a solution of (3-iodopropyl)amine hydroiodide **S1 (**365 mg, 1.17 mmol) in a 1:1 mixture of H_2O/MeOH was added potassium carbonate (1 g, 7.00 mmol) and the resultant suspension stirred until dissolution. Iodomethane (440 µL, 7.00 mmol) was added dropwise and the reaction mixture stirred for 1h at room temperature. Solvents were removed under reduced pressure and the resultant white solid was extracted with acetonitrile, taking the desired product into solution. Evaporation under reduced pressure followed by trituration with $CHCl_3$ (10 mL) afforded the title product as a delicate white powder (386 mg, 93%)

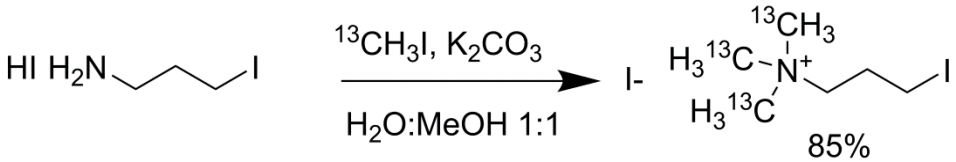
^{1}H NMR *δH* (400 MHz, DMSO) 3.32 (2H, m, CH_2NMe_3), 3.23 (2H, t, *J* = 6.7, CH_2I), 3.06 (9H, s, NMe_3) and 2.23 (2H, q, CH_2CH_2I)

^{13}C NMR *δC* (100 MHz, DMSO) 66.7 (CH_2NMe_3), 53.4 (NMe_3), 30.0 (CH_2CH_2I) and 2.8 (CH_2I)

IR (υ_{max}/cm^{-1}/KBr) 3061, 1469, 1240, 1182, 1013, 961 and 947

HRMS m/z (ESI$^+$): Found 228.0244 $[M]^+$; $C_6H_{15}IN$ requires 228.0242.

^{13}C (3-iodopropyl)-trimethylammonium iodide **^{13}1**

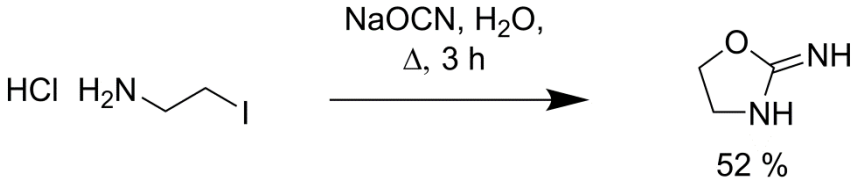


To a solution of (3-iodopropyl)amine hydroiodide **S1 (**365 mg, 1.17 mmol) in a 1:1 mixture of H_2O/MeOH was added potassium carbonate (1 g, 7.00 mmol) and the resultant suspension stirred until dissolution.^{13}C-Iodomethane (*Aldrich 277185*, 440 µL, 7.00 mmol) was added dropwise and the reaction mixture stirred for 1h at room temperature. Solvents were removed under reduced pressure and the resultant white solid was extracted with acetonitrile, taking the desired product into solution.

Evaporation under reduced pressure followed by trituration with $CHCl_3$ (10 mL) afforded the title product as a delicate white powder (355 mg, 85%)

^{1}H NMR δH (400 MHz, DMSO) 3.25 (2H, t, J = 6.8, CH_2NMe_3), 3.07 (9H, d, $^1J_{C\text{-}H}$ = 140.4 Hz, NMe_3), 2.89 (2H, t, J = 3.1, CH_2I) and 2.23 (2H, q, CH_2CH_2I)
^{13}C NMR δC (100 MHz, DMSO) 66.4 (CH_2NMe_3), 52.9 (*t,* NMe_3), 26.5 (CH_2CH_2I) and 2.3 (CH_2I)
HRMS m/z (ESI$^+$): Found 231.0344 $[M]^+$; $^{13}C_3{}^{12}C_3\,H_{15}IN$ requires 231.0344

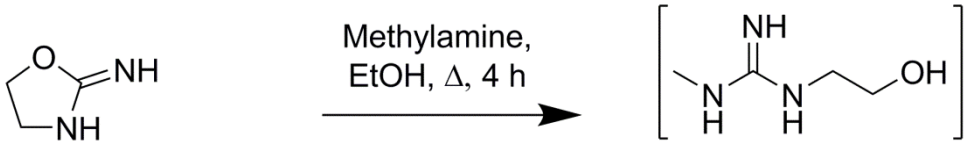
Oxazolidin-2-imine **3**



52 %

To a solution of sodium cyanate (2.00 g, 30 mmol) in water (10 mL) 2-chloroethylamine hydrochloride (3.50 g, 30 mmol) was added. The reaction mixture was heated to 90 °C and stirred for 3 h. Solvents were removed under reduced pressure and extraction performed on the resultant white residue with hot ethanol. The ethanolic fractions were filtered. The title product was precipitated by the addition of cold Et_2O to afford the title product as a white amorphous solid (1.90 g, 15.6 mmol, 52 %).
^{1}H NMR δH (400 MHz, MeOD) 4.84 (2H, t, J= 8.5 Hz, CH_2O) and 3.92 (2H, t, J= 5.7 Hz, CH_2N)
^{13}C NMR δC (100 MHz, MeOD) 163.7 (C=N), 70.91 (CH_2O) and 42.98 (CH_2N)

1-(2-hydroxyethyl)-3-methylguanidine



To an ethanolic (10.0 mL) solution of oxazolidin-2-imine (1.00 g, 11.6 mmol), methylamine in ethanol (33 %, 1.6 mL, 12.2 mmol) was added. The reaction mixture was then heated to 60 °C and stirred for 4 h. The reaction

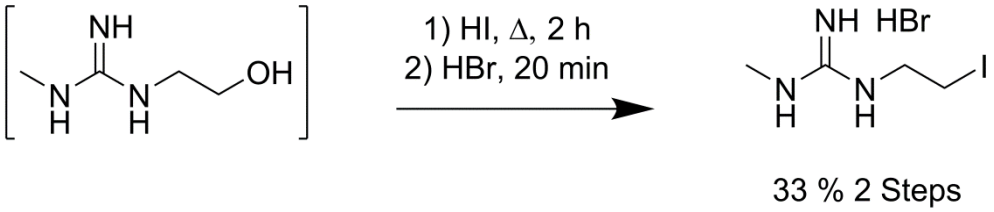
mixture was cooled to room temperature and solvents removed under reduced pressure to afford the title product as a colourless glue (1.31 g) which was used without further purification.

^{1}H NMR *δH* (400 MHz, MeOD) 3.72 (2H, t, *J*= 5.2 Hz, CH_2O), 3.39 (2H, t, *J*= 5.36, CH_2N) and 2.91 (3H, s, CH_3)

^{13}C NMR *δC* (100 MHz, MeOD) 157.5 (C=N), 60.29 (CH_2O), 43.96 (CH_2N) and 27.36 (CH_3)

LRMS: m/z (ESI+): Found 118.2$[M+H]^+$; $C_4H_{12}N_3O$ requires 118.10

1-(2-iodoethyl)-3-methylguanidine, bromide salt **6**



33 % 2 Steps

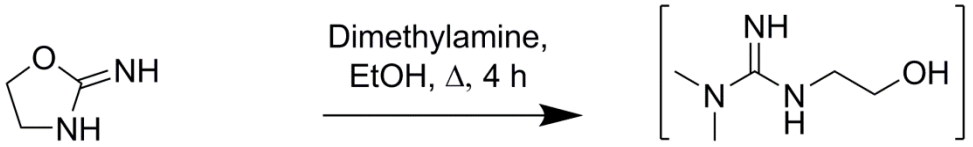
A solution of crude 1-(2-hydroxyethyl)-3-methylguanidine (1.29 g) in hydroiodic acid (55 %, 5 mL) was heated to 120 °C and stirred for 4 hr. The reaction mixture was cooled to room temperature and solvents removed under reduced pressure to afford a red brown residue which was purified by column chromatography (0 -> 20 % MeOH/DCM) to afford a hygroscopic red solid. The hydroiodide salt was dissolved in hydrobromic acid (47 %) and stirred for 20 mins. Solvents were removed under pressure to afford the title product as an extremely hygroscopic yellow solid (1.18 g, 3.83 mmol, 33 % (two steps)).

^{1}H NMR *δH* (400 MHz, MeOD) 3.54 (2H, t, *J*= 6.72 Hz, CH_2N), 3.29 (2H, t, *J*= 6.72, CH_2I) and 2.83 (3H, s, CH_3)

^{13}C NMR *δC* (100 MHz, MeOD) 44.20 (CH_2N), 27.82 (CH_3) and 1.42 (CH_2I)

HRMS: m/z (ESI+): Found 227.9993 $[M+H]^+$; $C_4H_{11}IN_3$ requires 227.9992

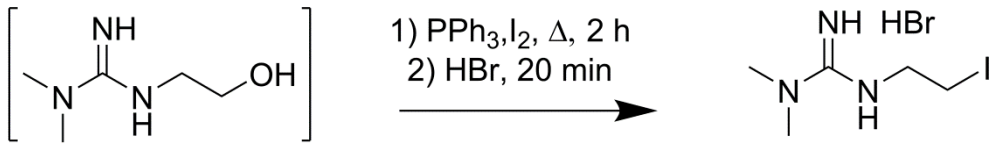
3-(2-hydroxyethyl)-1,1-dimethylguanidine



To an ethanolic (10 mL) solution of oxazolidin-2-imine (1.00 g, 11.6 mmol), dimethylamine in ethanol (33 %, 1.9 mL, 12.2 mmol) was added. The reaction mixture was then heated to 60 °C and stirred for 4 h. The reaction mixture was cooled to room temperature and solvents removed under reduced pressure to afford the title product as a colourless glue (1.38 g), which was used without further purification.

^{1}H NMR *δH* (400 MHz, MeOD) 3.67 (2H, t, *J*= 10.04 Hz, CH_2O), 3.36 (2H, t, *J*= 10.00, CH_2N), and 3.03 (6H, s, CH_3),
^{13}C NMR *δC* (100 MHz, MeOD) 157.29 (C=N), 60.70 (CH_2O), 46.97 (CH_2N) and 35.17 (CH_3)
LRMS m/z (ESI+): Found 132.1$[M+H]^+$; $C_5H_{14}N_3O$ requires 132.19

3-(2-iodoethyl)-1,1-dimethylguanidine, bromide salt **7**



14 % 2 Steps

To a suspension of 3-(2-hydroxyethyl)-1,1-dimethylguanidine (1.38 g, 10.5 mmol) in MeCN (100 mL) triphenylphosphine (4.12 g, 15. 7mmol) and iodine were sequentially added. The reaction mixture was heated to reflux and stirred for 2 h at which point all components were dissolved and analysis by TLC (20 % MeOH/DCM, potassium permanganate) indicated complete consumption of the starting material. Solvents were removed under reduced pressure to afford a think black residue which was purified by column chromatography (0->20 % MeOH/DCM) to afford a red solid. The hydroidodide salt was dissolved in hydrobromic acid (47 %) and stirred for 20 mins. Solvents were removed under reduced pressure to

afford the title product as an extremely hygroscopic amorphous yellow solid (0.540 g, 1.67 mmol, 14 % (2 steps)).

^{1}H NMR *δH* (400 MHz, MeOD) 3.53 (2H, t, *J*= 6.82 Hz, CH_2N), 3.27 (2H, t, *J*= 6.82, CH_2I) and 2.99 (6H, s, CH_3)
^{13}C NMR *δC* (100 MHz, MeOD) 156.58 (C=N), 44.84 (CH_2N), 38.05 (CH_3) and 1.42 (CH_2I)
HRMS m/z (ESI+): Found 242.0149 [M+H]$^{+}$; $C_5H_{13}IN_3$ requires 242.0149

Protein chemical modification

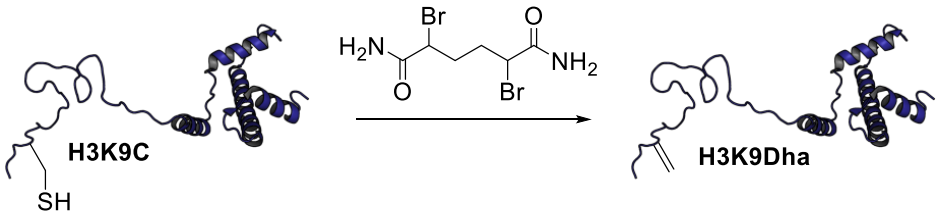
Histone H3 (K9C, C110A)

ARTKQTARCSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR
YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN
LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214

Observed mass = 15214

Histone H3 K9Dha

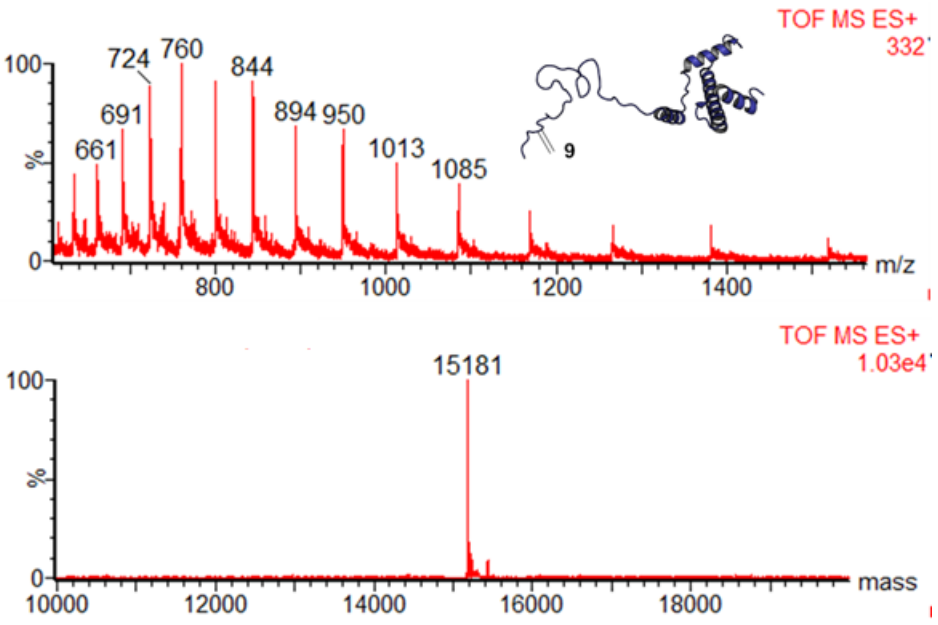


The dehydroalanine-containing histone was prepared according to general procedure **1.**

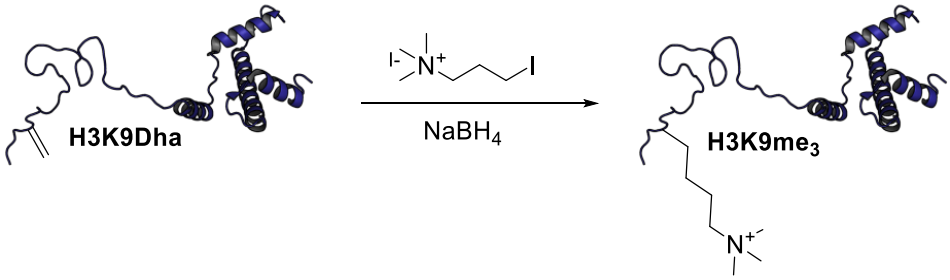
ESI-MS spectrum for the modified histone is shown below.

Calculated mass = 15180.7

Observed mass = 15181



Installation of trimethyllysine from dehydroalanine H3 K9



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K9Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

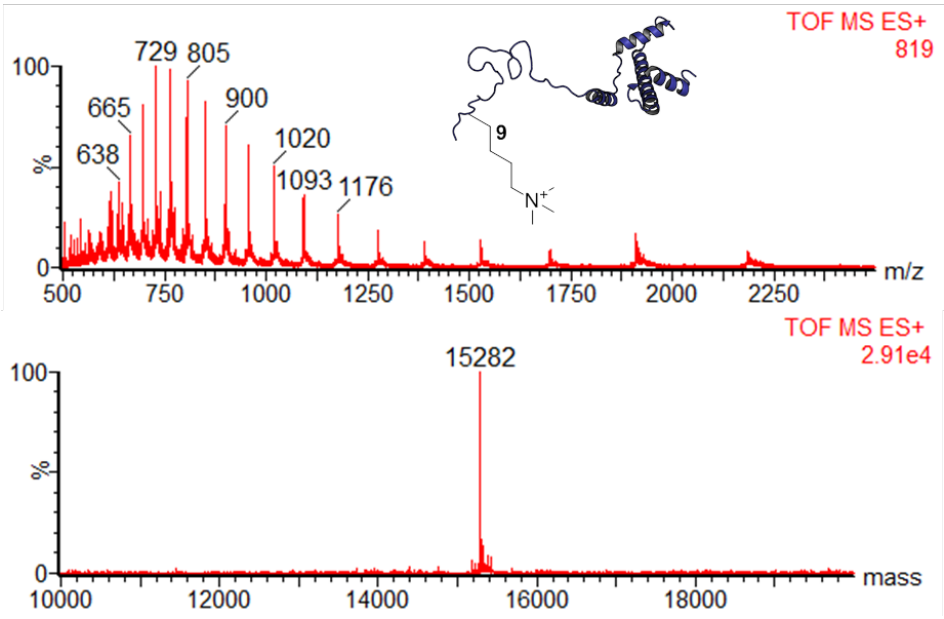
The following morning, within the glovebox, 5 mg solid (3-iodopropyl)trimethylamine iodide **S6** was dissolved in the degassed protein solution and the resultant pale solution gently mixed by pipette.

The entire volume (~200 µL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for one hour before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

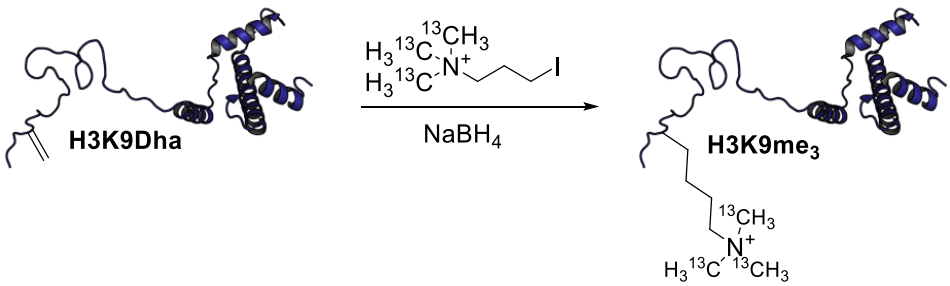
ESI-MS for the modified histone is shown below.

Calculated mass: 15282.8

Observed mass: 15282



Installation of ^{13}C-trimethyllysine from dehydroalanine H3 K9



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K9Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, 5 mg (3-iodopropyl)^{13}C-trimethylamine hydroiodide **S7** was dissolved in the degassed protein solution and the resultant pale solution gently mixed by pipette.

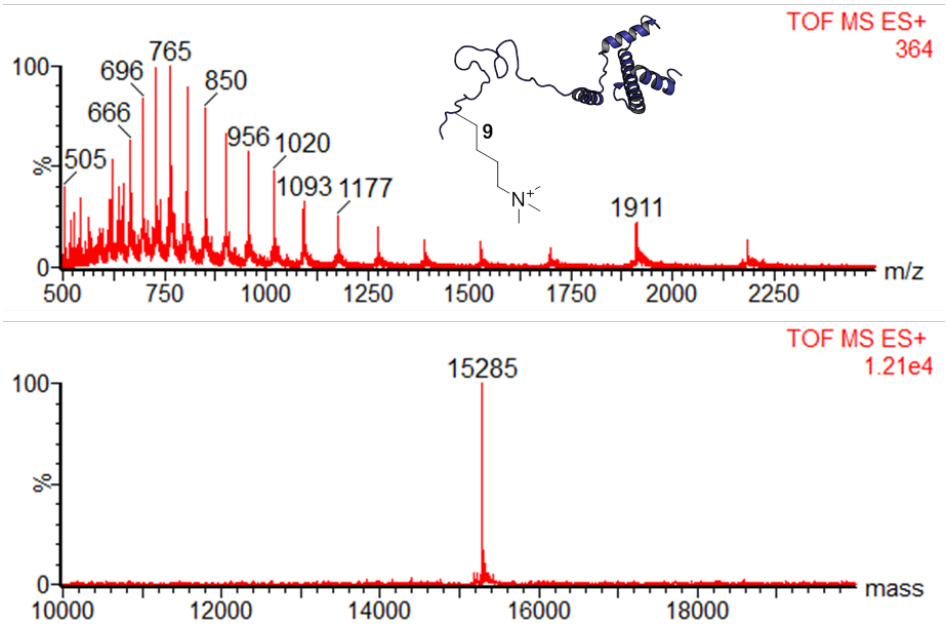
The entire volume (~200 µL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification.

The reaction was shaken at 4°C for one hour before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

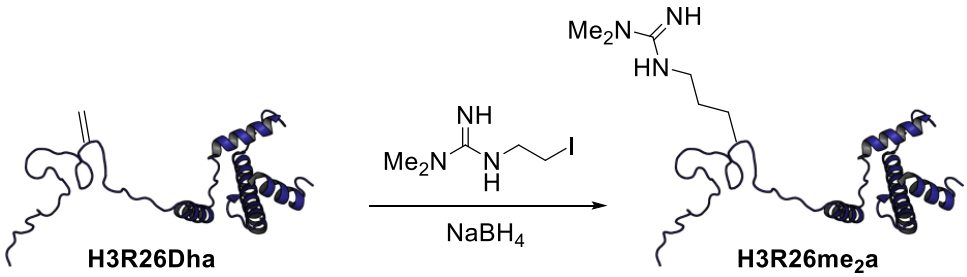
ESI-MS for the modified histone is shown below.

Calculated mass: 15284.9

Observed mass: 15285



Installation of asymmetric dimethylarginine from dehydroalanine H3 R26

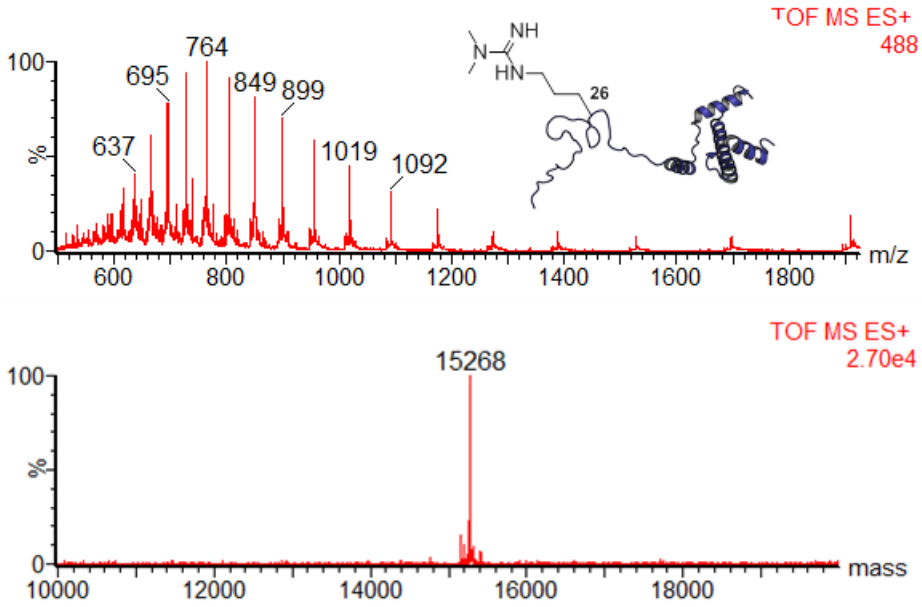


Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3R26Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer)

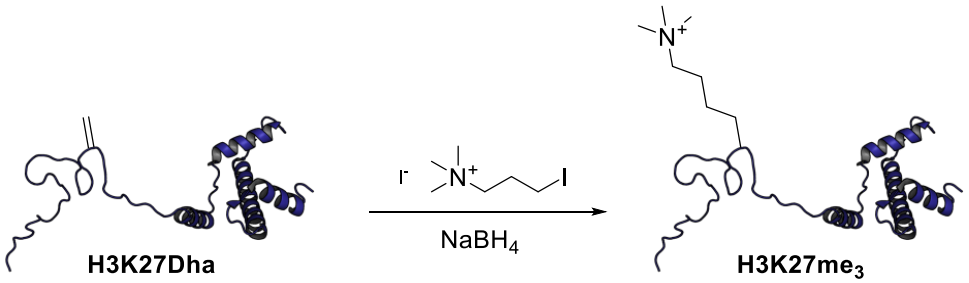
to give a final protein concentration of 1 mg/mL (~66 μM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was used to dissolve 5 mg solid 3-(2-iodoethyl)-1,1-dimethylguanidine **S17**. The resultant pale yellow solution was gently mixed by pipette.
The entire volume (~200 μL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification and decolouration. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below.
Calculated mass: 15268.9 (H3R26me$_2$ C110A)
Observed mass: 15268



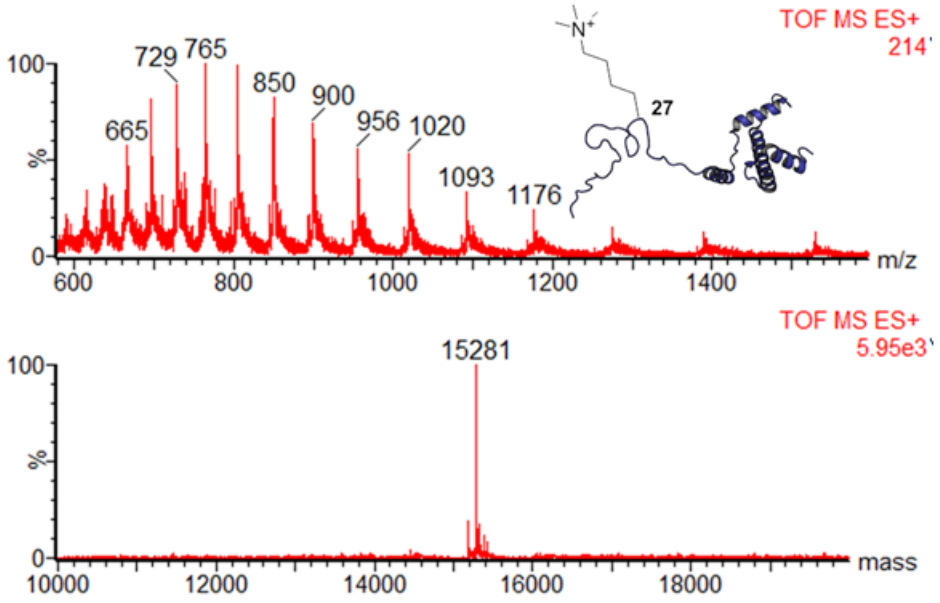
Installation of trimethyllysine H3 K27$_{me3}$



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3K27Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was used to dissolve 8 mg (3-iodopropyl)trimethylamine hydroiodide **S6**. The resultant pale solution was gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then rapidly added via pipette to a separate 2 mL eppendorf tube, kept at 4°C and containing 0.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution, followed by a rapid clarification. The reaction was shaken at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

ESI-MS for the modified histone is shown below:
Calculated mass: 15282.8 (H3 K27me$_3$, C110A)
Observed mass: 15281



Nucleosome reconstitution

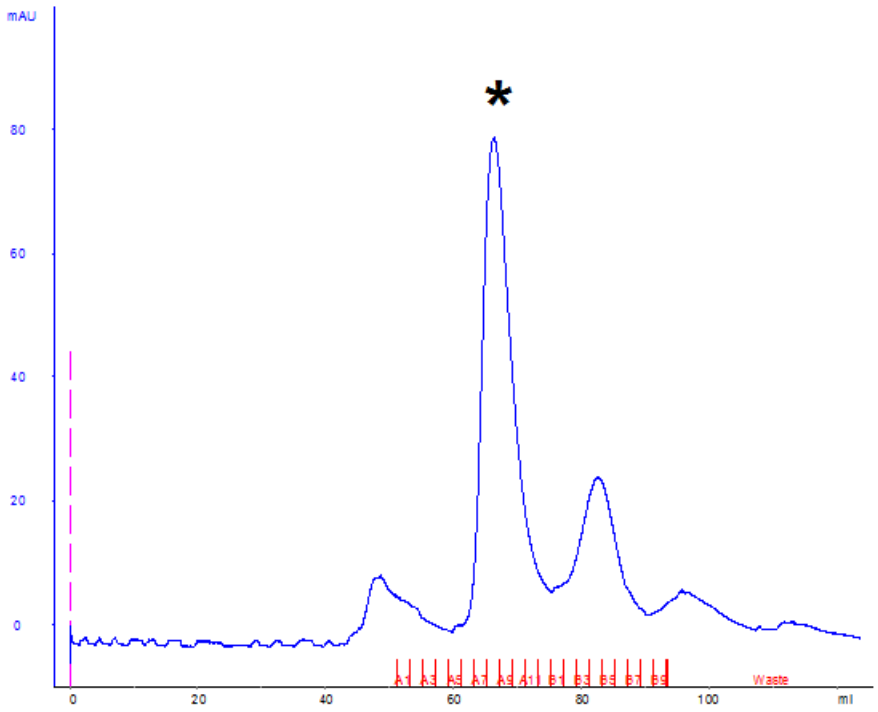
Starting from the modified histone proteins obtained post chemical mutagenesis reaction, modified nucleosomes were reconstituted in a step-wise manner as described earlier[61]. Briefly, the modified H3 protein (or wt H3 protein) along with other canonical wt histone proteins (H2A, H2B, and H4), in a molar ratio of 1.1, 1.1 and 1 respectively, were dissolved in unfolding buffer (7 M Gd.HCl, 50 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT) to a final concentration of 1-1.5 mg/mL. The protein mixture was dialyzed against the refolding buffer (2 M NaCl, 10 mM Tris pH 7.5, 1 mM DTT, 5 mM β-ME) thrice (for at least 3 hours each time). The resulting octamers were purified and analyzed by gel filtration chromatography (HiLoad 16/600 Superdex 200 column, (GE Healthcare Life Sciences)) in refolding buffer. The purified octamers were concentrated and stored in 50 % glycerol at –20 °C for further use.

The nucleosomes were reconstituted using either a 145 bp '601' DNA or a biotinylated 186 bp '601' DNA following salt gradient dialysis method. The DNAs were prepared as described[77]. Wt/modified octamer were combined with DNA in a molar ratio of 1:0.85 in refolding buffer. The samples were dialyzed against a series of step-gradients from high salt buffer (2 M KCl, 10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT) to low-salt buffer (250 mM KCl,

10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT). The reconstituted nucleosomes were analyzed by Native PAGE using Novex TBE Gels (6%, Invitrogen). The nucleosomes were stored at 4 °C for further use.

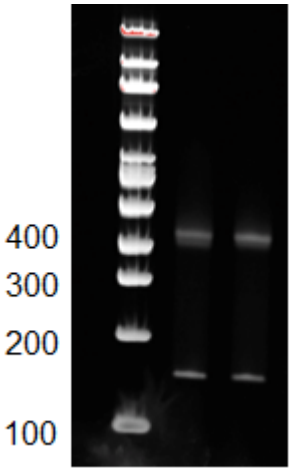
H3K9me$_3$(^{13}C) nucleosome formation

SEC trace of octamer assembly



***= H3K9me$_3$ ^{13}C** –bearing octamer

6% TBE gel analysis of reconstituted nucleosomes



Lanes represent nucleosomes in duplicate

KDM4a (JMJD2A, catalytic domain) protein expression and purification

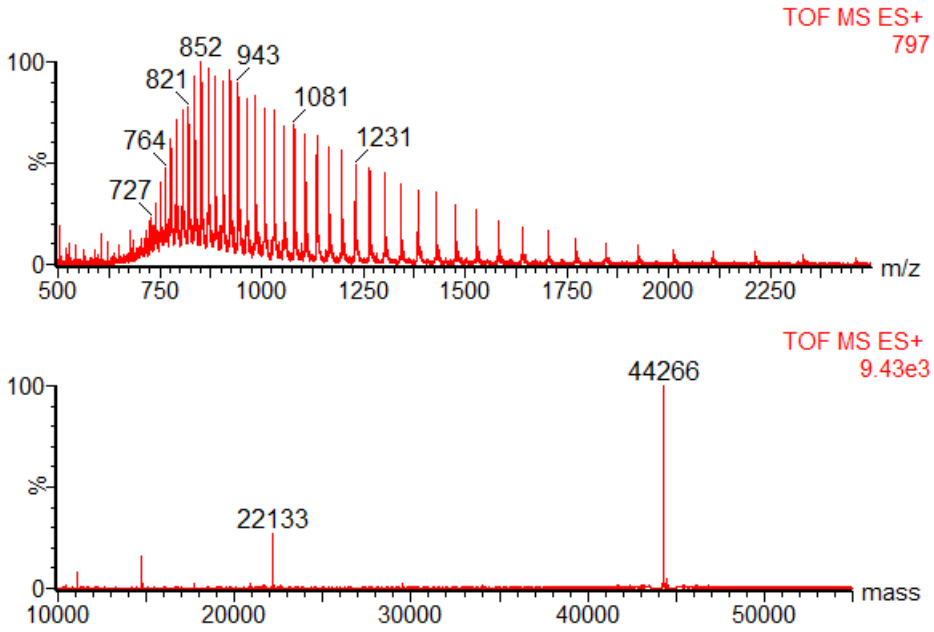
An expression plasmid (pNIC28-BSA4), containing a N-terminally His-tagged construct of the catalytic domain (1-359) of human KDM4a/JMJD2a, was transformed into *E. coli* BL21 (DE3) cells. Four litres of TB media containing 50 μg ml^{-1} kanamycin were inoculated with 10 ml L^{-1} overnight culture and grown at 37 °C. Expression was induced with 0.2 mM IPTG at A600 = 0.8 and incubatedfor 18 h at 18 °C. The cells were collected by centrifugation and frozen at -80 °C.

Frozen cell pellets were thawed and re-suspended in a total volume of 100 ml lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, DNAse, Complete protease inhibitor –EDTA tablets). Cells were disrupted by sonication and cell debris was removed by centrifugation. Recombinant JMJD2A was purified by Ni-affinity (5 ml HisTrap HP) and size-exclusion chromatography (120 ml HiLoad 16/60 Superdex 200) on an AKTA system. The supernatant was loaded onto the Ni-affinity column, washed with wash buffer (lysis buffer with 40 mM imidazole), and then eluted in a gradient with wash buffer containing 250 mM imidazole. Pooled fractions identified by SDS-PAGE were concentrated and loaded onto the gel filtration column, previously equilibrated in 50 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.5 mM TCEP.

Finally, the protein was exchanged into 10 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP, concentrated to 10 mg ml^{-1}, flash frozen in liquid N2 and stored at -80 °C.

ESI-MS for the purified demethylase is shown below

Observed mass: 44266

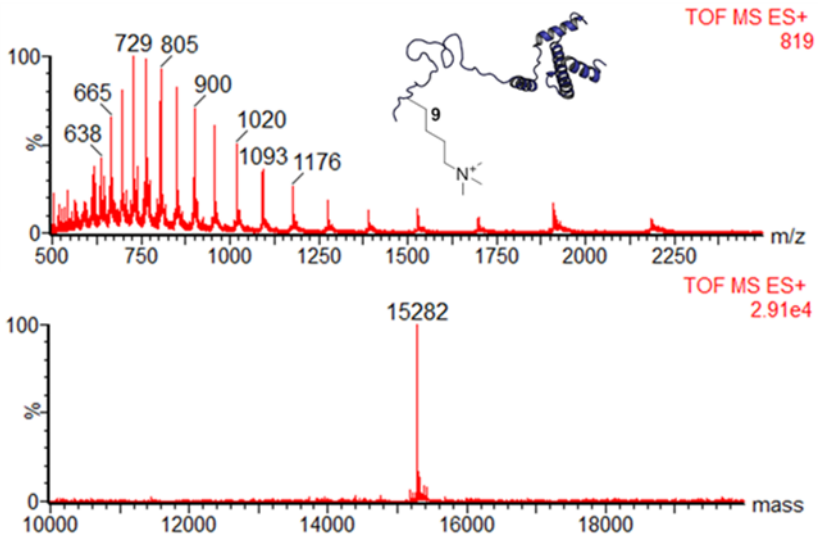


Demethylation of H3K9me$_3$ (^{12}C)

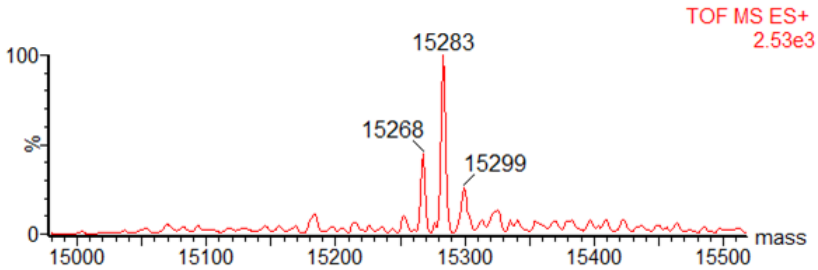
Demethylation of H3K9me$_3$ (^{12}C) was performed by incubating H3K9me$_3$ (15 µL of a 2.04 mg/mL (A280) solution in HEPES pH 7.5, giving a final histone concentration of 20 µM) with α-ketoglutarate (10 µL of a 10x stock, final concentration 50 µM), ascorbate (10 µL of a 10x stock, final concentration 100 µM) and $Fe(NH_4)_2(SO_4)_2$ (10 µL of a 10x stock, final concentration 10 µM). KDM4a catalytic domain (7.8 µL of a 1.17 mg/mL stock, giving a final concentration of 2 µM) was added and the solution adjusted to a final volume of 100 µL with HEPES pH 7.5. The demethylation reaction was shaken gently at 30 °C and aliquots taken at regular intervals for LC-MS analysis.

Example spectra:

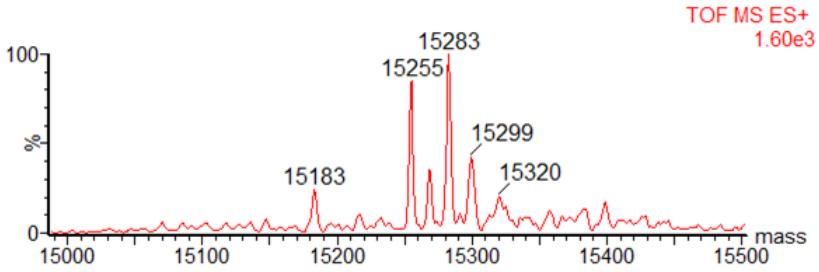
Starting material:



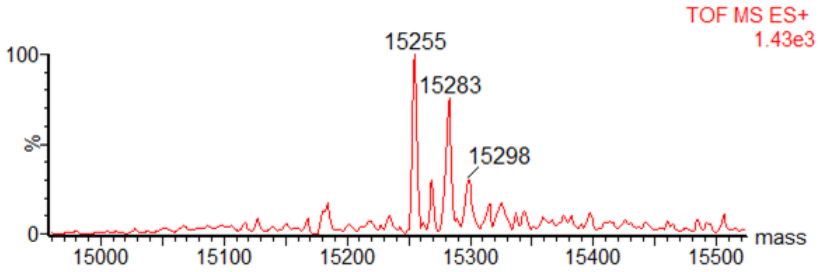
t = 1 minute



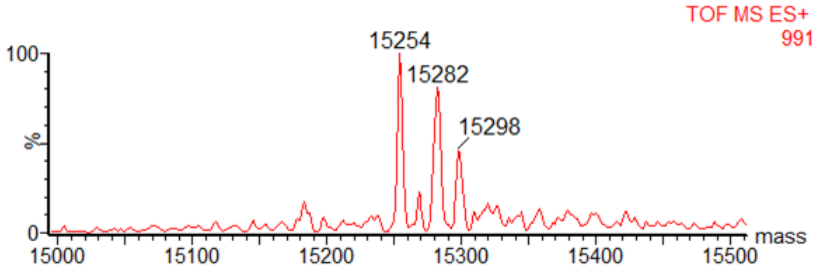
t = 2 minutes



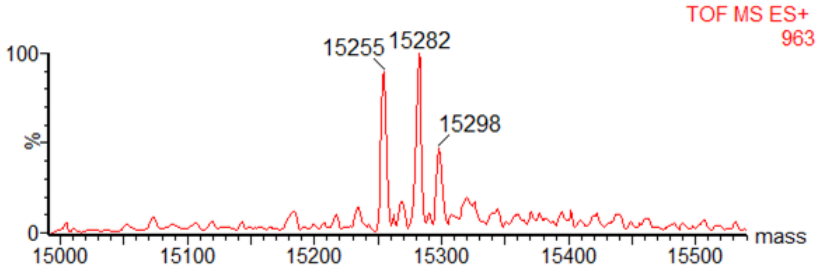
t = 4 minutes



t = 6 minutes



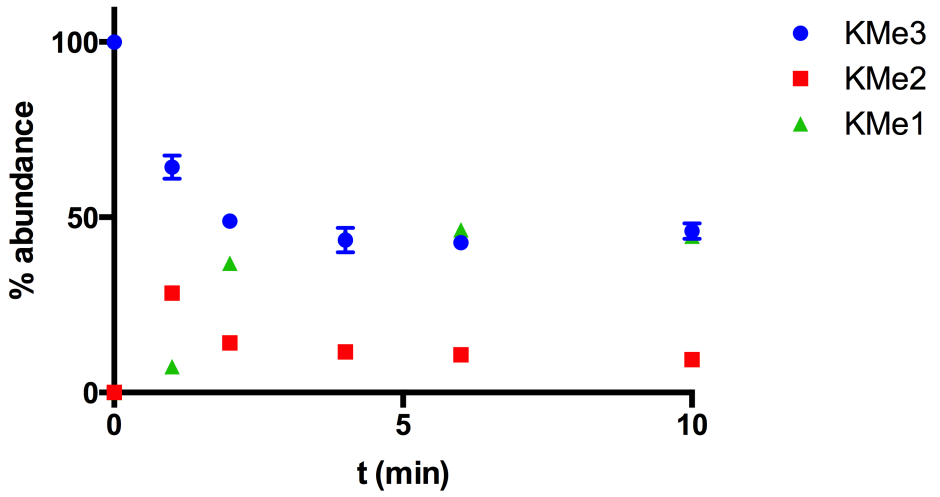
t = 10 minutes



Calculated masses: 15283 (H3 K9me$_3$), 15268 (H3 K9me$_2$), 15254 (H3 K9me)

Observed masses: 15282, 15268, 15255

The above data + 2 further replicates were plotted in graphpad Prism as the mean(+/- s.e.m) **(Figure 4, main text):**



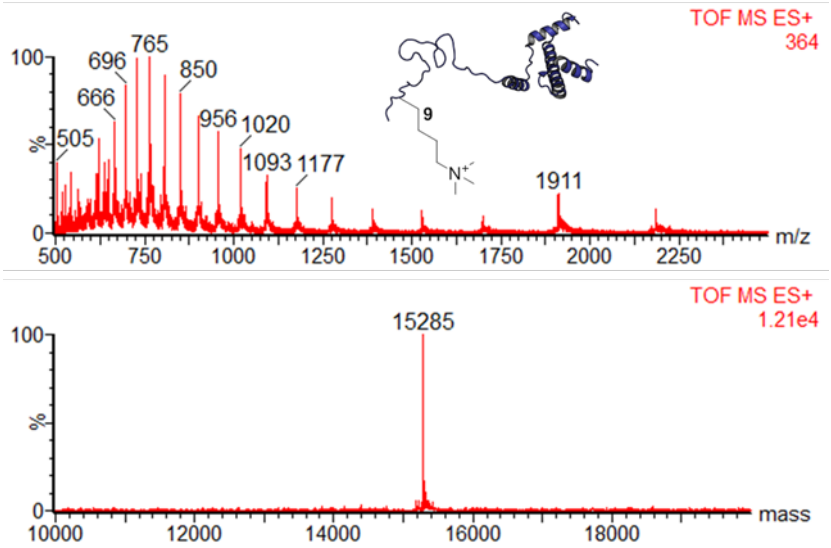
Note: Even with application of 'forcing' conditions (additional enzyme, substrate, heating), conversion of Kme_3 does not proceed beyond ~50%. We take this to indicate the inability of the substrate to either recognize or demethylate the D-epimer of the protein resultant from chemical mutation. In contrast, in western blot experiments on nucleosomes (see below) 100% of the signal resulting from Kme_3 is lost upon demethylation, which we take to indicate that the supplied antibody does not reconise the unnatural epimer.

Demethylation of H3K9me$_3$ (^{13}C)

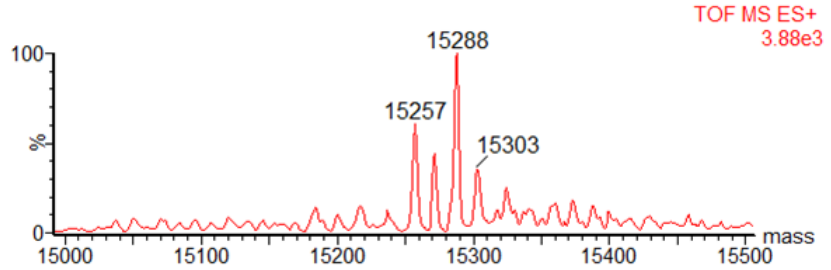
Demethylation of H3K9me$_3$ (^{13}C) was performed by incubating H3K9me$_3$ (18.52 µL of a 1.65 mg/mL (A280) solution in HEPES pH 7.5, giving a final histone concentration of 20 µM) with α-ketoglutarate (10 µL of a 10x stock, final concentration 50 µM), ascorbate (10 µL of a 10x stock, final concentration 100 µM) and $Fe(NH_4)_2(SO_4)_2$ (10 µL of a 10x stock, final concentration 10 µM). KDM4a catalytic domain (7.8 µL of a 1.17 mg/mL stock, giving a final concentration of 2 µM) was added and the solution adjusted to a final volume of 100 µL with HEPES pH 7.5. The demethylation reaction was shaken gently at 30 °C and aliquots taken at regular intervals for LC-MS analysis.

Example spectra:

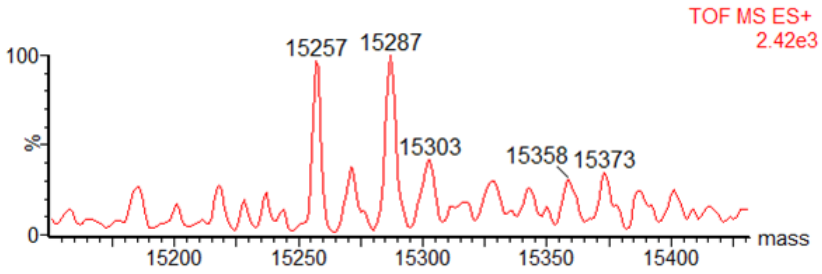
Starting material:



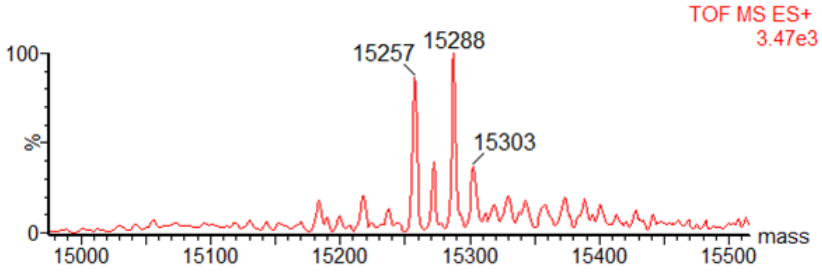
t = 1 minute



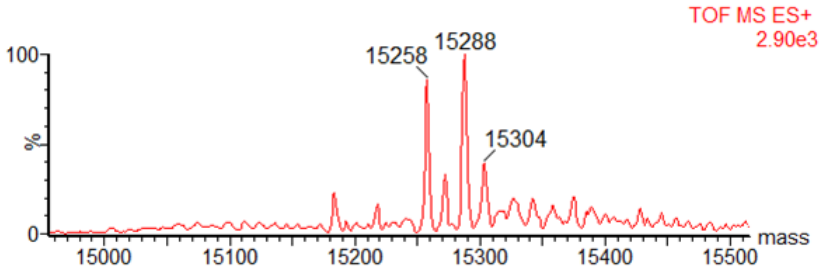
t = 2 minutes



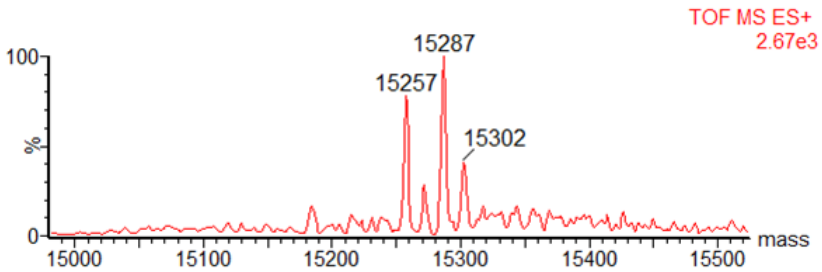
t = 4 minutes



t = 6 minutes



t = 10 minutes



^{13}C demethylation experiments

Final assay mixture contained:

125 μM H3K9me$_3$ (^{13}C)
180 μM 2-oxoglutarate
9 μM $Fe(NH_4)_2(SO_4)_2$
90 μM sodium ascorbate
3.8 μM KDM4a

Buffer: 50 mM ammonium formate pH 7.5

Experimental conditions

NMR experiments were conducted using a Aglient/Varian DD2 spectrometer (14.1 T/ 600 MHz) equipped with a 5 mm room-temperature triple resonance probe and carried out at 298 K. Experiments were carried out with 5 mm diameter NMR tubes (Norrell).The demethylation reaction was followed by back-to-back ^{13}C-edited 1D ^{1}H spectra, with each 1D spectrum consisting of 32 scans. Each spectrum employed a pre-scan delay of 1 s and acquisition time of 64 ms during which ^{13}C GARP decoupling was applied, giving an experimental time of approximately 35 s per 1D spectrum.

2D ^{13}C-^{1}H HSQC spectra were also obtained before and after the reaction. These employed 30 increments and 8 scans per increment, an acquisition time of 0.064 ms during which ^{13}C GARP decoupling was applied and a pre-scan delay of 1s, giving a total experimental time of approximately 10 minutes per 2D spectrum.

Demethylation of H3K9me$_3$ (^{13}C) nucleosomes

Nucleosome demethylation was performed by incubating H3 K9me$_3$ – containing nucleosomes (1.5 μL of a 7.25 μM solution in HEPES pH 7.5, final concentration 300 nM, added last) with α-ketoglutarate (1 mM), ascorbate (1 mM), KDM4a (6.4 μL of a 626 μM stock, giving a final concentration of 200 μM), and $Fe(NH_4)_2(SO_4)_2$ (400 μM) in a final assay volume of 20 μL (50 mM HEPES pH 7.5) at 37 °C overnight. A control reaction was performed in parallel with the same solution composition in the absence of KDM4a.

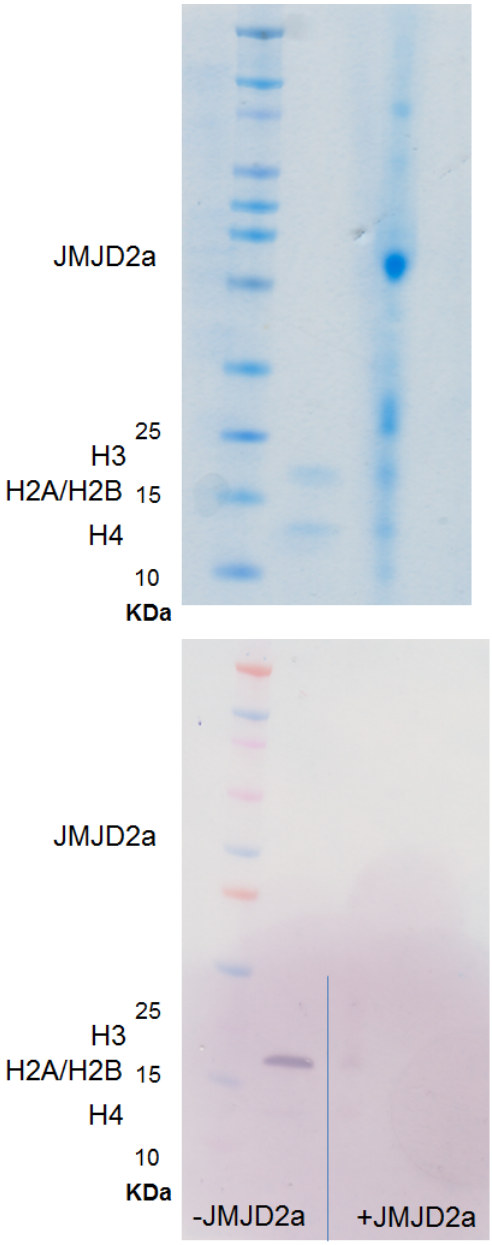
Reactions were quenched using 6× SDS sample loading buffer and boiled at 100°C for 2 min, before loading onto SDS-PAGE for analysis.

Western blot:

Two aliquots from each of the reactions prepared above were subjected to SDS-PAGE using a 12% NuPAGE pre-cast gel (Invitrogen) and MES running buffer (Invitrogen). One aliquot from each reaction was used for SDS-PAGE with Coomassie detection, the other was used for western blot. For the western blot, the protein was transferred to a PVDF membrane using an iBlot transfer apparatus (Invitrogen). The membrane was blocked with 30 mg/mL BSA in TBST buffer for 30 minutes at room temperature with rocking. After blocking, the membrane was washed with TBST buffer (5 ×15 mL). The primary antibody (anti-H3-K9Me3, rabbit polyclonal, Abcam ab8898) was diluted from 10 µL to 10 mL in TBST buffer.

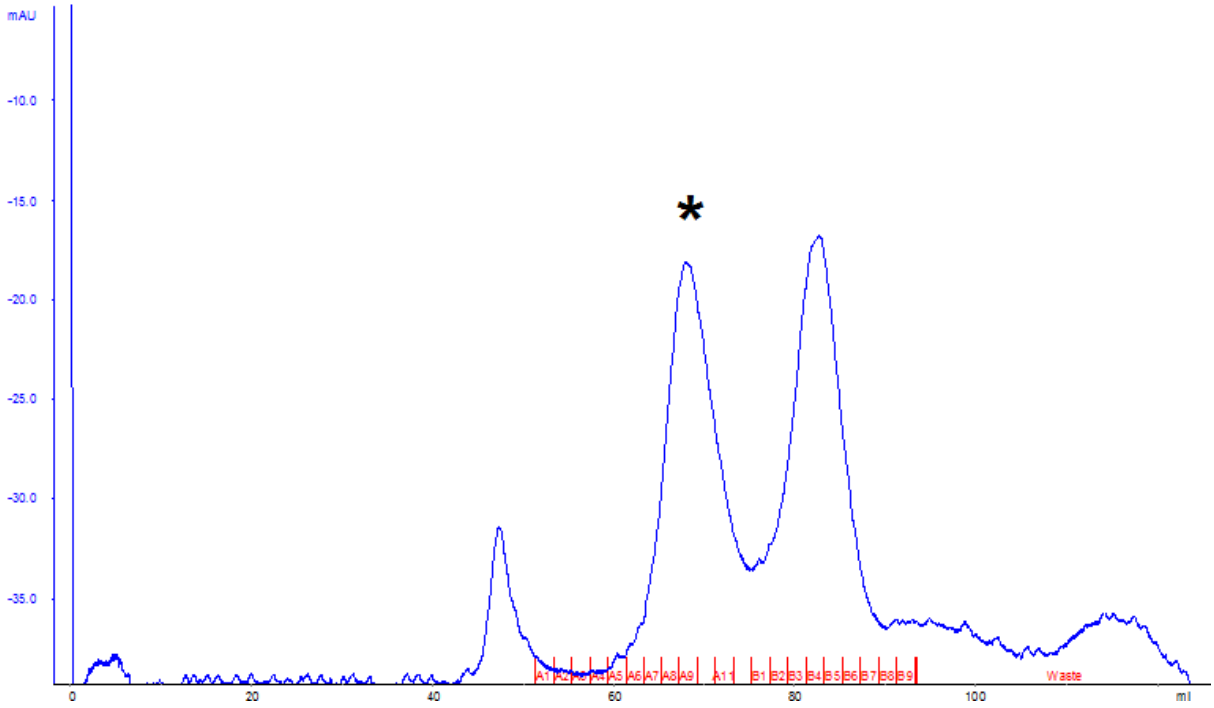
The membrane was then incubated with the primary antibody solution for 1 hour at room temperature with rocking. The membrane was then washed with TBST buffer (5 ×15 mL). The secondary antibody (anti-rabbit IgG produced in goat, conjugated with alkaline phosphatase, Sigma-Aldrich A3687) was diluted from 5 µL to 10 mL in TBST buffer. The membrane was incubated with the secondary antibody solution for 30 minutes at room temperature with rocking. The membrane was washed thoroughly with TBST buffer (10 ×15 mL) and then water (5 ×15 mL). The membrane was then developed using BCIP/NBT liquid substrate system (Sigma-Aldrich B1911). A clear band was observed for trimethylated histone from the control reaction (no demethylase), whereas no band was observed for the treated sample, indicating quantitative demethylation.

The SDS-PAGE gel with coomassie staining and western blot are shown below.


JMJD2a
25
H3
H2A/H2B
15
H4
10
KDa
JMJD2a
25
H3
H2A/H2B
15
H4
10
KDa
-JMJD2a
+JMJD2a

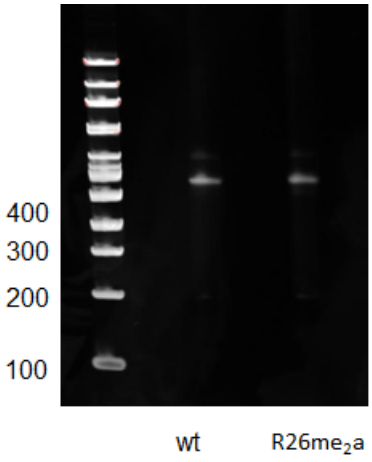
H3R26me$_2$a nucleosome formation

SEC trace of octamer assembly



***= H3R26me$_2$a – bearing nucleosomes**

6% TBE gel analysis of reconstituted nucleosomes



Interaction proteomics experiment

Affinity enrichment proteomics was performed as described[61]. 16 µg of each nucleosome were immobilized separately on 100 µL of magnetic streptavidin beads (Dynabeads® Streptavidin MyOne™ T1, Life Technologies) in the reconstitution buffer (10 mM Tris pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT with 0.1 % NP-40). Immobilized nucleosomes were washed with the binding buffer (20 mM HEPES pH 7.9, 150 mM NaCl, 0.2 mM EDTA, 20% Glycerol, 0.1% NP-40, 1 mM DTT supplemented with complete protease and phosphatase inhibitors) and then incubated with 1 mL of HeLa nuclear extract (0.5 mg/mL in binding buffer) for 4 hours at 4 °C. After 4 hours, non-specifically bound proteins were removed by washing with binding buffer (1 mL X 5 times). The interacting protein partners were eluted in 100 µL of 8 M Urea, 2 M NaCl.

The interacting protein partners were proteolytically digested using trypsin following Filter-Aided Sample Preparation (FASP) protocol[87]. The resulting peptides mixture were analyzed on an EASY-nLC 1000 UHPLC system (Proxeon) connected to a Q Exactive mass spectrometer (Thermo Fischer Scientific). Peptides were trapped on an in-house packed trap column (75 mm i.d. x 20 mm, reprosil C18, 3 µm, 120 Å) and separated on an EASY-spray Acclaim PepMap analytical column (75 mm i.d. x 500 mm, RSLC C18, 2 µm, 100 Å). Samples were loaded at a pressure of 500 bar with 100% solvent A (0.1% formic acid in water) and the peptides separated by a linear gradient (length: 120 min, 8–28% solvent B (0.1% formic acid in acetonitrile), flow rate: 200 nl/min). The separated peptides were electrosprayed directly into the Q Exactive operated in a data-dependent mode through an EASY-Spray nanoelectrospray ion source. Full-scan MS spectra were acquired in the Orbitrap (scan range 350–1500 m/z, resolution 70,000, AGC target 3e6, maximum injection time 100 ms). After the MS scans, the 20 most intense peaks were selected for HCD fragmentation at 30% of normalized collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17,500, AGC target 5e4, maximum injection time 120 ms).

The raw data files generated were processed using MaxQuant software (Version 1.4.3.17), integrated with Andromeda search engine[88,89]. The peak lists were searched against human protein database (Swiss-Prot, version 04/13) as well as list of common contaminants. Trypsin with a maximum number of 4 missed cleavages was selected as the protease. Acetylation (Protein N-term) and Oxidation (Methionine) were used as variable modifications while Carbamidomethylation (Cysteine) was set as a fixed modification. Protein and PSM false discovery rate (FDR) were set at 0.01. Label free quantification (LFQ) was performed using the in-built algorithm in the MaxQuant[82].Further downstream data analysis was performed using Perseus software (Version 1.5.0.18). LFQ intensity of all identified protein groups was uploaded; contaminants, reverse and protein groups only identified by site were removed. Protein groups identified in at least 4 replicate experiments among 6 replicate experiments of either bait (modified) or control (wt) were used. LFQ intensities were logarithmized and missing LFQ intensities were replaced by the average of minimum LFQ intensity observed among each experiment. False discovery rate (FDR) based paired t-test (FDR = 0.05, S0 = 1) was used to generate fold-change difference as well as a *p* value, in turn annotating significant interacting protein partners. The data was visualised in a volcano plot of ratio change versus log (*p* value)

References: Chapter 5

1 Ambler, R. ε-N-Methyl-lysine in bacterial flagellar protein. *Nature* **184**, 56-57 (1959).

2 Lee, D. Y., Teyssier, C., Strahl, B. D. & Stallcup, M. R. Role of protein methylation in regulation of transcription. *Endocrine reviews* **26**, 147-170 (2005).

3 Stallcup, M. R. Role of protein methylation in chromatin remodeling and transcriptional regulation. *Oncogene* **20**, 3014 (2001).

4 Huang, J. & Berger, S. L. The emerging field of dynamic lysine methylation of non-histone proteins. *Current opinion in genetics & development* **18**, 152-158 (2008).

5 Shi, Y. & Whetstine, J. R. Dynamic regulation of histone lysine methylation by demethylases. *Molecular cell* **25**, 1-14 (2007).

6 Seet, B. T., Dikic, I., Zhou, M.-M. & Pawson, T. Reading protein modifications with interaction domains. *Nature reviews Molecular cell biology* **7**, 473-483 (2006).

7 Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* **25**, 15-30 (2007).

8 Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* **13**, 343-357 (2012).

9 Luger, K. & Richmond, T. J. The histone tails of the nucleosome. *Current opinion in genetics & development* **8**, 140-146 (1998).

10 Gordon, F., Luger, K. & Hansen, J. C. The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. *Journal of Biological Chemistry* **280**, 33701-33706 (2005).

11 Woodcock, C. L. & Ghosh, R. P. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2**, a000596 (2010).

12 Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Current opinion in genetics & development* **21**, 175-186 (2011).

13 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381-395 (2011).

14 Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology* **6**, 227 (2005).

15 Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature reviews Molecular cell biology* **6**, 838-849 (2005).

16 Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Molecular cell* **48**, 491-507 (2012).

17 Kubicek, S. *et al.* Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Molecular cell* **25**, 473-481 (2007).

18 Trojer, P. & Reinberg, D. Histone lysine demethylases and their impact on epigenetics. *Cell* **125**, 213-217 (2006).

19 Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-481 (2006).

20 Shi, Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nature reviews genetics* **8**, 829-833 (2007).

21 Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development* **17**, 1870-1881 (2003).

22 Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology* **14**, 1025-1040 (2007).

23 Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-980 (2010).

24 Canzio, D. *et al.* Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Molecular cell* **41**, 67-81 (2011).

25 Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87-91 (2007).

26 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080 (2001).

27 Garcia, B. A., Shabanowitz, J. & Hunt, D. F. Characterization of histones and their post-translational modifications by mass spectrometry. *Current opinion in chemical biology* **11**, 66-73 (2007).

28 Graur, D. *et al.* On the immortality of television sets:"function" in the human genome according to the evolution-free gospel of ENCODE. *Genome biology and evolution* **5**, 578-590 (2013).

29 Fierz, B. & Muir, T. W. Chromatin as an expansive canvas for chemical biology. *Nature chemical biology* **8**, 417-427 (2012).

30 Müller, M. M. & Muir, T. W. Histones: at the crossroads of peptide and protein chemistry. *Chemical reviews* **115**, 2296 (2015).

31 Shimko, J. C., North, J. A., Bruns, A. N., Poirier, M. G. & Ottesen, J. J. Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein binding within nucleosomes. *Journal of molecular biology* **408**, 187-204 (2011).

32 Li, J. *et al.* One-pot native chemical ligation of peptide hydrazides enables total synthesis of modified histones. *Organic & biomolecular chemistry* **12**, 5435-5441 (2014).

33 Siman, P., Karthikeyan, S. V., Nikolov, M., Fischle, W. & Brik, A. Convergent chemical synthesis of histone H2B protein for the site‐specific ubiquitination at Lys34. *Angewandte Chemie International Edition* **52**, 8059-8063 (2013).

34 Seenaiah, M., Jbara, M., Mali, S. M. & Brik, A. Convergent versus sequential protein synthesis: the case of ubiquitinated and glycosylated H2B. *Angewandte Chemie International Edition* **54**, 12374-12378 (2015).

35 Simon, M. D. *et al.* Rapid Flow‐Based Peptide Synthesis. *ChemBioChem* **15**, 713-720 (2014).

36 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**, 6705-6710 (1998).

37 Nguyen, U. T. *et al.* Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. *Nature methods* **11**, 834-840 (2014).

38 Li, W.-T. *et al.* Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *Journal of molecular biology* **385**, 1156-1164 (2009).

39 Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically Encoding N ϵ-Methyl-l-lysine in Recombinant Histones. *Journal of the American Chemical Society* **131**, 14194-14195 (2009).

40 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).

41 Wang, Y.-S. *et al.* A genetically encoded photocaged Nε-methyl-l-lysine. *Molecular bioSystems* **6**, 1557-1560 (2010).

42 Groff, D., Chen, P. R., Peters, F. B. & Schultz, P. G. A Genetically Encoded ε‐N‐Methyl Lysine in Mammalian Cells. *ChemBioChem* **11**, 1066-1068 (2010).

43 Chalker, J. M., Bernardes, G. a. J. & Davis, B. G. A "tag-and-modify" approach to site-selective protein modification. *Accounts of chemical research* **44**, 730-741 (2011).

44 Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003-1012 (2007).

45 Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post‐Translational Modifications. *Angewandte Chemie International Edition* **51**, 1835-1839 (2012).

46 Hajduk, P. J. *et al.* NMR-based screening of proteins containing 13C-labeled methyl groups. *Journal of the American Chemical Society* **122**, 7898-7904 (2000).

47 Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *Journal of biomolecular NMR* **20**, 71-75 (2001).

48 Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature protocols* **1**, 749-754 (2006).

49 Kerfah, R., Plevin, M. J., Sounier, R., Gans, P. & Boisbouvier, J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Current opinion in structural biology* **32**, 113-122 (2015).

50 Jones, D. H. *et al.* Site-specific labeling of proteins with NMR-active unnatural amino acids. *Journal of biomolecular NMR* **46**, 89 (2010).

51 Ong, S.-E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* **1**, 2650-2660 (2006).

52 Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. & Schofield, C. J. Monitoring the Activity of 2‐Oxoglutarate Dependent Histone Demethylases by NMR Spectroscopy: Direct Observation of Formaldehyde. *ChemBioChem* **11**, 506-510 (2010).

53 Liokatis, S., Dose, A., Schwarzer, D. & Selenko, P. Simultaneous detection of protein phosphorylation and acetylation by high-resolution NMR spectroscopy. *Journal of the American Chemical Society* **132**, 14704-14705 (2010).

54 Theillet, F. o.-X. *et al.* Site-specific mapping and time-resolved monitoring of lysine methylation by high-resolution NMR spectroscopy. *Journal of the American Chemical Society* **134**, 7616-7619 (2012).

55 Stützer, A. *et al.* Modulations of DNA contacts by linker histones and post-translational modifications determine the mobility and modifiability of nucleosomal H3 tails. *Molecular cell* **61**, 247-259 (2016).

56 Frederick, K. K. *et al.* Sensitivity-enhanced NMR reveals alterations in protein structure by cellular milieus. *Cell* **163**, 620-628 (2015).

57 Thongwichian, R. *et al.* A multiplexed NMR-reporter approach to measure cellular kinase and phosphatase activities in real-time. *Journal of the American Chemical Society* **137**, 6468-6471 (2015).

58 Freedberg, D. I. & Selenko, P. Live cell NMR. *Annual review of biophysics* **43**, 171-192 (2014).

59 Williams, S. T. *et al.* Studies on the catalytic domains of multiple JmjC oxygenases using peptide substrates. *Epigenetics* **9**, 1596-1603 (2014).

60 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Chromatin Protocols*, 1-16 (1999).

61 Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nature communications* **6** (2015).

62 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).

63 Shiau, C. *et al.* Reconstitution of nucleosome demethylation and catalytic properties of a Jumonji histone demethylase. *Chemistry & biology* **20**, 494-499 (2013).

64 Baldwin, G. & Carnegie, P. Specific enzymic methylation of an arginine in the experimental allergic encephalomyelitis protein from human myelin. *Science* **171**, 579-581 (1971).

65 Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Molecular cell* **33**, 1-13 (2009).

66 Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. *Nature Reviews Cancer* **13**, 37-50 (2013).

67 Le, D. D., Cortesi, A. T., Myers, S. A., Burlingame, A. L. & Fujimori, D. G. Site-specific and regiospecific installation of methylarginine analogues into recombinant histones and insights into effector protein binding. *Journal of the American Chemical Society* **135**, 2879 (2013).

68 Chen, D. *et al.* Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174-2177 (1999).

69 Ma, H. *et al.* Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Current Biology* **11**, 1981-1985 (2001).

70 Schurter, B. T. *et al.* Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry* **40**, 5747-5756 (2001).

71 Jacques, S. L. *et al.* Carm1 preferentially methylates h3r17 over h3r26 through a random kinetic mechanism. *Biochemistry* **55**, 1635-1644 (2016).

72 Wei, H., Mundade, R., Lange, K. & Lu, T. Protein arginine methylation of non-histone proteins and its role in diseases. *Cell Cycle* **13**, 32-41 (2014).

73 Xu, W. *et al.* A methylation-mediator complex in hormone signaling. *Genes & development* **18**, 144-156 (2004).

74 Chen, D., Huang, S.-M. & Stallcup, M. R. Synergistic, p160 coactivator-dependent enhancement of estrogen receptor function by CARM1 and p300. *Journal of Biological Chemistry* **275**, 40810-40816 (2000).

75 Daujat, S. *et al.* Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Current Biology* **12**, 2090-2097 (2002).

76 Torres-Padilla, M.-E., Parfitt, D.-E., Kouzarides, T. & Zernicka-Goetz, M. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445**, 214-218 (2007).

77 Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-484 (2010).

78 Chang, B., Chen, Y., Zhao, Y. & Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science* **318**, 444-447 (2007).

79 Webby, C. J. *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90-93 (2009).

80 Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nature communications* **7** (2016).

81 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

82 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics* **13**, 2513-2526 (2014).

83 Saksouk, N. *et al.* Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Molecular cell* **56**, 580-594 (2014).

84 Khan, A. *et al.* BEND3 represses rDNA transcription by stabilizing a NoRC component via USP21 deubiquitinase. *Proceedings of the National Academy of Sciences* **112**, 8338-8343 (2015).

85 Abhiman, S., Iyer, L. M. & Aravind, L. BEN: a novel domain in chromatin factors and DNA viral proteins. *Bioinformatics* **24**, 458-461 (2008).

86 Chalker, J. *Reaction Engineering for Protein Modification: Tools for Chemistry and Biology.* DPhil Thesis thesis, University of Oxford, (2011).

87 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature methods* **6**, 359 (2009).

88 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).

89 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805 (2011).

Chapter 6: Post-translational mutagenesis for protein phosphorylation

Introduction

The phosphorylation of proteins is a regulatory mechanism employed by all domains of life[1]. Transfer of the phosphoryl moiety to an acceptor residue by a kinase, and its removal by a complementary phosphatase, allows cycles of signaling that underlie many of the key processes of biology[2]. Phosphorylation of residues (including, but not limited to*, serine, threonine and tyrosine, **Figure 6.1**) within a protein adds negative charge[3], increases residue size and installs new hydrogen bond acceptors[4-6]. These alterations in side-chain composition can alter local protein structure, global folding and protein dynamics and often lead to a conformational 'switch'[7] and/or change in binding partner interactions[8,9]. On a global level, phosphorylation events are the molecular signals underlying cell development, growth and division[10]. Aberrant kinase signalling is a hallmark of cancer[11,12], underscoring the potential therapeutic benefit to understanding the effects of this PTM.

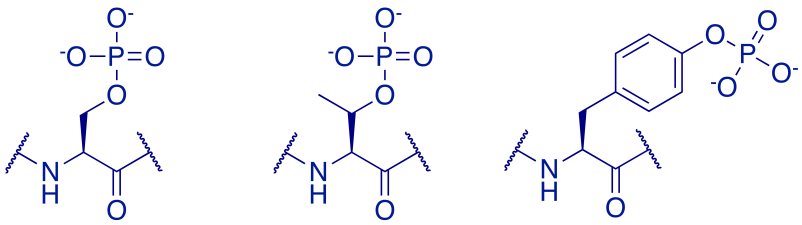


Figure 6.1: The most abundant phosphorylated amino acids in proteins

Phosphorylation has been reported on all four canonical histones and the linker histone H1[13]. Histone H3-Ser10 phosphorylation (H3-S10Ph)[14] illustrates the context-dependent and varied roles to which protein phosphorylation has been put during evolution. This modification was originally identified[15,16] as a key signal for chromatin condensation during mitosis[17] but was later shown to also play an important role in

transcriptional activation during interphase[18,19]. These seemingly contradictory effects on chromatin structure and function indicate that simplistic concepts of 1:1 correspondence between a single PTM and a biological readout are probably misguided[20]. Instead, the context of the modification (stoichiometry, presence of other PTMs, cell-cycle state, protein complex occupancy etc.) must be considered. In order to untangle the complexities of a given phosphorylation event, biophysical, molecular level understanding must therefore be combined with identification of binding partners and cellular context.

Methods for synthetic protein phosphorylation

Access to site-specifically and quantitatively phosphorylated proteins has long been a bottleneck to unraveling the biological complexity of this modification[21]. Treatment of recombinant proteins with kinases in vitro typically results in a heterogeneous mixture of differentially phosphorylated species[22]. Many studies in biology still utilize mutations to acidic residues such as aspartate and glutamate to simulate the effect of phosphorylation[23]. This is clearly inadequate from a chemist's point-of-view as the singly anionic carboxylate can only crudely approximate the dianionic phosphate, bearing not just a different charge but a different size and geometry[3]. As for other PTMs, chemical biologists have pursued total chemical synthesis, semi-synthesis and biological approaches such as non-canonical amino acid mutagenesis in order to install phosphorylated amino acids into proteins[21]. Further, the rapid turnover of phosphoryl groups in cells and lysates has motivated the development of 'phosphatase-resistant' phospho-mimetics as chemical probes for recognition of this PTM[24].

Native chemical ligation involving one or more synthetic phosphopeptide fragments has been used to access proteins by total synthesis[21]. For example, two consecutive NCL reactions were used to synthesise the p53-binding domain of the E3 ubiquitin ligase, MDM2, incorporating a phosphorylated serine residue at Ser-17[25]. Intriguingly, N-terminal phosphoamino acids have also been used in lieu of cysteine for chemical

ligation, an approach termed 'phosphate-assisted ligation'[26]. However, total protein synthesis remains limited to small proteins (<300 residues) and requires a final refolding step that can be problematic. As a consequence, semi-synthetic strategies have proven more popular for the targeted incorporation of phosphorylated amino acids into intact proteins.

The most common method for protein semi-synthesis, expressed protein ligation (EPL)[27], involves the ligation of a synthetic phosphopeptide with a recombinant protein fragment that can be of any size. EPL has been used to install site-specific phosphorylation (on serine, threonine or tyrosine) to a range of proteins, including Src kinase, casein kinase II[28] and α-synuclein[29]. Indeed, in the original 1998 report of EPL, a phosphotyrosine containing, N-terminal cysteinyl peptide was ligated with an expressed *C*-terminal thioester variant of C-terminal Src kinase (Csk), to generate Csk with an appended phosphotyrosine 'tail' (**Figure 6.2**)[27]. Characterisation of the semi-synthetic Csk-pTyr revealed an interaction between the appended pTyr-containing tail and a SH2 domain within the protein, and an unexpected increase in kinase activity. Notably, Csk (at 50 kDa, 450 residues), and many proteins that have been modified by EPL since, are beyond the reach of purely chemical approaches.

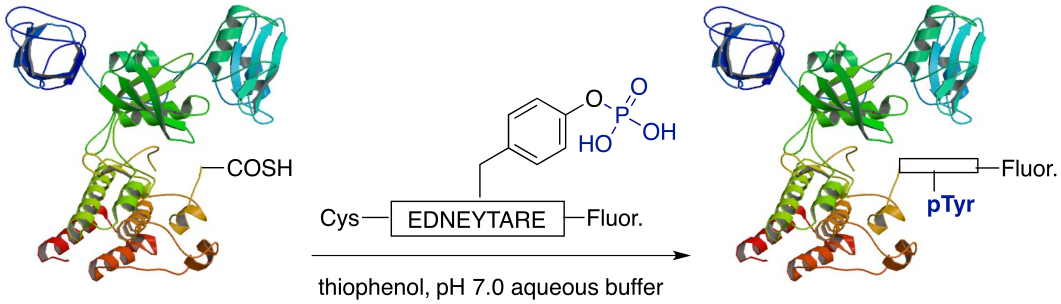


Figure 6.2: pTyr incorporation into the *C*-terminus of Csk by expressed protein ligation[27]

A key advantage of EPL relative to existing biosynthetic and 'tag-and-modify' approaches is the ability to install multiple modifications on a single synthetic peptide, limited only by the chemistry of SPPS[30]. This

attribute renders EPL the most efficient current method for introducing multiple identical modications or examining cross talk between different modifications. As a recent illustration, the Cole group reported an EPL approach to the 47 kDa phosphatase enzyme, PTEN, containing a tetraphosphorylated C-terminal tail[31]. A range of biophysical experiments using the semi-synthetic protein demonstrated that tetra-phosphorylation at this location results in a closed conformation with altered catalytic activity.

In the context of histone modifications, both phosphorylated histone H3[32], H2A[33] and H2B[34] have been prepared by EPL approaches. Semi-synthetic H3-S10Ph was further used to reconstitute nucleosome arrays as substrates for enzymatic assays, revealing a discrepancy in the acetyltransferase activity of the SAGA complex when nucleosomal, as opposed to peptide substrates, were used[32]. This study was thus one of the first to illustrate the necessity of using physiologically relevant protein substrates in such assays.

The Cole laboratory has pioneered the incorporation of unnatural, non-hydrolysable analogues of phosphorylated residues ('phospho-mimetics') by EPL[21]. In one elegant study, the enzyme serotonin *N*-acetyltransferase was prepared containing the unnatural phospho-mimetic phosphonodifluoromethylene alanine (**Pfa**) at position 205, in place of the natural PTM phosphoserine[35]. Replacement of the native O-P linkage with CF_2-P was shown to confer enhanced cellular stability upon microinjection.

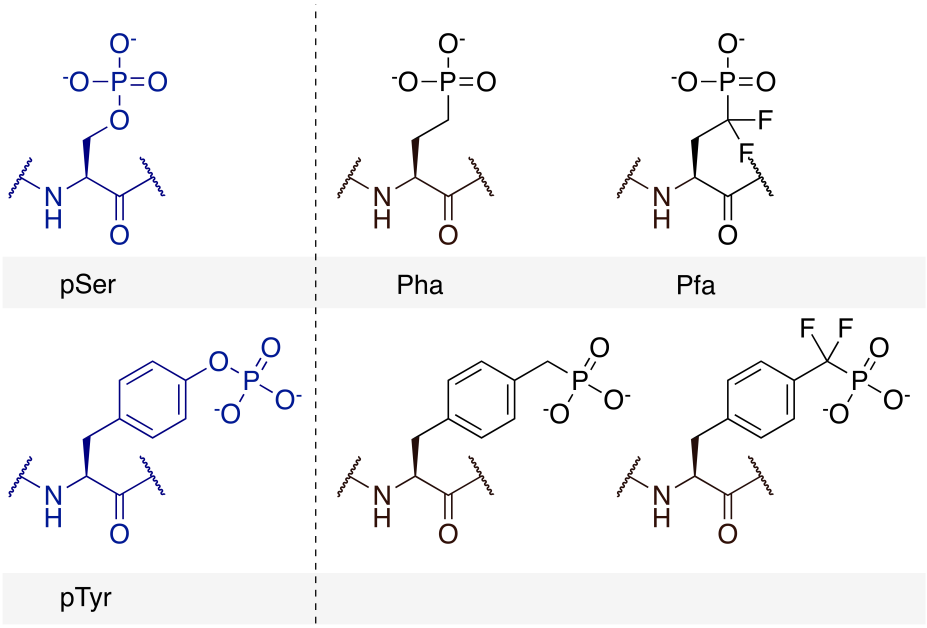


Figure 6.3: pSer, pTyr (dark blue) and their unnatural carba-mimetics (black)

'Tag-and-modify' approaches offer an alternative to ligation chemistry for protein semi-synthesis[36]. Relatively few chemistries have been applied to install phospho-mimetics via the chemoselective modification of expressed proteins[21]. The Hackenberger group has explored the use of p-azidophenylalanine (azido-Phe) as a precursor for phosphoramidate **1** as a metabolically stable analog of pTyr (**Figure 6.4a**)[37]. In their initial report the model protein SecB was first expressed with a single azido-Phe residue via amber codon suppression. A Staudinger-phosphite reaction with a photocaged, PEGylated phosphite derivative was then used to convert azido-Phe to the protected phosphoramidate **2**. UV irradiation removed the PEG chains and liberated the desired phosphoramidate **1. 1** was recognized by a pan-specific pTyr antibody, at least partially validating mimicry of the native pTyr. More recently, the Davis group[38] has reported the use of dehydroalanine as a precursor to phosphorylated cysteine pCys, itself a natural PTM. pCys has also been proposed as a mimic of pSer and pThr. Indeed, histone H3 containing pCys at position 10 was recognised by the cognate antibody for pSer and by a 14-3-3 domain 'reader' for this residue[39]. Further, synthetic phosphorylation of p38α kinase to install pCys as a mimic of pThr at position 180 was demonstrated to at least partially recapitulate the kinase activity of the activated enzyme (**Figure 6.4b**)[22].

While promising, the stability of the *S*-phosphate linkage in cellular media remains uncertain, motivating exploration of methods to install alternate stable linkages.

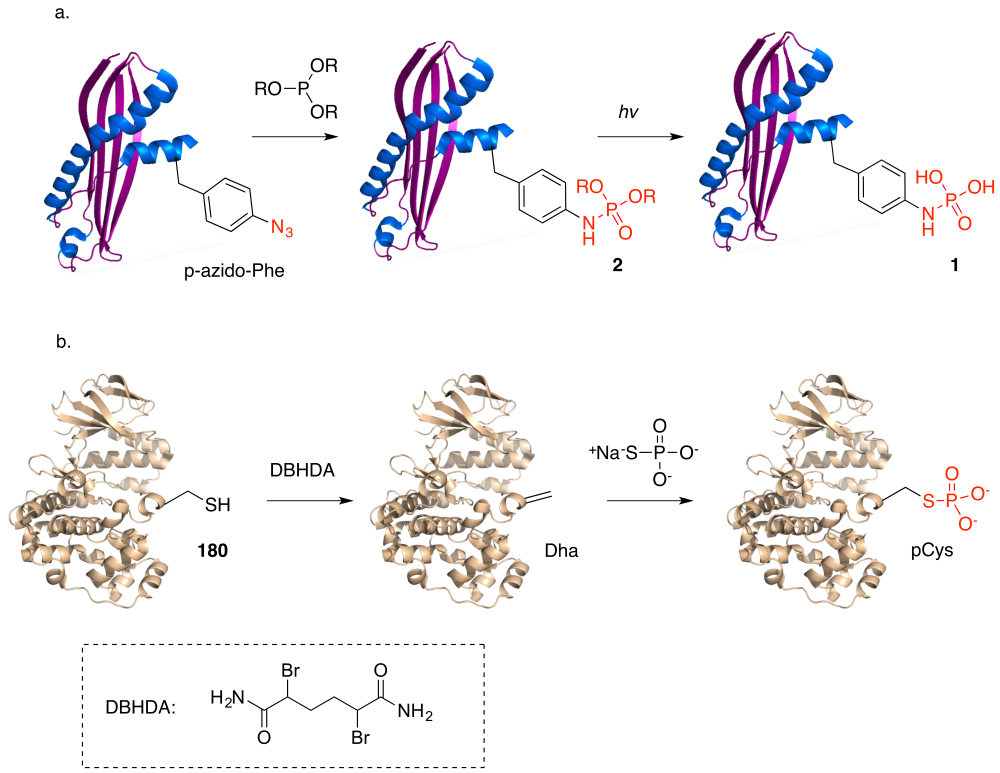


Figure 6.4: 'Tag and modify' approaches to site-specific protein phosphorylation via a. Staudinger-phosphite reaction[37] and b. Thia-Michael addition to dehydroalanine[22]

Biological approaches to the incorporation of phosphorylated amino acids

Sidney Hecht reported the first genetically programmed incorporation of a phosphorylated amino acid into a protein[40]. Using an *in vitro* translation system, phosphoserine and phosphotyrosine were incorporated into firefly luciferase by chemically misacylated tRNA, in response to nonsense codons. Photocaged phosphoproteins could also be incorporated via this method[41], allowing temporal control of signalling events controlled by phosphorylation.

The identification of Sep-tRNA synthetase (phosphoserine-tRNA synthetase, SepRS) as an unusual aminoacyl-tRNA synthetase from methanogenic bacteria[42] opened a novel approach towards the genetic

incorporation of phosphoserine into proteins. In these cells, SepRS is responsible for charging $tRNA^{Cys}_{GCA}$ with phosphoserine, generating pSer-$tRNA^{Cys}_{GCA}$. This pSer- containing tRNA does not act in ribosomal protein synthesis but instead acts as the substrate for enzymatic conversion to Cys-$tRNA^{Cys}$ [42]. It is the latter species that is ultimately transported to the ribosome, directing incorporation of cysteine residues into proteins.

A number of researchers soon realized that by engineering an appropriate $tRNA^{Cys}$ to recognize an amber codon, this codon could be reassigned for pSer incorporation into proteins[43,44]. Indeed, initial proof-of-principle studies demonstrated the feasibility of this approach, with site-specifically phosphorylated MEK1 protein being produced in *E coli*, albeit with impractically low yields[43]. Challenges in engineering the SepRS/$tRNA^{Cys}_{GCA}$ pair arise from the fact that swapping the $_{GCA}$ anticodon for the amber $_{CUA}$ renders the tRNA a poor substrate for aminoacylation with pSer and also because the aminoacylated pSer-tRNA itself is a poor subtrate for the EF-Tu proteins required for transport to the ribosome[45]. Extensive engineering and evolution of the SepRS, the tRNA, and auxiliary proteins such as EF-TU has now resulted in 'practical' systems for programmed incorporation of pSer in response to amber codons in *E coli*[44,46]. Notably, by also altering the expression of biosynthetic enzymes responsible for endogenous pSer production, the non-hydrolyzable pSer analogue phosphohomoalanine (Pha) could also be in incorporated into proteins using the orthogonal SepRS/tRNA pair[44].

Goals of this chapter

With reliable methods for C-C bond formation at dehydroalanine in hand, we sought to extend our chemical approach to the synthesis of phosphoproteins. In particular, we reasoned that C-C bond formation could enable post-expression installation of the metabolically stable carba-analogues of pSer previously employed by the Cole laboratory[21]. Access to these useful PTM-mimics via selective chemical modification has not yet been reported and would enable their installation at protein sites

inaccessible to ligation approaches without requiring engineering of biosynthetic enzymes.

Results

Post-translational mutagenesis enables installation of phosphoserine mimetics

The chemically labile nature of the O-P linkage in phosphoserine has inspired a number of efforts towards phosphoserine mimetics containing more stable linkages[21]. In most cases, these stable pSer analogues have been incorporated via peptide synthesis followed by ligation of the phosphopeptide fragment to a recombinant protein. We sought a complementary and perhaps more general approach to protein phosphorylation in which stable analogues of pSer, containing a single oxygen→methylene substitution, could be installed site-specifically to an intact protein (**Figure 6.5,** a and b). We reasoned that the C-C bond in phosphonate **3,** phosphohomoalanine[44] (Pha, or carba-phosphoSer) could potentially be installed via C-C bond forming chemical mutation using protein-Dha and the reported iodophosphonate **4** (**Figure 6.5b**).

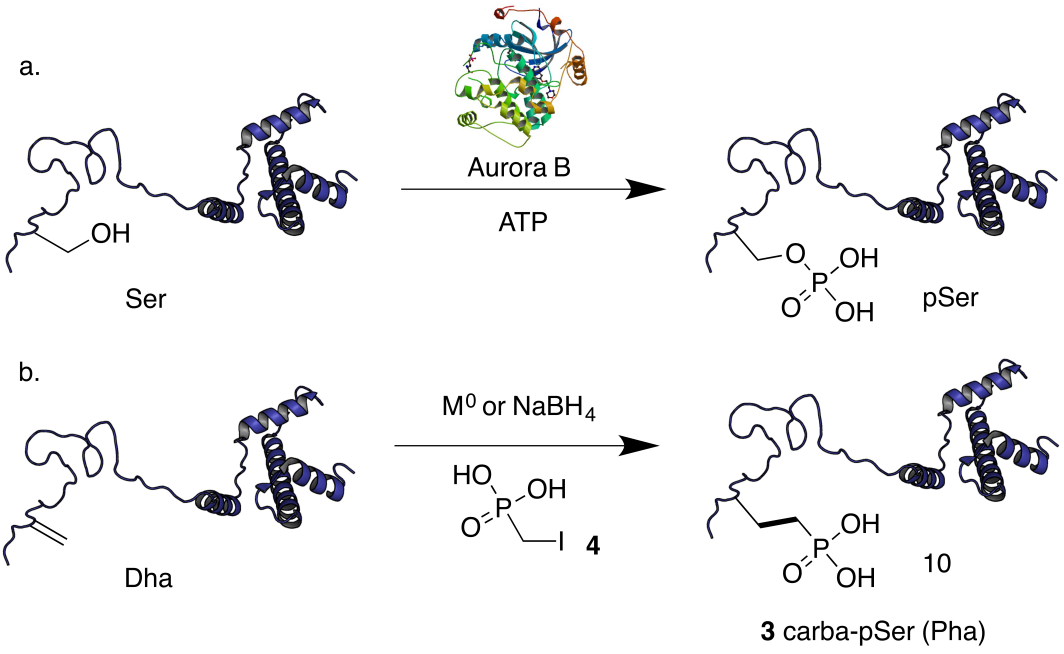


Figure 6.5: Phosphorylation of H3-S10 in nature (a.) and synthetic phosphorylation via C-C bond forming 'chemical mutation' (b.) The installed carba- linkage is highlighted in black.

As an initial protein model, we again employed histone H3-Dha, in this case targeting H3-S10Ph as the PTM of interest (**Figure 6.5b**)[14,47]. We chose this site as known 'readers' have been reported[48,49], enabling biochemical validation of the pSer mimic. Furthermore, the Davis group has explored the consequences of H3-S10Ph on nucleosome structure and dynamics[50]; access to chemically modified analogues may provide novel tools for this endeavor. Recombinant *Xenopus laevis h*istone H3-S10C was expressed in *E. coli* and purified by size-exclusion (SEC) and ion-exchange chromatography[51]. Treatment of the cysteine mutant protein with DBHDA[39], followed by heating at 37 °C, afforded H3-S10Dha (>95% conversion) for use in chemical mutation (see **Experimental Section**). The requisite side-chain 'donor' reagent, iodophosphonate **4,** was synthesized by acidic hydrolysis of the corresponding phosphonate ester**.** Initial attempts at chemical mutation using indium powder and reagent **4** displayed some evidence for product formation by LC-MS, although in the context of complex product mixtures and protein degradation.

In contrast, reactions mediated by $NaBH_4$ afforded clear conversion to the desired carba-pSer modified protein. However, reduced conversions were observed, at identical iodide loadings, relative to other iodo- reagents previously tested. After extensive attempts at optimization, it was discovered that a large excess of borohydride (10k eq. relative to protein, previous reactions typically employed 1000 eq.) was necessary to promote full conversion to the desired species (**Figure 6.6,** H3-S10Pha calc. 15317 obs. 15318). The hydrolysis of this quantity of borohydride in the aqueous reaction mixture is accompanied by vigorous effervescence, which complicates protein desalting and isolation. Although this transformation was unusually capricious, these unusual conditions could reproducibly achieve excellent conversions, if variable yields. The cause of the poor efficiency of chemical mutation employing **4** remains unknown.

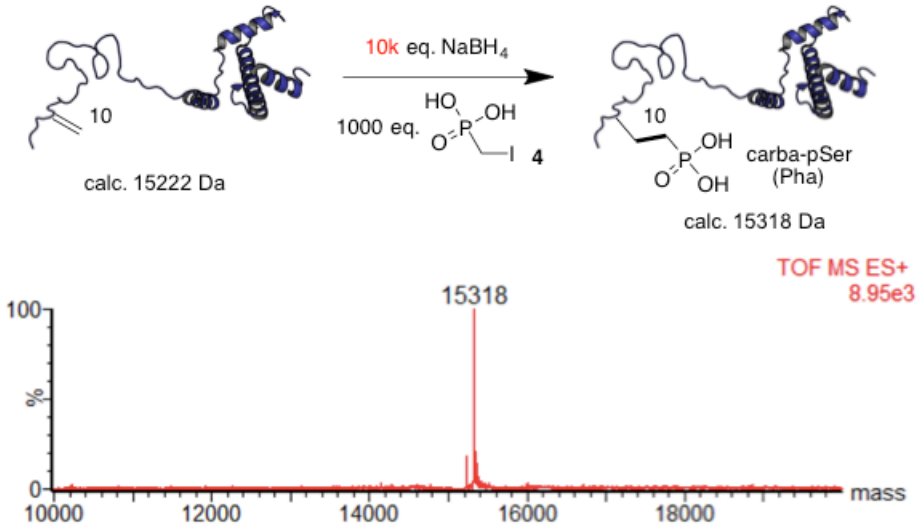
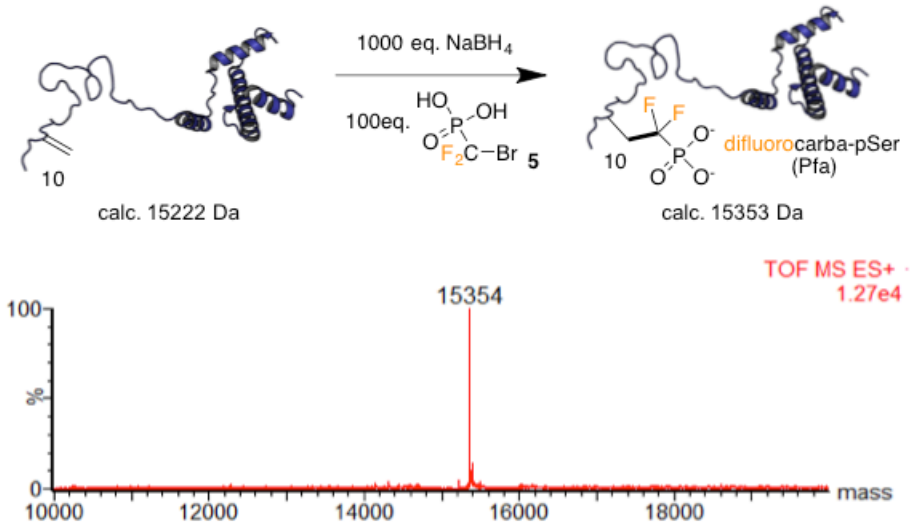


Figure 6.6: Installation of H3-S10Pha by borohydride-mediated chemical mutation

The difficulties associated with this initial carba-phosphorylation reaction motivated us to explore alternative iodo-precursors for related pSer mimetics. The difluoromethylene analogue of Pha **3,** phosphodifluoromethylalanine (Pfa), has been utilized by the Cole laboratory[35] and others, building on the proposal that α-fluorinated phosphonates are isopolar mimetics of phosphate esters[24]. Usefully, the difluoromethylene group has been reported to impart nucleophilic-type polarity to carbon-centred radicals[52], suggesting the reagent might be more suitably matched with Dha acceptors than the previously-employed methylene analogue[53-55]. Pleasingly, upon treatment of H3-S10Dha with (bromodifluoromethyl) phosphonic acid **5** and $NaBH_4$ the desired product, (Pfa, or difluorocarba-pSer)**,** was observed by LC-MS of the reaction mixture (**Figure 6.7**). Strikingly, **5** proved a remarkably efficient 'donor' for the $NaBH_4$-mediated chemical mutation reaction, requiring only 100 eq. for complete conversion under optimized conditions. Relative to the previous reaction with iodophosphonate **4,** reactions using **5** also required greatly reduced borohydride loadings (500-1000 eq.). These results may point to the suggested enhanced 'nucleophilicity', and corresponding reactivity with electrophilic acceptors, of carbon-centred radicals formed at difluoromethylene groups.



Figurer 6.7: Installation of H3-S10Pfa by borohydride-mediated chemical mutation

Characterisation of phosphoserine mimetics of H3-S10Ph

Having identified suitable conditions for the precise installation of carba-pSer and difluorocarba-pSer into histone proteins, we next sought to characterize the recognition of these pSer side-chain mimetics by protein 'readers' of the natural modification. Western blotting experiments revealed faithful recognition of carba-pSer and difluorocarba-pSer containing histones by a commercial antibody raised against H3-S10Ph, supporting their effective mimicry of this residue (**Figure 6.8**). The corresponding wild-type histone, bearing no phosphorylation at serine 10, was not recognized.

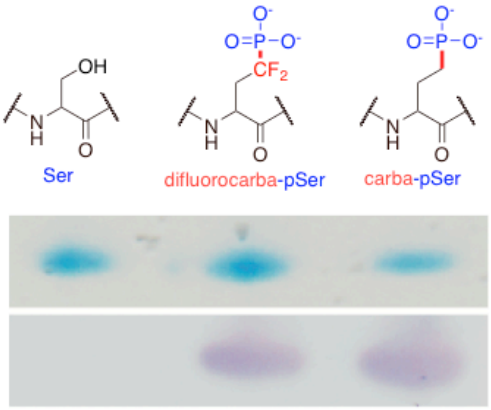


Figure 6.8: H3-S10Pha and H3-S10Pfa are recognized by a commercial antibody for H3-S10pSer, suggesting effective mimicry of this PTM

In order to probe recognition by a naturally occurring 'reader' domain, we chose to investigate the interaction of MORC3, with our synthetically phosphorylated histones. MORC3 (MORC family CW-type zinc finger protein 3) was recently identified by interaction proteomics as a preferred binding partner for H3-S10Ph containing peptides[49]. A competition assay based on Alphascreen™ technology[56] was used to assess binding between the modified H3-S10 proteins and MORC3. Alphascreen is a bead-based assay that relies on singlet oxygen diffusion between a donor and acceptor bead in close proximity to generate signal (emission at 615 nm) upon irradiation. Donor and acceptor beads are brought into close proximity by the interaction of immobilized complementary protein or peptide interaction partners.

We designed a competition assay[39] in which a biotinylated H3-S10Ph peptide and His-tagged MORC3 protein were mixed with streptavidin donor and nickel chelate-conjugated acceptor beads to generate signal in an interaction-dependent manner. First, the concentrations of assay components were titrated to obtain a robust signal. For the competition experiment, addition of MORC3-binding 'competitors' should lead to a reduction in the emission signal due to sequestration of MORC3, and thus Ni-conjugated acceptor beads. Indeed, upon adding wild-type, carba-pSer and difluorocarba-pSer to the assay mixture, clear reduction in signal was observed (**Figure 6.9**). Notably, the decline in signal with increased competitor loading followed the order wild-type< carba-pSer< difluorocarba-pSer. This result corroborates the previously reported enhanced interaction between MORC3 and histone H3 upon phosphorylation at H3-S10[49]. Further, it suggests difluorocarba-pSer may indeed act as a more suitable mimic of pSer than carba-pSer, as previously proposed[24]*****. The molecular details of these differences could be investigated further by biophysical means, a prelude to which is discussed below.

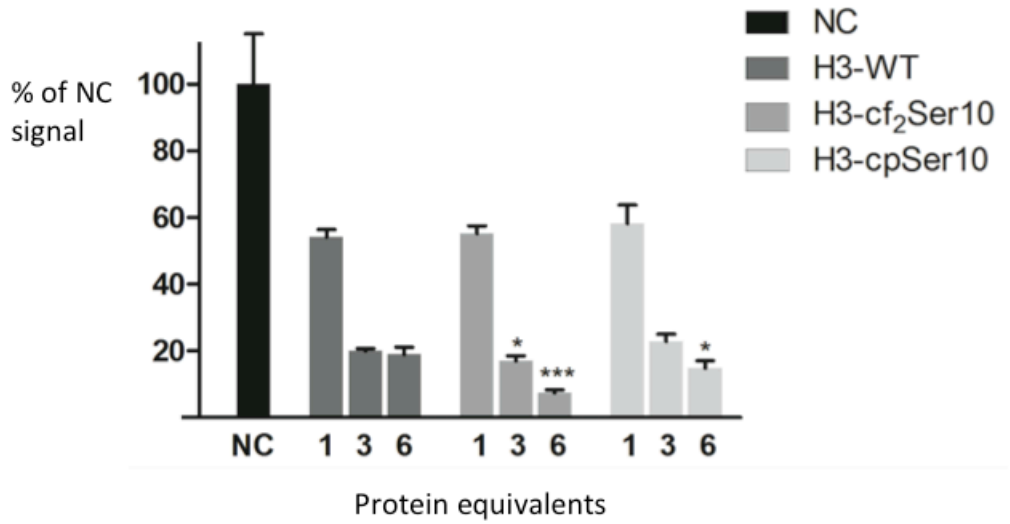


Figure 6.9: A bead-based competition assay reveals enhanced binding of histone H3 and MORC3 upon phosphorylation **NC** = no competitor.

***As the modified pSer-mimetic residues were present in the experiment as epimeric mixtures resultant from radical modification, it would be of interest to compare the 'pure' $_L$ and $_D$ forms of these residues. These residues could be introduced by EPL incorporating chemically synthesized histone phosphopeptides or by recently developed methods for site-specific genetic incorporation of pSer and analogues. Further, to get a strict assessment as to which pSer mimetic more faithfully recapitulates the binding behavior of natural H3-S10Ph, homogenous samples of H3-S10Ph must also be prepared, an endeavor outside of the scope of this current work.*

^{19}F NMR studies of H3-S10Pfa: Regioselectivity of chemical mutation

The presence of two ^{19}F nuclei in Pfa, entirely absent in proteins naturally and possessing unique NMR properties[57] (gyromagnetic ratio close to proton, large chemical shift dispersion and extreme sensitivity to local chemical environment), and the biological relevance of this stable mimetic of phosphoserine, prompted the choice of this residue for further characterization by protein NMR.

It was anticipated that the unique connectivity of the ^{19}F atoms, adjacent to the C-C bond constructed during the chemical mutation reaction, could allow insight into the diastereoselectivity and regioselectivity of the chemical mutation reaction. In previous chapters, the site of reaction (at Dha) and chemoselectivity (no modification observed at residues other than Dha) of the transformation were confirmed by intact protein MS and proteolytic digest followed by LC-MS/MS analysis. However, no information as to the atomic connectivity of the introduced modification is typically gained from these techniques. Protein NMR is one technique that can provide unambiguous atomic connectivity data through the analysis of through-bond (scalar) couplings. Indeed, it is a mainstay of structural biology to analyse proteins in this manner[58]. Surprisingly, this technique has rarely been used by protein chemists for characterization purposes[59], likely due to the perceived (and largely accurate) requirement for large amounts of protein material, additional requirements for isotopically labeled proteins/ reagents, and limited access to the requisite high-field spectrometers.

As protein NMR typically requires highly concentrated samples, we sought to further optimize reaction conditions for the conversion of H3-S10Dha to H3-S10Pfa on multi-milligram samples. Remarkably, using 10 milligrams of H3-S10Dha (in 1 mL buffer, 660 μM) only 75 equivalents of (bromodifluoromethyl)phosphonic acid **5** are required to effect almost full (89% based on LC-MS analysis) conversion to the H3-S10Pfa product in the presence of sodium borohydride (**Figure 6.10**). In these experiments, no

problematic di-alkylation products or protein degradation were observed. Typical yields after protein desalting and dialysis/concentration into NMR buffer were ~60%.

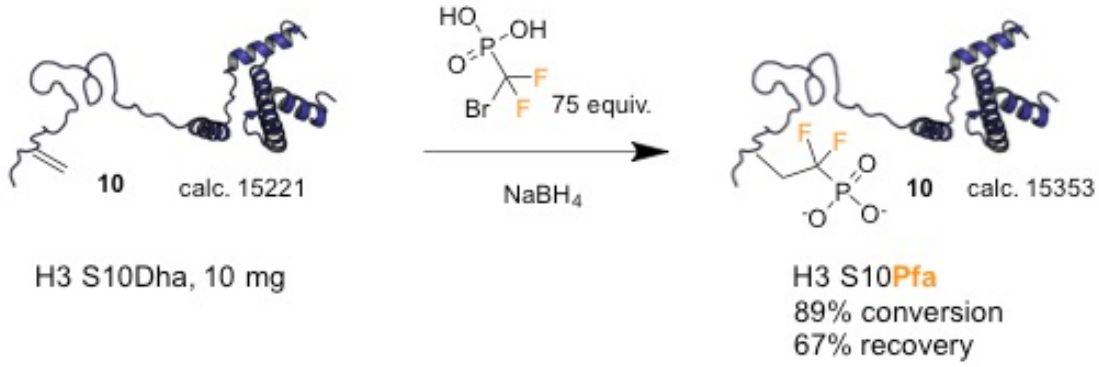


Figure 6.10: Large-scale synthesis of H3-S10Pfa for protein NMR

Initial NMR experiments (performed by WL Ng and T. Claridge) employed a sample containing the H3-cf2pSer10 (160 μM) in HEPES buffer (25 mM, pH 7.5) supplemented with 100 mM NaCl and 10% D_2O. ^{19}F {^{31}P} spectra at this sample concentration required a large number of scans (2048) but revealed a number of clear features. Although ^{1}H-couplings are not resolved, the coupling pattern observed can be explained by the initially proposed side-chain structure (**Figure 6.11a**) wherein 2 diastereotopic fluorines are strongly coupled, generating four ^{19}F resonances for each protein diastereoisomer, with the chemical shifts at least partially overlapping**.** Simulated spectra (using parameters extracted from the experimental system) reproduced the observed spectra when the two diastereomers were assumed to exist in equal concentrations (**Figure 6.11b**). Notably, the chemical shift and couplings observed for the protein Pfa were consistent with the previously reported values for the free amino acid[60].

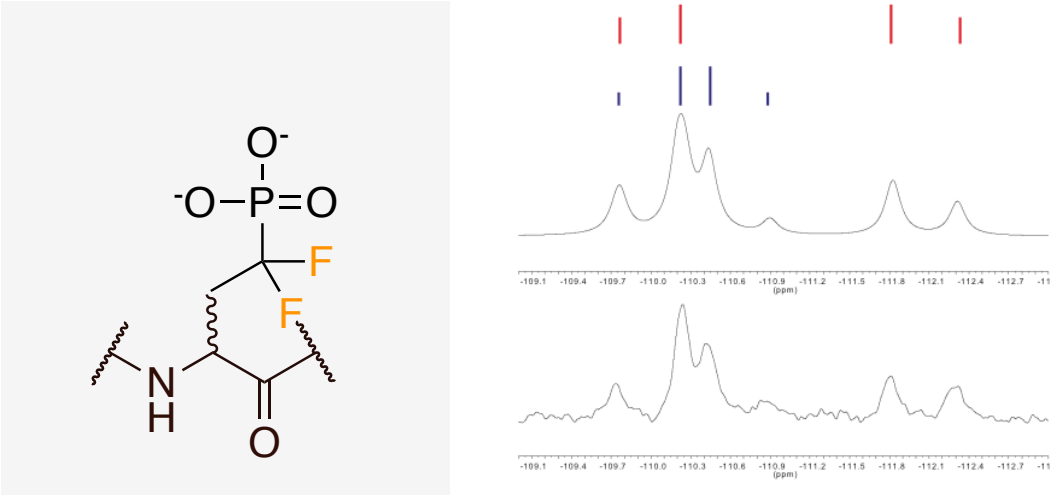


Figure 6.11: a. Structure of Pfa residue highlighting two ^{19}F nuclei. b. Spectra of H3-S10Pfa (bottom), compared with simulated (top). See Experimental Section for details.

Having demonstrated the use of ^{19}F NMR for analysis of the diastereoselectivity of protein chemical modification, it was anticipated that observation of the expected three-bond scalar coupling between the γ-fluorines and β-protons would provide a ready means of confirming the desired regioselectivity of reaction (**Figure 6.12**).

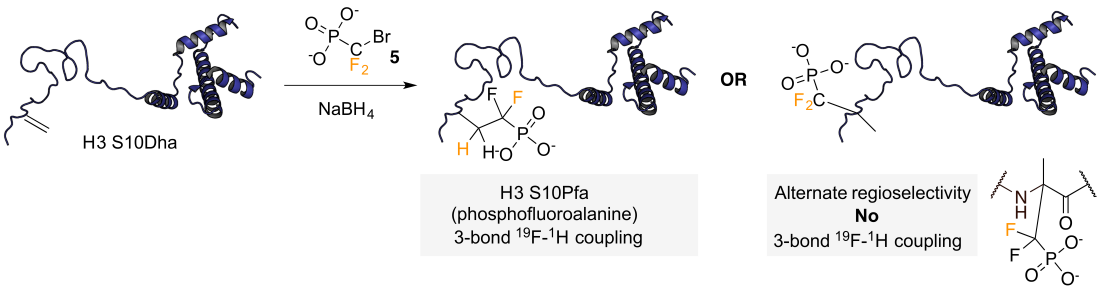


Figure 6.12: ^{19}F NMR suggests a method for determining the regioselectivity of the chemical mutation reaction 'on-protein'

The large number of scans required for the initial 1D ^{19}F data indicated that a more concentrated sample was necessary for the proposed ^{19}F-^{1}H 3-bond coupling experiments. A fresh H3-S10Pfa sample was therefore prepared and concentrated by centrifugal concentration (Vivaspin 500 MWCO=3 kDa) to a concentration of 1.2 mM (18.6 mg/mL) in 50 mM ammonium formate (D_2O). Routine 1D spectra could be readily attained at this protein

concentration. Notably, the observed $^{19}F\{^{31}P\}$ spectra reproduced the earlier spectra, despite use of a different instrument and buffer.

The small-molecule Pfa model compound **6** was prepared in two steps (synthesis performed by Mateusz Imiolek) by zinc-mediated addition of diethyl (bromodifluoromethyl)phosphonate to Dha model system Ac-Dha-Me, followed by hydrolysis with TMS-iodide (**Figure 6.13**, see **Experimental Section** for further details). Unexpectedly, the ^{19}F spectra of this compound did not exhibit the coupling expected from two chemically inequivalent, diastereotopic fluorine nuclei, which is observed in the protein system, and notably, in previously reported spectra for the free acid form of Pfa[60]. The cause of this divergent NMR behavior of the *N*-methyl amide form of Pfa is unknown. Critically, the Pfa *N*-methyl amide synthesized by zinc-mediated radical addition did display the expected ^{19}F-^{1}H γ-fluorine and β-proton three bond coupling in 2D correlation spectra, proving the desired regioselectivity for addition in the model system (see **Experimental Section**).

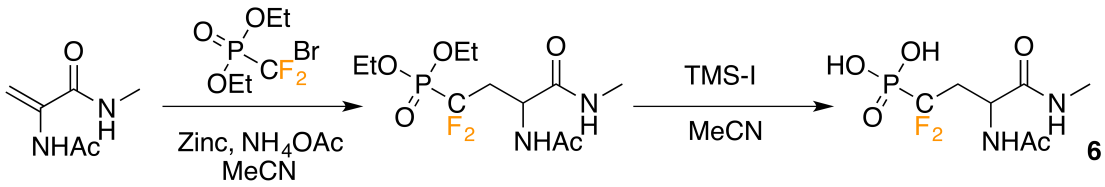


Figure 6.13: Synthesis of small molecule Pfa model.

The three-bond scalar coupling experiment envisioned for use on protein-Pfa was then optimized on the *N*-methyl amide Pfa model compound **6**. A typical 1D heteronuclear-edited HSQC-type pulse sequence was used, with ^{19}F substituting for ^{15}N, and ^{31}P decoupling applied throughout. The results are shown below (**Figure 6.14**) and match exactly with the 2D ^{1}H, ^{19}F HSQC of the model compound previously recorded, with a clear scalar coupling shown between the γ-fluorines and β-protons (^{1}H δH 2.5 ppm). In the case of the model compound, the 5-bond coupling between the γ-fluorine and α-proton is also (weakly) observed (^{1}H δH 4.6 ppm).

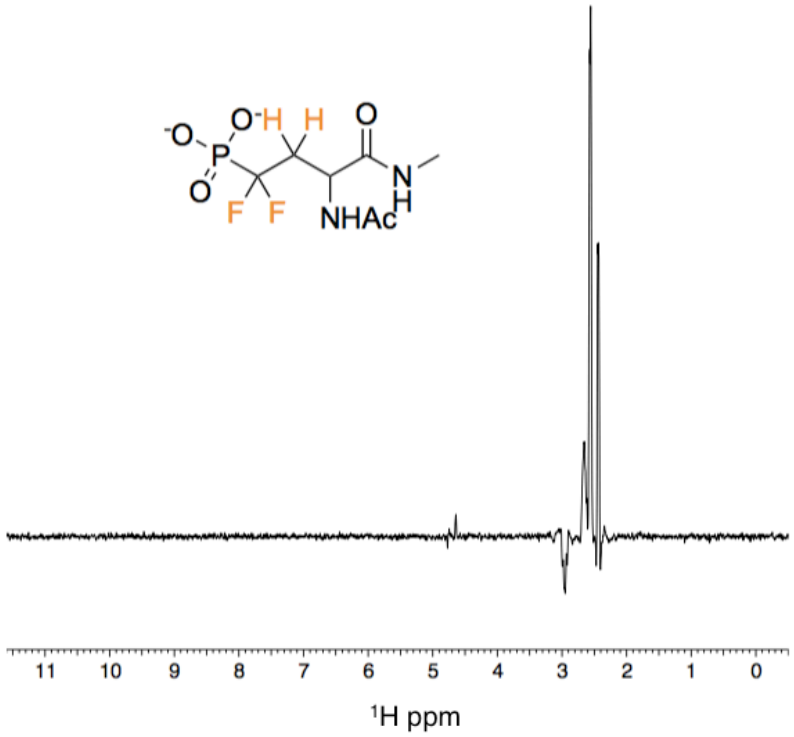


Figure 6.14: 1D ^{19}F-^{1}H scalar-coupling experiment on small molecule Pfa model **6**.

The 1D scalar-coupling experiment was then performed on the protein sample, with 16,384 scans being recorded (**Figure 6.15** and **Experimental Section**). Pleasingly, a clear signal is observed at ~2.5, consistent with that observed on the small molecule. This result reveals clear coupling between the introduced fluorine and a proton separated by three-bonds and consistent with the β protons (^{1}H δH 2.5 ppm) derived from Dha. This result thus confirms the expected regioselectivity of 'chemical mutation' with sodium borohydride, 'on protein', at least for the (bromodifluoromethylene)phosphate reagent used here.

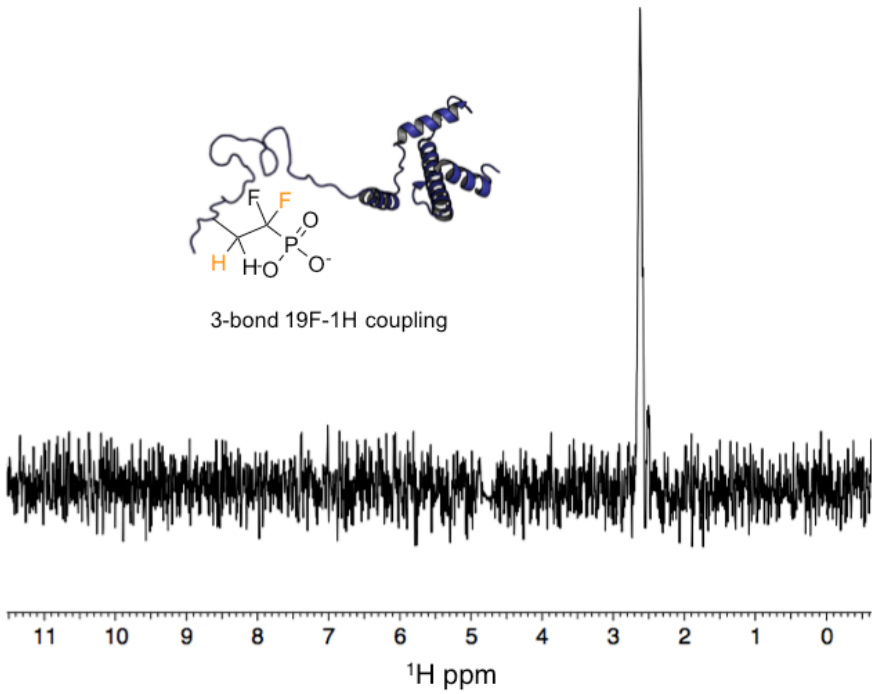


Figure 6.15: 1D-edited HSQC reveals the highlighted 3-bond F-H coupling on H3-S10Pfa, confirming the regioselectivity of chemical mutation

Discussion and future prospects

In this chapter, a facile route for access to carba-pSer (Pha) and difluorocarba-pSer (Pfa) on histone proteins was reported. Strikingly, the use of the difluoromethylene analogue allowed both milder conditions for chemical mutation and enhanced binding to a known pSer 'reader' protein, relative to the methylene-linked side-chain. The latter result suggests support for previous proposals that fluorophosphonates act as more effective phosphate mimics by virtue of polarity, geometry, and pKa[24].

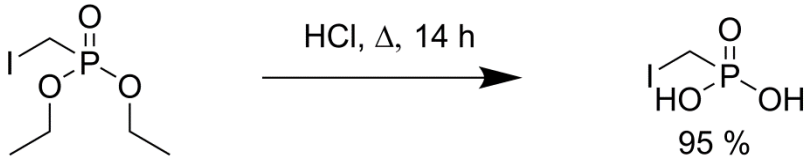
By analogy to the previously described work using ^{13}C NMR methyl groups as reporters of PTM 'eraser' activity, the use of the ^{19}F nuclei in Pfa as a novel 'isotope PTM' for biophysical investigation of protein phosphorylation, can be envisioned; work in such a direction is underway.

Experimental Section Chapter 6

Reagent synthesis

(Iodomethyl)phosphonic acid **4**

Reagent supplied by Sebastian Galan. Procedure is included for reference, originally reported in[61]



To a solution of diethyl (iodomethyl)phosphonate (1.00 g, 3.60 mmol) in water (2.12 mL), concentrated hydrochloric acid (4.25 mL)was added. The reaction mixture was heated to reflux for 14 h. The reaction mixture was cooled to room temperature, concentrated under reduced pressure then dried azeotropically with toluene (3 x 10 mL). The oily residue was flash frozen then dried by lyophilisation to afford the title product as a white amorphous solid (0.751 g, 3.42 mmol, 95 %).

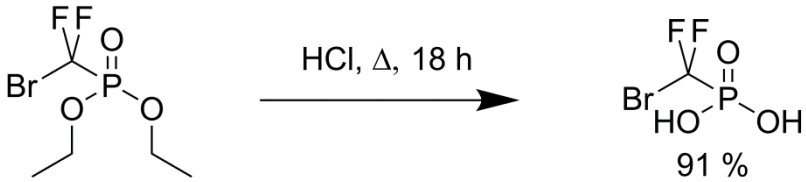
^{1}H NMR *δH* (400 MHz, DMSO) 7.30 (br. s, 2 H, OH) and 2.97 (d, J=10.0 Hz, 2 H, CH2)

^{13}C NMR *δC* (101 MHz, DMSO) -8.11 (d, J=145.4 Hz);

^{31}P NMR *δP* (162 MHz, DMSO) 15.75 (s);

HRMS m/z (ESI-): Found 220.88676 [M-H]-; $CH_3O_3{}^{127}IP$ requires 220.88700 [M-H]-.

(Bromodifluoromethyl)phosphonic acid 5

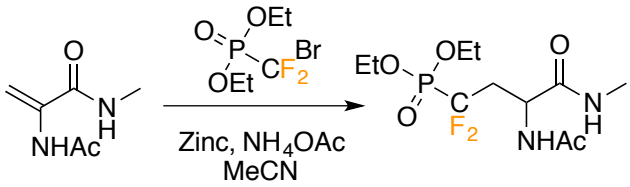


Diethyl (bromodifluoromethyl)phosphonate (1.00 g, 3.75 mmol) was treated with concentrated hydrochloric acid solution (10.0 mL) then heated to reflux for 18 hours. The reaction mixture was cooled to room temperature. The solution was evaporated under reduced pressure. The residue was dried azeotropically with toluene (3 x 10 mL) and dried extensively under high vacuum, to afford the title product as a colourless oil (0.720 g, 0.341 mmol, 91%).

^{1}H NMR *δH* (400 MHz, acetone-d6) 4.4 (2H, s)
^{19}F NMR *δF* (376.5 MHz, acetone-d6) -62.7 (d,$^{2}J_{PF}$ 88 Hz, 2 F)
^{31}P NMR *δP* (162 MHz, acetone-d6) -2.3 (t, $^{2}J_{PF}$ = 88 Hz, 1 P)
HRMS m/z (ESI+): Found 210.8967 [M+H]+; $CH_3O_3{}^{79}BrP$ requires 210.8966 [M+H]+.

Synthesis of Ac-Pfa-Nme (Pfa model compound) 6

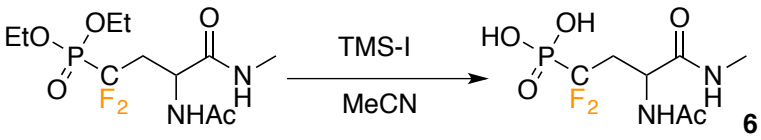
Synthesis performed by Mateusz Imiolek



Ac-Dha-Nme (80 mg, 0.563 mmol) was added to a round-bottomed flask equipped with a magnetic stir bar and suspended in 2 mL acetonitrile. 0.42 mL of 5 M NH_4OAc (aq., pH 6) was added and the mixture was stirred until all of the material had dissolved. Zinc powder (295 mg ,4.5 mmol) was added and the reaction mixture stirred and cooled on ice. Diethyl (bromodifluoromethyl)phosphonate (200 µL, 1.1 mmol) was slowly added over a ten minute period. The reaction was stirred on ice for a further 50 minutes at which point TLC (5% MeOH in DCM, ninhydrin) revealed no remaining starting material. The solution was filtered through a cotton plug

and extracted 3 times with EtOAc. The organic phase was dried *in vacuo* and purified on silica gel ((10 % MeOH in DCM) to provide the diethylphosphonate product.

^{1}H NMR *δH* (400 MHz, chloroform-d) δ 7.16 (s, 1H) 7.00 (d, J = 3.7 Hz, 1H) 4.83 (dd, J = 13.1, 6.8 Hz, 1H) 4.37 – 4.18 (m, 4H), 2.77 (d, J = 3.4 Hz, 3H), 2.65 – 2.44 (m, 2H), 2.00 (s, 3H), 1.38 (dt, J = 9.5, 7.0 Hz, 6H). ^{19}F NMR *δF* (376 MHz, chloroform-d) δ -105.07 (dt, J = 100.3, 15.1 Hz), -105.90 (dt, J = 101.6, 14.0 Hz), -109.58 (dt, J = 109.5, 23.3 Hz), -110.41 (dt, J = 109.8, 22.2 Hz).



6

The above diethylphosphonate intermediate (21.3 mg, 64 mmol) was dissolved in 1 mL acetonitrile. N_2 gas was bubbled through the solution for 10 minutes and the mixture was cooled on ice. TMSI (276 µL, 30 eq) was added dropwise and the mixture stirred for 100 min at RT. The solvent was removed *in vacuo* and the resultant residue treated with a cold solution of 10% AcOH. The solution was stirred at RT for 60 min, at which point the organic phase was extracted multiple times with DCM until no longer pink. Lyophilization of the aqueous fraction gave 42 mg of a brown solid containing the desired product, Ac-Pfa-NMe, **6**.

^{1}H NMR *δH* (400 MHz, deuterium oxide) δ 4.62 (dd, J = 8.8, 4.4 Hz, 1H), 2.69 (s, 3H), 2.55 – 2.33 (m, 2H), 1.99 (s, 3H). unreferenced ^{31}P NMR *δP* (162 MHz, deuterium oxide) δ 3.97 (broad). ^{19}F NMR *δF* (376 MHz, deuterium oxide) δ -113.39 (dt, J = 93.5, 19.4 Hz). ^{13}C NMR *δC* (101 MHz, deuterium oxide) δ 48.25 (-CH) 34.78 (- CH_2) 25.95 (-CH_3), 21.74 (-CH_3) ESI (-) calcd

$[M-H]^-$ 273.05 found: 273.0

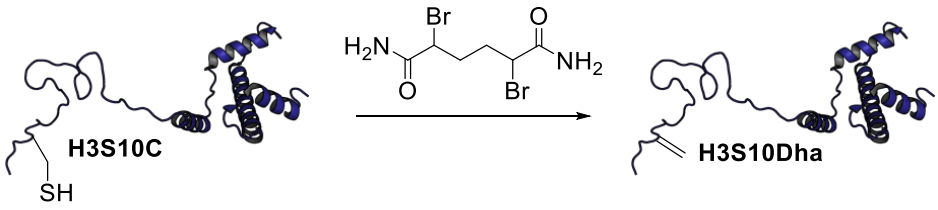
Protein modification

Histone H3 (S10C, C110A)

ARTKQTARK**C**TGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR
YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN
LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15254.8
Observed mass = 15255

Dehydroalanine formation

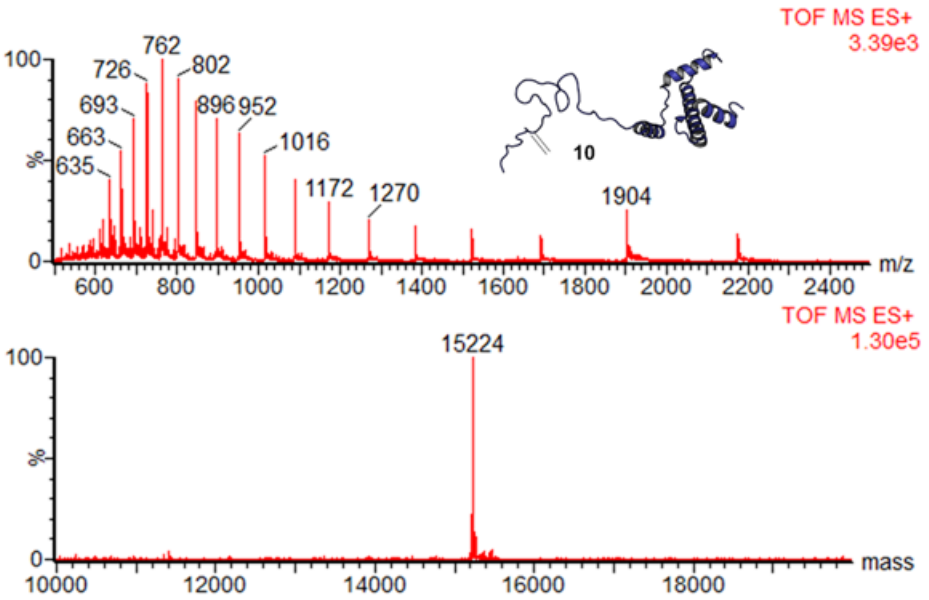


The dehydroalanine-containing histone was prepared according to general procedure **1.**

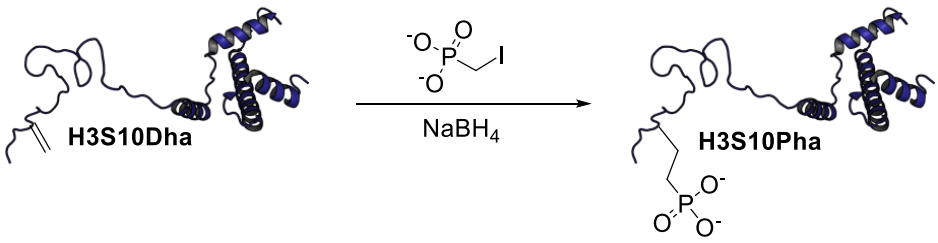
ESI-MS spectrum for the modified histone is shown below.

Calculated mass: 15220.7

Observed mass: 15224



Installation of phospho-homoalanine from dehydroalanine H3 S10



A 160 µL aliquot of histone H3S10Dha (0.88 mg/mL, 58 µM in reaction buffer, 500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0) was added to a N_2 atmosphere glovebox and left overnight at room temperature.

The following morning, within the glovebox, the degassed protein solution was combined with 2.1 mg of iodo-methylenephosphonic acid **4.** The resultant solution was gently mixed by pipette.

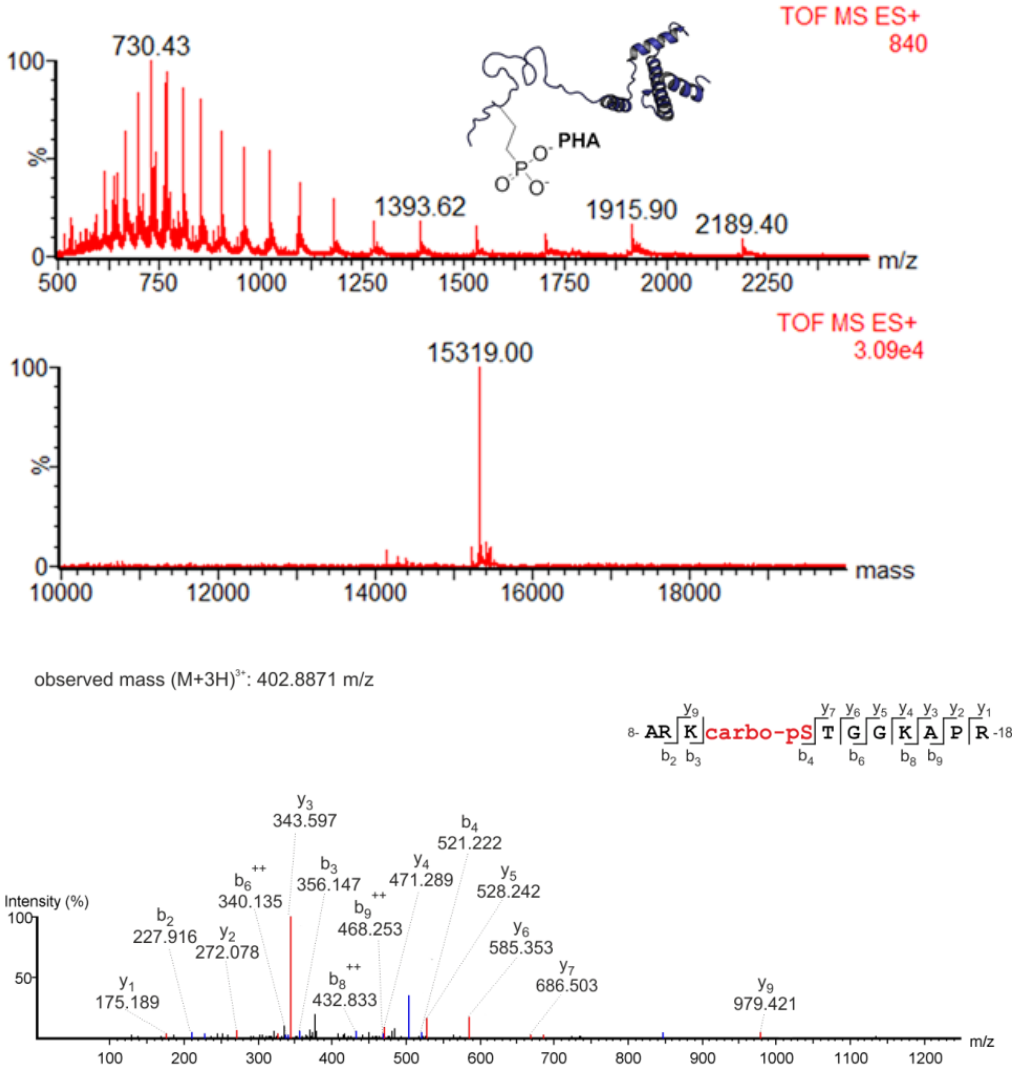
The entire volume (~200 µL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 3.5 mg powdered $NaBH_4$. The tube was gently shaken by hand, causing vigorous effervescence/ foaming to spread through the solution. The reaction was gently shaken at 4°C for one hour before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

It should be noted that for this reaction the equivalents of $NaBH_4$ used was 10 times greater than proved necessary for other halide reagents.

ESI-MS for the modified histone is shown below.

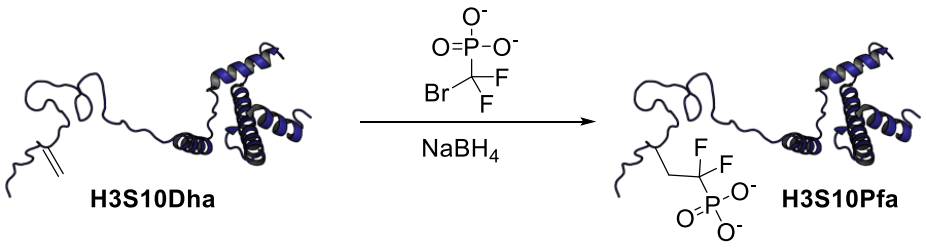
Calculated mass: 15316.8

Observed mass: 15319



MS/MS spectrum of the peptide $_{8}$**ARK$_{cp}$STGGKAPR$_{18}$**

Installation of phospho-difluoromethylalanine from dehydroalanine H3 S10



Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to a 2 mL eppendorf tube, followed by 20 µL of a 10mg/mL histone H3S10Dha stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to a N_2 atmosphere glovebox and left overnight at room temperature.

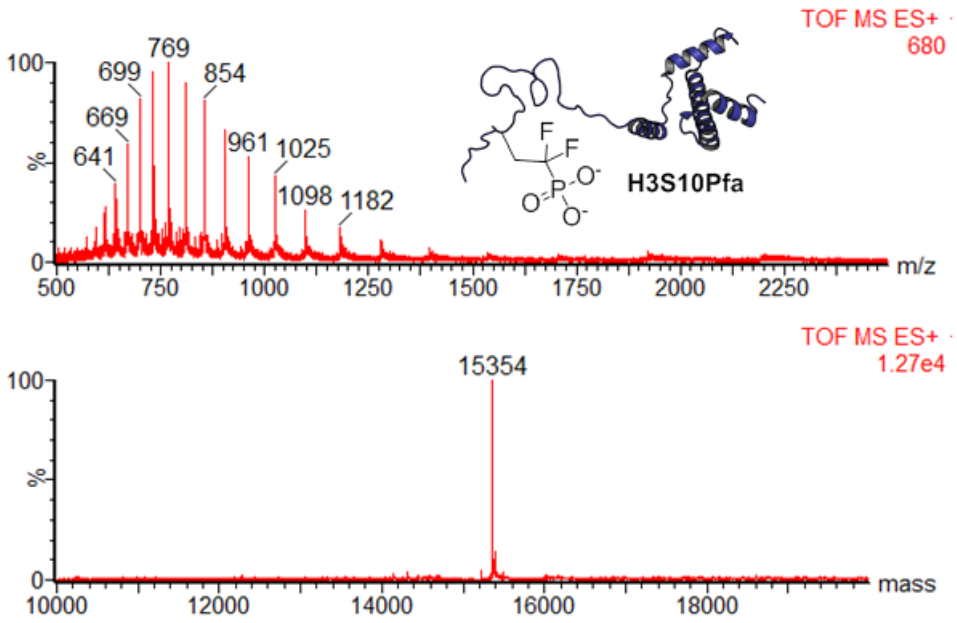
A stock solution of (Bromodifluoromethyl)phosphonic acid **5** (prepared by dissolving 55 mg **5** in 275 µL water and adjusting the pH to 6.0) was also ported into the glovebox.

The following morning, within the glovebox, 2.5 µL of the above stock solution was pipetted into the protein solution, and the resultant solution gently mixed by pipette. The entire volume (~200 µL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 1 mg of solid $NaBH_4$. Sustained effervescence was observed. The reaction was incubated at 4°C for 30 minutes before being removed from the glovebox and quenched by elution through a SpinTrap™ G-25 desalting column (pre-equilibrated with fresh reaction buffer), affording the desired chemical mutagenesis product. The desalted solution was analysed by LC-MS.

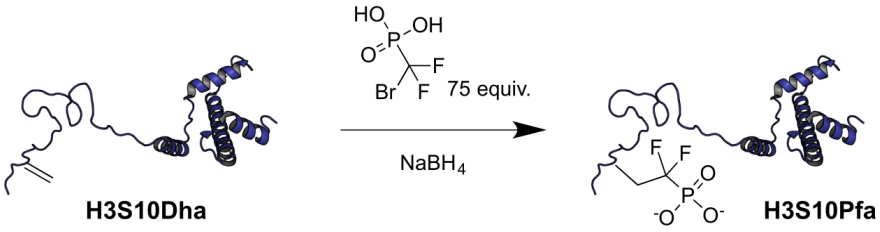
ESI-MS for the modified histone is shown below.

Calculated mass: 15352.8 (H3S10Pfa, C110A)

Observed mass: 15354



Multi-milligram scale installation of phospho-difluoromethylalanine (Pfa)



A 2 mL eppendorf tube containing a solution of H3S10Dha (10 mg/mL, 1 mL, 0.66 mM) in reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0) was ported into a N_2 atmosphere glovebox (Belle Technologies) and equilibrated to anaerobic conditions for 6 hours at room temperature.

A stock solution of (Bromodifluoromethyl)phosphonic acid **S9** (prepared by dissolving 55 mg **S9** in 275 µL water and adjusting the pH to 6.0) was also ported into the glovebox.

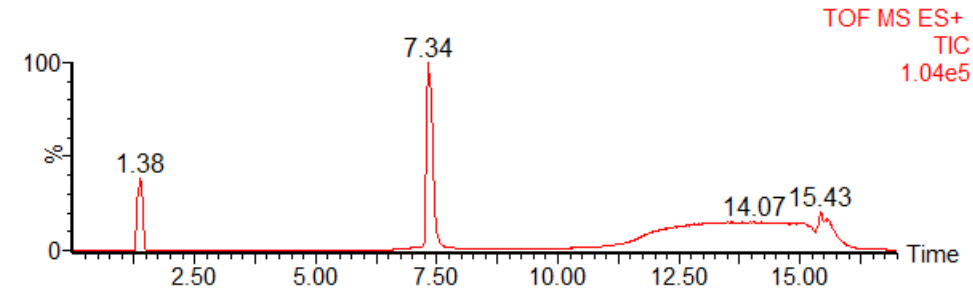
50 µL of the above stock solution was pipetted into the protein solution, and the resultant solution gently mixed by pipette. The entire volume (~1 mL) of the protein solution was then added to a separate 2 mL eppendorf tube, kept at 4°C and containing 5 mg of solid $NaBH_4$. Sustained

effervescence is observed. The reaction was incubated at 4°C for 30 minutes before being removed from the glovebox.

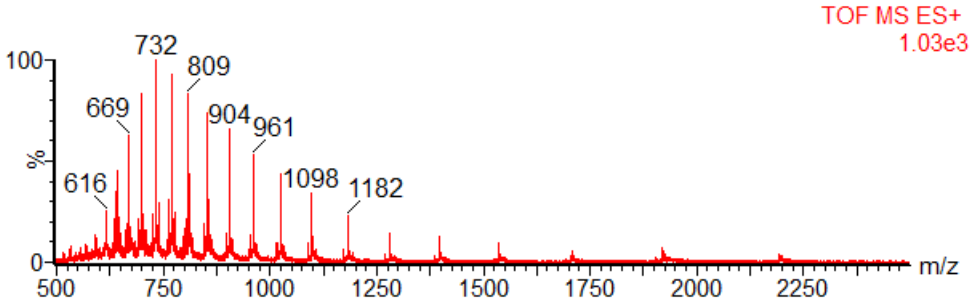
Outside the glovebox, the reaction mixture was concentrated to 500 µL using a Vivaspin 500 concentrator (MWCO 10 kDa) and excess reagent was removed by passing the solution through a PD MiniTrap G25 (equilibrated with reaction buffer). Eluting with 1 mL of the same buffer provided the Pfa-containing protein. Protein concentration determination was performed using a NanoDrop 1000 spectrophotometer, which allowed the calculation of yield (6.75 mg recovered, 67.5% yield).

ESI-MS for the modified histone is shown below.
Calculated mass: 15352.8
Observed mass: 15355

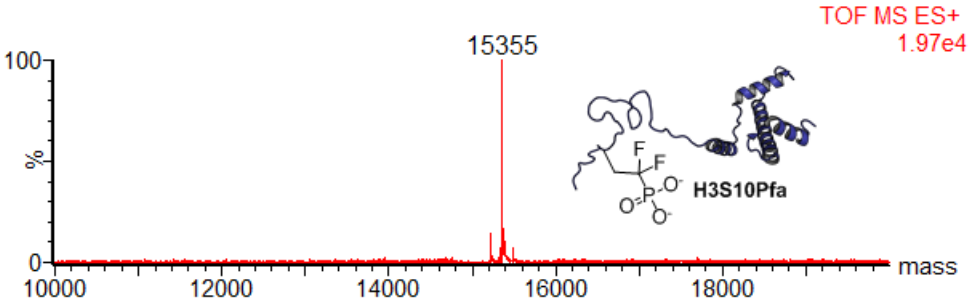
TIC chromatogram



Ion series (peak at 7.34 min)



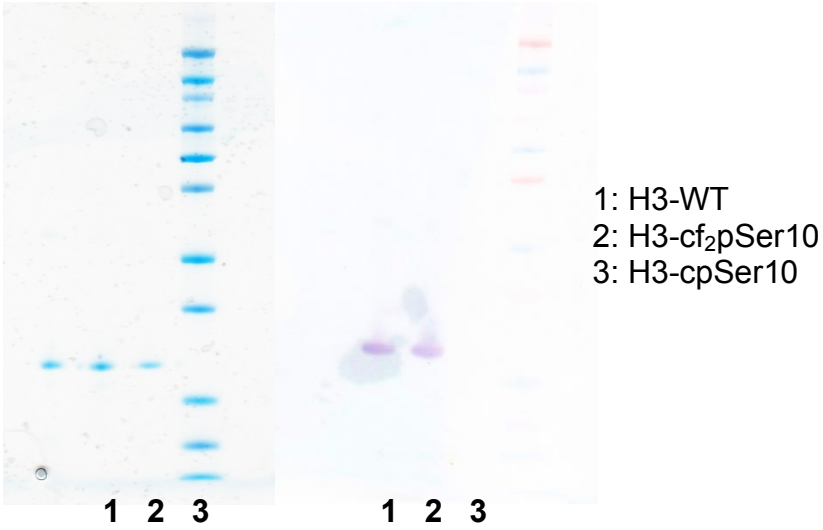
Deconvoluted spectrum



Western blot with Anti-H3S10Phos

Performed by WL Ng

Two aliquots of the H3 S10Phos mimics were subjected to SDS-PAGE using a 12% NuPAGE pre-cast gel (Invitrogen) and MOPS running buffer (Invitrogen). One aliquot was visualized using Instant Blue (coomassie) staining. The other sample was used for western blot analysis. A lane with unmodified H3 (S10C, C110A) was used alongside both samples as a control along with standard molecular weight markers. For the western blot, the protein was transferred to a PVDF membrane using an iBlot transfer apparatus (Invitrogen). The membrane was blocked with 20 mg/mL BSA in TBST buffer for 60 minutes at room temperature with rocking (TBST = 50 mM Tris, pH 7.5 with 150 mM NaCl and 0.05% (w/w) Tween 20). After blocking, the membrane was washed with TBST buffer (5 × 40 mL). The primary antibody (anti-H3-S10Phos, rabbit polyclonal, Abcam ab5176) was diluted from 6.6 µL to 20 mL in TBST buffer (with 2% BSA). The membrane was incubated with the primary antibody solution overnight at 4^{o}C with rocking. The membrane was then washed with TBST buffer (5 × 40 mL). The secondary antibody (anti-rabbit IgG produced in goat, conjugated with alkaline phosphatase, Sigma-Aldrich A3687) was diluted from 1 µL to 20 mL in TBST buffer (with 2% BSA). The membrane was incubated with the secondary antibody solution for 15 minutes at room temperature with rocking. The membrane was washed with TBST buffer (5 × 40 mL) and then water (2 × 40 mL). The membrane was then developed using BCIP/NBT liquid substrate system (Sigma-Aldrich B1911). A clear band was observed at both modified proteins and no band was observed at the unmodified H3 (S10C, C110A).



Full gel and western blot

Supplementary Figure 29. Western blot analysis of H3wt, H3-cpS10 and H3-cf$_2$pS10

(rabbit polyclonal, Abcam ab5176)

H3 S10ph mimetics binding assay

Performed by WL Ng and Maola Khan.

Histone binding assay was performed using AlphaScreen kit with minor modifications of the manufacturer's protocol (PerkinElmer, USA). All reagents were diluted in the recommended buffer (25 mM HEPES, 100 mM NaCl, 0.1% BSA; pH 7.4) and allowed to equilibrate to room temperature prior to addition to the plate. Biotinylated H3S10ph peptide was purchased from Anaspec, USA and human recombinant MORC family CW-type zinc finger protein 3 (MORC3) was purchased from MyBioSource, USA.

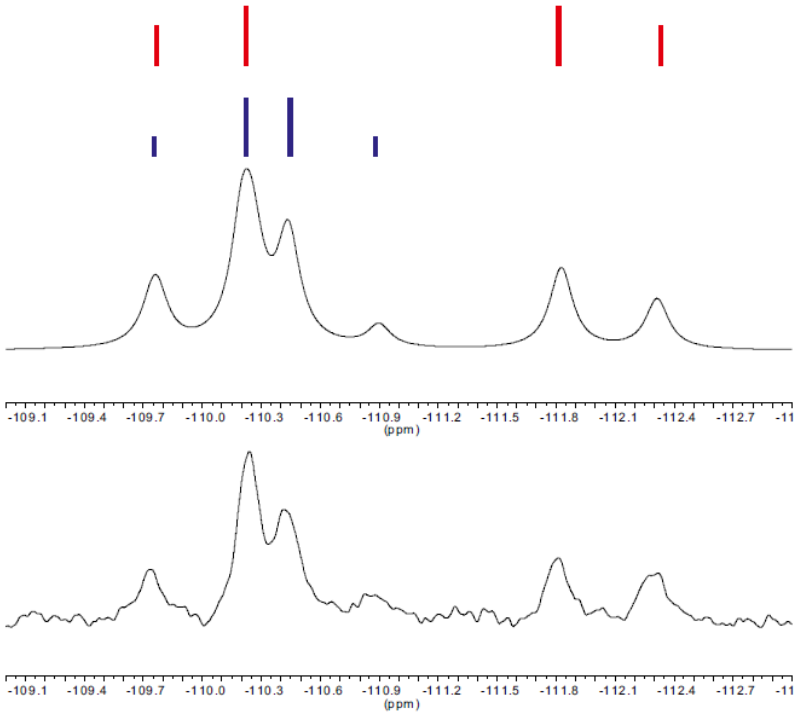
For the competition assay, 3 µL MORC3 (final conc. 40 nM) was mixed with 4 µL biotinylated H3S10ph (final conc. 400 nM) in a 384-well plate. The plate was sealed and incubated for 30 min at room temperature. Following incubation, 3 µL containing 1, 3 and 6 equivalent (400, 1200 and 2400 nM, respectively) of H3-WT, H3-cf$_2$Ser10, H3-cpSer10 were added and incubated for 45 min at room temperature to compete with H3S10ph. In the control wells, buffer was used instead of competitor. AlphaScreen streptavidin-conjugated donor and nickel chelate-conjugated acceptor beads were prepared as 2 x stock (50 µg/mL) and 10 µL were added to the wells to give a final concentration of 25 µg/ml and incubated for a further 45 min in the dark at room temperature. The plate was analysed using an Envision (PerkinElmer) plate reader. Measurements were performed in at least triplicates. The background was subtracted from the obtained reading and the values were normalized with respect to the control. Graph was drawn using GraphPad Prism and analysed by student's t-test.

^{19}F NMR procedure 1 (WL Ng and Tim Claridge)

NMR measurements were conducted using a Bruker AVIII 600 spectrometer equipped with a 5mm z-gradient broadband (BBFO) N_2-cooled Prodigy probe operating at 298K using 5 mm diameter NMR tubes (Norell). ^{19}F{^{31}P} decoupled spectra were recorded using inverse-gated GARP ^{31}P decoupling. These spectra were ^{1}H coupled since fluorine spectra were recorded on the shared ^{1}H/^{19}F coil, hence simultaneous ^{1}H

decoupling was not possible. ^{19}F 90° pulse lengths were 15 µs and spectra were typically obtained using 2048 scans, a recovery delay of 1s and and acquisition time of 0.6 s. Data were processed with 20 Hz Lorentzian line broadening using TopSpin 3.5 software (Bruker) and were referenced externally to $CFCl_3$ @ 0 ppm. Samples contained the H3-cf2pSer10 (160 µM) in HEPES buffer (25 mM, pH 7.5) supplemented with 100 mM NaCl and 10% D_2O.

Simulations of ^{19}F spectra were performed using MNova (v10.0.2) using data extracted from the experimental spectra.



A= -110.04, -112.04. J= 275Hz
B= -110.06, -110.58, J= 260 Hz

^{31}P-decoupled (^{1}H-coupled) ^{19}F NMR (1H couplings not resolved) (bottom) compared with simulated, based on the indicated parameters for A and B and equal concentrations.

^{19}F experiments 2 – Regioselectivity by ^{19}F-^{1}H HSQC with Gogulan Karunanithy

NMR experiments were conducted using a Aglient/Varian DD2 spectrometer (14.1 T/ 600 MHz) equipped with a 5 mm room-temperature triple resonance probe and carried out at 298 K. Experiments were carried out using 5 mm diameter Shigemi NMR microtubes (magnetic susceptibility matched to D2O) to maximise sample concentration. A ^{19}F edited ^{1}H spectrum is carried out to verify the regioselectivity of the reaction. The spectrum was recorded using 16384 scans, an acquisition time of 0.2 s and a pre-scan delay of 2 s giving a total experimental time of approximately 10 hours.

References: Chapter 6

1 Leonard, C. J., Aravind, L. & Koonin, E. V. Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily. *Genome research* **8**, 1038-1047 (1998).

2 Hunter, T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **80**, 225-236 (1995).

3 Nishi, H., Shaytan, A. & Panchenko, A. R. Physicochemical mechanisms of protein regulation by phosphorylation. *Frontiers in genetics* **5**, 270 (2014).

4 Waksman, G. *et al.* Crystal structure of the phosphotyrosine recognition domain SH 2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358**, 646-653 (1992).

5 Mandell, D. J. *et al.* Strengths of hydrogen bonds involving phosphorylated amino acid side chains. *Journal of the American Chemical Society* **129**, 820-827 (2007).

6 Hunter, T. Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 2513-2516, doi:10.1098/rstb.2012.0013 (2012).

7 Bah, A. *et al.* Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106-109 (2015).

8 Cohen, P. The regulation of protein function by multisite phosphorylation–a 25 year update. *Trends in biochemical sciences* **25**, 596-601 (2000).

9 Nishi, H., Hashimoto, K. & Panchenko, A. R. Phosphorylation in protein-protein binding: effect on stability and function. *Structure* **19**, 1807-1815 (2011).

10 Olsen, J. V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).

11 Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355-365 (2001).

12 Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279-3290 (2007).

13 Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098-1108 (2012).

14 Johansen, K. M. & Johansen, J. Regulation of chromatin structure by histone H3S10 phosphorylation. *Chromosome Research* **14**, 393-404 (2006).

15 Bradbury, E. M., Inglis, R. J., Matthews, H. R. & Sarner, N. Phosphorylation of Very - Lysine - Rich Histone in Physarum polycephalum. *European Journal of Biochemistry* **33**, 131-139 (1973).

16 Gurley, L. R., Walters, R. A. & Tobey, R. A. Cell cycle-specific changes in histone phosphorylation associated with cell proliferation and chromosome condensation. *The Journal of cell biology* **60**, 356 (1974).

17 Wei, Y., Yu, L., Bowen, J., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**, 99-109 (1999).

18 Clayton, A. L., Rose, S., Barratt, M. J. & Mahadevan, L. C. Phosphoacetylation of histone H3 on c - fos - and c - jun - associated nucleosomes upon gene activation. *The EMBO journal* **19**, 3714-3726 (2000).

19 Cheung, P. *et al.* Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Molecular cell* **5**, 905-915 (2000).

20 Gardner, K. E., Allis, C. D. & Strahl, B. D. Operating on chromatin, a colorful language where context matters. *Journal of molecular biology* **409**, 36-46 (2011).

21 Chen, Z. & Cole, P. A. Synthetic approaches to protein phosphorylation. *Current opinion in chemical biology* **28**, 115-122 (2015).

22 Chooi, K. P. *et al.* Synthetic phosphorylation of p38α recapitulates protein kinase activity. *Journal of the American Chemical Society* **136**, 1698 (2014).

23 Pearlman, S. M., Serber, Z. & Ferrell, J. E. A mechanism for the evolution of phosphorylation sites. *Cell* **147**, 934-946 (2011).

24 Berkowitz, D. B., Eggen, M., Shen, Q. & Shoemaker, R. K. Ready access to fluorinated phosphonate mimics of secondary phosphates. Synthesis of the (α, α-difluoroalkyl) phosphonate analogues of l-phosphoserine, l-phosphoallothreonine, and l-phosphothreonine. *The Journal of organic chemistry* **61**, 4666-4675 (1996).

25 Zhan, C., Varney, K., Yuan, W., Zhao, L. & Lu, W. Interrogation of MDM2 phosphorylation in p53 activation using native chemical ligation: the functional role of Ser17 phosphorylation in MDM2 reexamined. *Journal of the American Chemical Society* **134**, 6855 (2012).

26 Thomas, G. L. & Payne, R. J. Phosphate-assisted peptide ligation. *Chemical Communications*, 4260-4262 (2009).

27 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**, 6705-6710 (1998).

28 Tarrant, M. K. *et al.* Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. *Nature chemical biology* **8**, 262-269 (2012).

29 Hejjaoui, M. *et al.* Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: α-

synuclein phosphorylation at tyrosine 125. *Journal of the American Chemical Society* **134**, 5196 (2012).

30 Huse, M., Holford, M. N., Kuriyan, J. & Muir, T. W. Semisynthesis of hyperphosphorylated type I TGFβ receptor: addressing the mechanism of kinase activation. *Journal of the American Chemical Society* **122**, 8337-8338 (2000).

31 Bolduc, D. *et al.* Phosphorylation-mediated PTEN conformational closure and deactivation revealed with protein semisynthesis. *Elife* **2**, e00691 (2013).

32 Shogren-Knaak, M. A., Fry, C. J. & Peterson, C. L. A native peptide ligation strategy for deciphering nucleosomal histone modifications. *Journal of Biological Chemistry* **278**, 15744-15748 (2003).

33 Jbara, M., Maity, S. K., Morgan, M., Wolberger, C. & Brik, A. Chemical Synthesis of Phosphorylated Histone H2A at Tyr57 Reveals Insight into the Inhibition Mode of the SAGA Deubiquitinating Module. *Angewandte Chemie International Edition* **55**, 4972-4976 (2016).

34 Chiang, K. P., Jensen, M. S., McGinty, R. K. & Muir, T. W. A semisynthetic strategy to generate phosphorylated and acetylated histone H2B. *ChemBioChem* **10**, 2182-2187 (2009).

35 Zheng, W. *et al.* Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *Journal of Biological Chemistry* **280**, 10462-10467 (2005).

36 Chalker, J. M., Bernardes, G. a. J. & Davis, B. G. A "tag-and-modify" approach to site-selective protein modification. *Accounts of chemical research* **44**, 730-741 (2011).

37 Serwa, R. *et al.* Chemoselective Staudinger‐Phosphite Reaction of Azides for the Phosphorylation of Proteins. *Angewandte Chemie International Edition* **48**, 8234-8239 (2009).

38 Bernardes, G. J., 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).

39 Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post‐Translational Modifications. *Angewandte Chemie International Edition* **51**, 1835-1839 (2012).

40 Arslan, T., Mamaev, S. V., Mamaeva, N. V. & Hecht, S. M. Structurally modified firefly luciferase. Effects of amino acid substitution at position 286. *Journal of the American Chemical Society* **119**, 10877-10887 (1997).

41 Rothman, D. M. *et al.* Caged phosphoproteins. *Journal of the American Chemical Society* **127**, 846-847 (2005).

42 Sauerwald, A. *et al.* RNA-dependent cysteine biosynthesis in archaea. *Science* **307**, 1969-1972 (2005).

43 Park, H.-S. *et al.* Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **333**, 1151-1154 (2011).

44 Rogerson, D. T. *et al.* Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nature chemical biology* **11**, 496-503 (2015).

45 Heinemann, I. U. *et al.* Enhanced phosphoserine insertion during Escherichia coli protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS letters* **586**, 3716-3722 (2012).

46 Lee, S. *et al.* A facile strategy for selective incorporation of phosphoserine into histones. *Angewandte Chemie International Edition* **52**, 5771-5775 (2013).

47 Zippo, A. *et al.* Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* **138**, 1122-1136 (2009).

48 Klingberg, R. *et al.* Analysis of phosphorylation-dependent protein–protein interactions of histone H3. *ACS chemical biology* **10**, 138-145 (2014).

49 Kunowska, N., Rotival, M., Yu, L., Choudhary, J. & Dillon, N. Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. *Nucleic acids research*, gku1350 (2015).

50 Lercher, L. *Chemical tools for the study of epigenetic mechanisms* DPhil thesis, University of Oxford, (2014).

51 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Chromatin Protocols*, 1-16 (1999).

52 Tedder, J. M. & Walton, J. C. The importance of polarity and steric effects in determining the rate and orientation of free radical addition to olefins: Rules for determining the rate and preferred orientation. *Tetrahedron* **36**, 701-707 (1980).

53 Lequeux, T. *et al.* Sulfanyl-and selanyldifluoromethylphosphonates as a source of phosphonodifluoromethyl radicals and their addition onto alkenes. *Organic letters* **3**, 185-188 (2001).

54 Pignard, S., Lopin, C., Gouhier, G. & Piettre, S. R. Phosphonodifluoromethyl and phosphonothiodifluoromethyl radicals. Generation and addition onto alkenes and alkynes. *The Journal of organic chemistry* **71**, 31-37 (2006).

55 De Schutter, C., Pfund, E. & Lequeux, T. Radical conjugate addition of ambiphilic fluorinated free radicals. *Tetrahedron* **69**, 5920-5926 (2013).

56 Wigle, T. J. *et al.* Screening for Inhibitors of Low-Affinity Epigenetic Peptide-Protein Interactions: An AlphaScreen™-Based Assay for Antagonists of Methyl-Lysine Binding Proteins. *Journal of biomolecular screening* **15**, 62-71 (2010).

57 Marsh, E. N. G. & Suzuki, Y. Using 19F NMR to probe biological interactions of proteins and peptides. *ACS chemical biology* **9**, 1242-1250 (2014).

58 Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C. & Szyperski, T. Protein NMR spectroscopy in structural genomics. *Nature Structural & Molecular Biology* **7**, 982-985 (2000).

59 McFarland, J. M., Joshi, N. S. & Francis, M. B. Characterization of a three-component coupling reaction on proteins by isotopic labeling and nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **130**, 7639-7644 (2008).

60 Kawamoto, A. M. & Campbell, M. M. A new method for the synthesis of a phosphonic acid analogue of phosphoserine via a novel 1,1-difluorophosphonate intermediate. *Journal of Fluorine Chemistry* **81**, 181-186 (1997).

61 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

Chapter 7: 'Photoredox mutagenesis' for radical-mediated protein modification

Introduction

The previous chapters have demonstrated the utility of radical additions to dehydroalanine residues on proteins. Although undeniably powerful, the described method, relying on borohydride-mediated reaction of Dha with alkyl radicals generated from alkyl halides, is not without certain limitations. The 'dual' character of borohydride in aqueous solution (acting as single-electron initiator or mediator[1] while also acting as a nucleophilic hydride source[2-4]) limits conversions in the case of bulky or electronically disfavoured alkyl halides, as reduction of Dha to alanine becomes a competing reaction. Further, to date, the substrate scope of the borohydride chemistry appears limited to bromo- and iodide reagents. This can be problematic in the cases of unstable halides such as the urea derivatives used for installation of citrulline residues, which readily cyclise in solution. We hypothesized that alternative means of radical generation could allow expansion of the radical donors available for protein modification and potentially circumvent the reductive conversion to alanine.

The application of radical chemistry to synthetic problems has undergone a renaissance in recent years[5-8]. This has been driven, in part, by the ready accessibility of metal complexes capable of driving single-electron transfer processes under mild conditions, using visible wavelength light at ambient temperature. One area poised to benefit from these developments is protein chemistry. The stringent limitations on reaction design for protein substrates, imposed by the necessity of maintaining near-physiological conditions of solvent, temperature and pH, in addition to the need for exquisite chemoselectivity in order to avoid unintended modification of protein nucleophiles, have inspired interest in single-electron chemistry for protein modification[9].

Yet perhaps surprisingly, the use of visible light photoredox catalysts for protein modification has not been widely explored. In an early report, the generation of tyrosyl radicals by $[Ru(bpy)_3]^{2+}$ in the presence of ammonium persulfate (APS) and visible light allowed a novel approach to protein-crosslinking[10]. Tyrosyl radicals were proposed to be trapped by adjacent tyrosine residues, in an arene coupling, or by nucleophilic residues such as cysteine and lysine, accounting for the observed protein homo- and hetero-multimers observed upon irradiation with visible light. Sato and Nakamura later reported a ligand-directed modification of this method[11]. Tethering the photoredox catalyst to protein target-selective ligands enabled selective generation of tyrosyl radicals that were productively captured in an oxidative radical coupling by designed trapping agents. The ligand-directed nature of the photoredox catalyst allowed selective labelling of the target protein in cell lysate.

Chromophore-assisted light inactivation (CALI) has been advanced as a rapid method for modulation of target protein activity within cells[12]. In CALI, a chromophore generates singlet oxgygen upon irradiation. The short-lived nature of singlet oxygen allows selective oxidation, and thus inactivation, of protein species in the vicinity of the chromophore. $Ru(bpy)_3^{2+}$ has been reported as a uniquely efficient photocatalyst for CALI[13]. Indeed, suitable $Ru(bpy)_3^{2+}$ complexes can be rendered cell permeable, allowing targeted inactivation of proteins in living cells. The mechanism of inactivation in such experiments was clarified by identification of oxidation at a range of residues of targeted proteins including methionine, histidine and tryptophan. Further, it was demonstrated that protein knockdown by CALI can be regulated by external radical trapping agents, which instead promote the above-described tyrosyl radical 'labelling' pathway, highlighting $Ru(bpy)_3^{2+}$ complexes as bifunctional protein modification catalysts[11].

Recently, the Macmillan group reported an approach to peptide macrocyclisation using photoredox catalysis[14]. In the reported method,

single-electron transfer from iridium photoredox catalyst [Ir{dF(CF_3)ppy}$_2$(dtbpy)]PF_6 **1** induces decarboxylation of the C-terminus of a linear peptide, generating peptidic α-amino radicals that can be trapped by N-terminal acrylamides, thus effecting macrocyclisation. However, similar reactions utilising intact proteins as substrates have not yet been reported.

Dehydroalanine as 'tag' for photoredox mediated side-chain functionalisation

Inspired by these examples, and the previously described utility of protein dehydroalanine as a 'SOMO'-phile 'tag' for radical modification[9,15,16], potential photoredox transformations were sought that could complement and improve the existing chemical mutation reaction or deliver access to novel side-chain motifs that could prove useful in chemical biology. In this chapter, two candidate transformations are explored that may meet these criteria. Alkyl radical generation by single electron oxidation of alkyl trifluoroborates[17,18] is shown to enable a photoredox variant of the Dha→Leu chemical mutation. Further, nucleophilic α-aminoalkyl radicals[19] are shown to be a novel class of side-chain 'donor' reagents for facile radical modification at dehydroalanine.

Results

Design

Following preliminary experiments, two candidate transformations were chosen to illustrate the utility of photoredox catalysts in protein modification (**Figure 7.1**). Oxidation of alkyl (or alkoxy-) trifluoroborates or dimethylaniline species by the excited state of Ir[dF(CF_3)ppy]$_2$(dtbbpy)PF_6 **1** (**Figure 7.2**) should generate nucleophilic alkyl or α-amino alkyl radicals, respectively, that should react readily with protein-bound 'SOMOphiles' such as dehydroalanine. Whilst mechanistically similar, these reactions offer access to usefully complimentary side-chain classes.

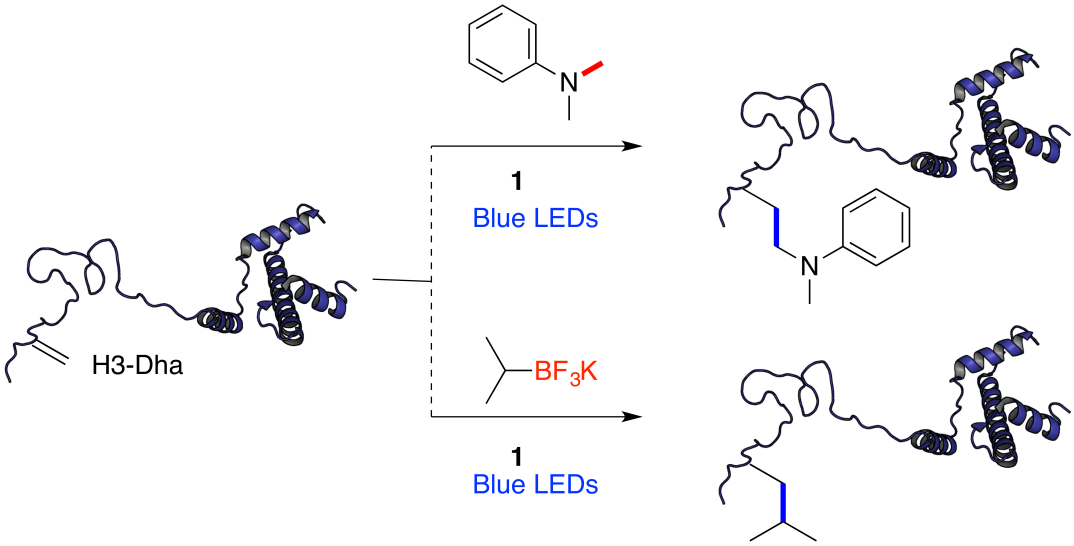


Figure 7.1: Proposed photoredox transformations for the modification of proteins at dehydroalanine. Bold blue line represents newly-constructed bond. Red represents radical-reactive moiety.

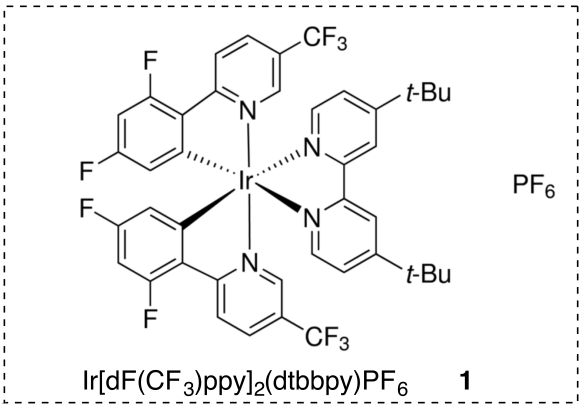


Figure 7.2: Ir-complex **1**

Alkyl trifluoroborates

It was hypothesized that oxidative methods for the generation of radical intermediates could circumvent the stability and reactivity problems associated with the use of organohalides as precursors for protein modification. Consideration of solubility, commercial availability and atom-economy led to the identification of potassium trifluoroborates as potential 'donors' for chemical mutagenesis[20]. The single-electron oxidation of secondary and benzylic trifluoroborates by transition metal photocatalysts, generating alkyl radicals, has been reported by the Akita[21,22] and Molander[18] groups; employing the generated radicals in Giese-type conjugate additions has also been reported[22]. However, peptide or protein

variants of these reactions under aqueous conditions have not yet been reported.

Consideration of redox potentials led to selection of iridium complex $Ir[dF(CF_3)ppy]_2(dtbbpy)PF_6$ [23] **1** as photocatalyst for initial protein experiments. The excited state complex of this variant is the most strongly oxidizing of the readily available iridium photocatalysts and should thus have the widest scope for oxidative generation of radicals[24]. The use of $Ru(bpy)_3Cl_2.6H_2O$ was also explored as, despite decreased oxidizing potential, the improved aqueous solubility and affordability of this catalyst could support wider protein substrate scope and uptake by non-specialist researchers.

As side-chain donor, initial experiments utilized the commercially available potassium isopropyltrifluoroborate, which, as a source of isopropyl radicals, should react with dehydroalanine, constituting a 'chemical mutation' to leucine (**Figure 7.3**). The analogous reaction utilizing isopropyl iodide as a source of alkyl radicals has been extensively discussed in previous chapters. Initial experiments employing a large molar excess of **1** and isopropyl trifluoroborate, relative to histone H3-K9Dha as model protein, under anaerobic conditions and using a simple desktop blue LED lamp, gave clear indication of the desired reactivity as judged by LC-MS after four hours reaction time (**Figure 7.3**). Additional side-products resulting from radical addition (dialkylation and protein cleavage at the reaction site) were also detected in these initial experiments (**Figure 7.3**). Control reactions confirmed the importance of both light and photocatalyst **1** for successful reaction (**Table 1**, see **Experimental Section**). In line with the designed oxidative manifold, no reduction of dehydroalanine was observed by chemical probing with thiol nucleophiles[25] (see **Experimental Section**). Pleasingly, despite the oxidative conditions, no above-background oxidation of the two H3 methionine residues was observed (see **Experimental Section**). Notably, the loading of isopropyl trifluoroborate could be reduced significantly, to only 125 equivalents relative to protein,

albeit with longer reaction times. No organic co-solvent was required, beyond that necessary to prepare a stock solution of **1.** The observation of protein cleavage within a strictly anaerobic environment is striking and suggests an alternate cleavage mechanism to the oxidative cleavage observed in Chapter 3. Preliminary results indicate that cleavage can be avoided by the use of higher starting concentrations of isopropyl trifluoroborate, although with the risk of increased di-alkylation products. Further understanding of the cleavage mechanism should enable further development of this reaction, perhaps through the incorporation of redox active additives, into a practical method for chemical mutation.

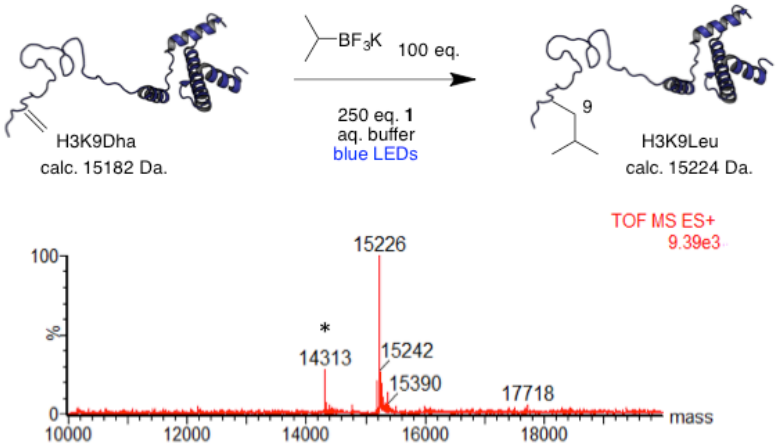


Figure 7.3: Chemical mutation Dha→ Leu using photoredox catalyst **1**. Cleavage product is highlighted by *.

1	***i*-PrBF$_3$K**	**Light**	**Reaction**
No	Yes	Yes	No
Yes	No	Yes	No
Yes	Yes	No	No

Table 1: Control reactions

Alternative alkyl trifluroborates were examined as potential substrates in the protein oxidative Giese-type addition to dehydroalanine (**Figure 7.4,** also see **Experimental Section**). Unfortunately, under the examined

reaction conditions, none of the trifluoroborates tested gave promising results. Benzyl trifluoroborate, intriguingly, gave ~20% conversion to a di-alkylated product when reacted with **1** and histone H3-S10Dha. The di-alkylation is presumed to take place at Dha although benzyl radicals may also react with other amino acids.

In Chapter 4 benzyl radicals were also shown to be poor substrates for mono-addition to dehydroalanine. Cyclohexyl trifluoroborate also gave poor results, with <10% conversion to the desired product being observed under any conditions.

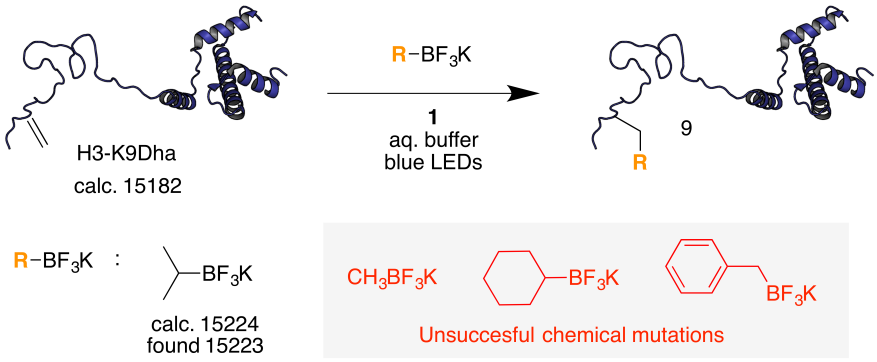


Figure 7.4: Side-chain scope for chemical mutation *via* oxidative generation of radicals from alkyl trifluoroborates.

Addition of α-aminoalkyl radicals to dehydroalanine residues on histone H3

α-aminoalkyl radicals are unusual intermediates, accessed via single electron oxidation of activated (typically aromatic) tertiary amines[26], which enable functionalization of an sp^3 C–H bond adjacent to the amine. α-aminoalkyl radicals are regarded as nucleophilic and have been shown to engage in conjugate additions to Michael acceptors[19,27], although reaction with dehydroalanine, or indeed on peptide or protein substrates, has not been previously reported.

Preliminary experiments involving irradiation of a buffered aqueous solution of dimethylaniline **2**, Ir-complex **1** and H3-K9Dha yielded no

evidence of product formation by LC-MS/MS. It was speculated that the poor aqueous solubility of either **2** or Ir-complex **1** could hinder successful reaction. Indeed, addition of organic solvent to the reaction mixture proved immediately fruitful; irradiation of a H3-K9Dha solution containing a large molar excess of dimethylaniline[19] **2** and Ir-complex **1** with blue LEDs led to observation of ~40% of the expected product in the presence of 20**%** acetonitrile (**Figure 7.5a**).

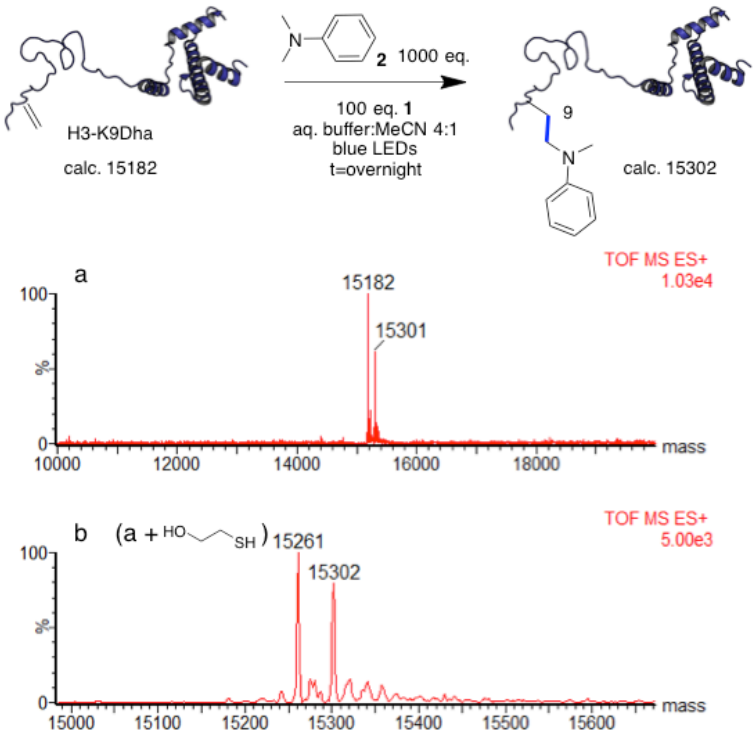


Figure 7.5: Reaction of H3-Dha with dimethylaniline **2** in the presence of light and catalyst **1** affords ~40% conversion when employing MeCN as co-solvent. a. LC-MS of reaction mixture. b. above reaction mixture treated with β-mercaptoethanol at pH 8.0. The shift in the peak corresponding to H3-Dha indicates no reduction to alanine has occurred, while the lack of a shift for the product peak indicates the photoredox reaction has occurred at the Dha site.

Chemical probing confirmed the site-selectivity of the conjugate addition – no residues other than the expected dehydroalanine were modified (**Figure 7.5b**). Further, no reduction to alanine was observed. Switching to DMF as co-solvent led to improved yields of modified protein and shorter reaction times, enabling >85% conversions. Control reactions demonstrated the dependence of the transformation on Ir-complex **1**

(**Experimental Section**). Notably, attempting the reaction outside of the glovebox resulted in protein degradation, presumably due to the generation of peroxyl-type radical species in an aerobic solution[28] (**Experimental Section**).

$Ru(Bpy)_3Cl_2.H_2O$ is a cheap, readily accessible and water soluble transition-metal photoredox catalyst that has previously been employed both for α-aminoalkyl radical generation[27] and in protein modification[11] and chemical biology[28]. Substitution of $Ru(bpy)_3Cl_2.6H_2O$ for **1** could enable greater protein substrate scope through use of a less harshly oxidizing[24], and fully water soluble, catalyst. Initial attempts using $Ru(bpy)_3Cl_2.6H_2O$ for α-aminoalkyl radical addition to histone-Dha, employing strictly aqueous conditions, were notably unsuccessful. However, reaction did proceed in the presence of DMF. This effect of co-solvent, despite the full aqueous solubility of the Ru-complex, may imply that DMF can act as a sensitizer or redox mediator[29] in this reaction. Indeed, the addition of 100 equivalents (relative to protein, 0.1 relative to dimethylaniline) tetramethylethylenediamine (TMEDA, previously[30] employed as a redox mediator in photoredox catalysis) to the reaction mixture enabled conversion under fully aqueous conditions, albeit at reduced rates (**Figure 7.6a**).

The use of 'catalytic' (sub-stoichiometric relative to protein) loading of reagents is rare in protein chemistry, where the low concentrations of protein substrates employed often require large excesses of reagent to force reactions to full conversion in realistic time-frames. Pleasingly, the use of TMEDA in conjunction with DMF enabled the lowering of the $Ru(bpy)_3Cl_2.6H_2O$ loading to just 0.5 equivalents relative to the protein substrate (**Figure 7.6b**).

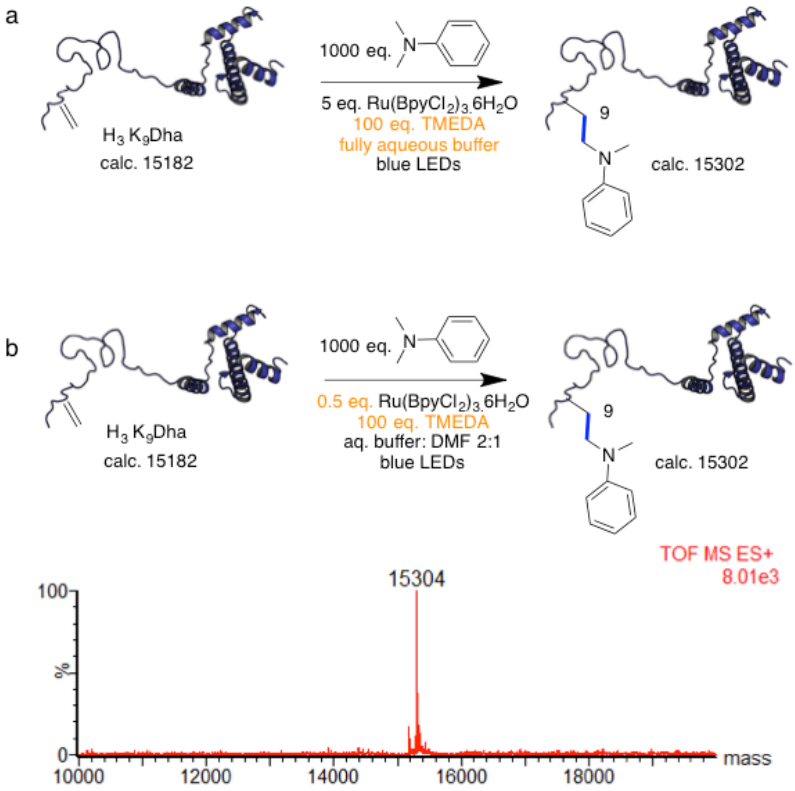


Figure 7.6: Optimised conditions for the photoredox Giese-type reaction of **2** with H3-Dha. b. Successful protein modification using catalytic photoredox catalyst.

Having identified a range of successful conditions with model substrate **2**, the substrate scope of the addition of α-amino radicals to Dha was examined (**Figure 7.7** and **Experimental Section**). Dimethylaminobenzoic acid reacted well using both **1** and $Ru(bpy)_3Cl_2.6H_2O$, affording 80% conversion to the conjugated product. Other dimethylanilines with electron-withdrawing substituents are likely to react well, although the electron-donating phenol in p-dimethylaminophenol precluded any reaction. Pleasingly, aromatic heterocycles such as DMAP could also be employed. DMAP has been used as an acylation catalyst for affinity labeling of proteins[31] – this direct method for its introduction to proteins may thus be of use in chemical biology.

Attention then turned to alkyl tertiary amines (**Figure 7.7**, see **Experimental Section** for full details). TMEDA itself, in the absence of competing dimethylamino-containing substrates, was readily capable of

addition to Dha under photoredox conditions. Reaction of (dimethylamino)propylamine was more challenging, with only ~15% conversion attainable after 48h of reaction using $Ru(bpy)_3Cl_2.6H_2O$ (5 eq.). Although switching to the strongly oxidizing **1** improved conversion, a number of additional peaks were seen in the LC-MS spectra, indicating competing reactivity. Reactions performed with (dimethylamino)acetone or dimethylaminopropanediol resulted in complex mixtures of products.

Activation of amines for single-electron oxidation has been reported via incorporation of adjacent sacrificial trifluoroborates[32] or carboxylic acids[33]. Indeed, this strategy has been shown to enable the use of secondary amines in reactions including the Giese-type addition here[33]. Experimenting with this approach, the reaction of N-phenylglycine with histone-Dha under various photoredox conditions was attempted. Unfortunately, no addition could be observed. Surprisingly, use of sarcosine in place of N-phenylgylcine provided immediate results, with the dimethylamine-conjugated product readily observed. Optimisation by incorporation of small amounts (<50 eq., TMEDA addition is competitive at higher concentrations) TMEDA improved conversions to around 70%, with ~20% TMEDA addition still observed. Further optimization of this reaction ands its extension to other α-amino acids may provide a route towards a vastly expanded substrate scope for this methodology.

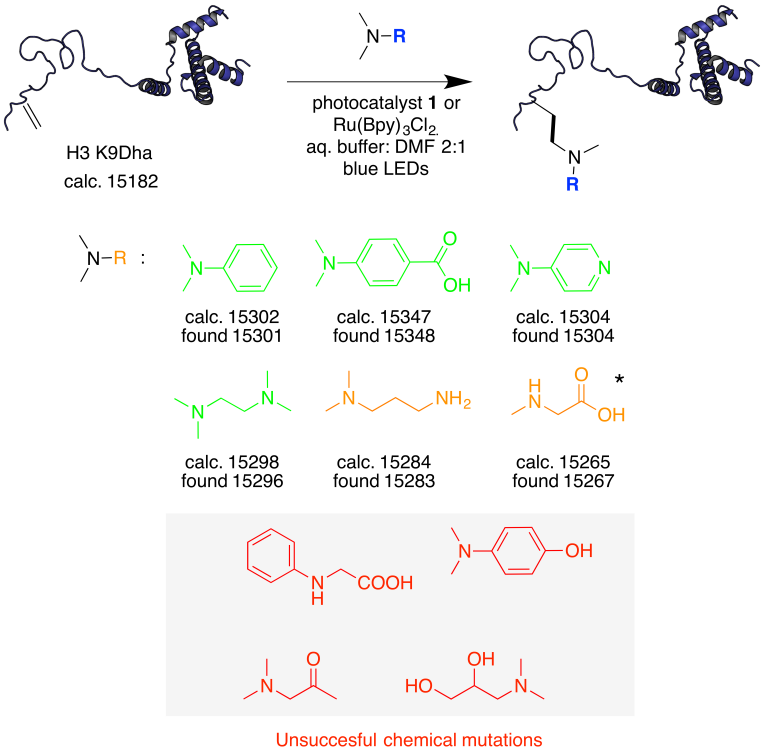


Figure 7.7: Substrate scope for the photoredox Giese-type reaction of tertiary amines with histone-Dha. Green = successful substrates (>80% conversion with minimal side-products). Orange = Promising substrates. Red = Unsuccessful reactions (no reaction or only side-products detected).

Discussion and future prospects

In this chapter, two novel classes of side-chain 'donor' were explored for radical modification at dehydroalanine; alkyl trifluoroborates and dimethylamino-containing compounds. The oxidative generation of alkyl radicals from alkyl trifluoroborates, mediated by photoredox catalyst **1**, was shown to be a viable strategy for chemical mutation to leucine from Dha. A novel protein cleavage reaction was identified as a significant side-reaction that remains to be overcome. Further, the substrate scope of this reaction to date remains limited to potassium isopropyltrifluoroborate; more oxidizing photocatalysts may enable extension to other substrates, although the enhanced risk of protein damage must be considered. Intriguingly, no protein cleavage was observed under the extremely similar reaction conditions employed for the Giese-type reaction of α-aminoalkyl radicals with protein-Dha. This fact is suggestive of dissimilarity in the

operative mechanism that may provide fruitful pathways for further optimization of the alkyl trifluoroborate reaction.

The reaction of α-aminoalkyl radicals with protein-Dha is more immediately promising. Full conversions could be obtained on model substrate dimethylaniline **2** using **1** or the milder $Ru(bpy)_3Cl_2.6H_2O$ as photocatalyst and TMEDA as redox mediator. With the developed conditions, extension of the scope of this reaction to other proteins and side-chains can be readily envisioned. While the γ-N linkage resultant from this reaction is not obviously conducive to installation of natural PTMs, applications in chemical biology can still be envisioned. For example, the direct conjugation of DMAP to histone-Dha was demonstrated using this system, suggesting immediate application in 'pull-down' experiments for 'reader' identification[31]. More speculatively, Nme_2-containing sugars could provide access to stable carba-mimetics of protein N-linked glycosylation.

Experimental Section: Chapter 7

Protein chemical modification

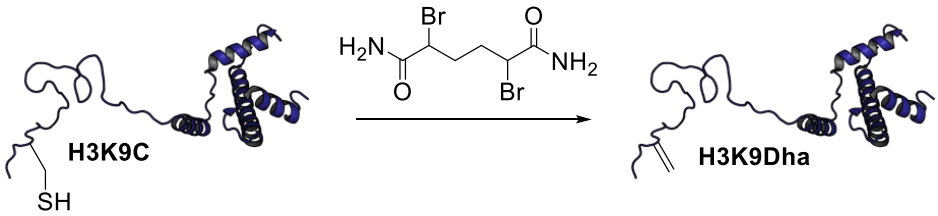
Histone H3 (K9C, C110A)

ARTKQTARCSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR
YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN
LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15214

Observed mass = 15214

Histone H3 K9Dha

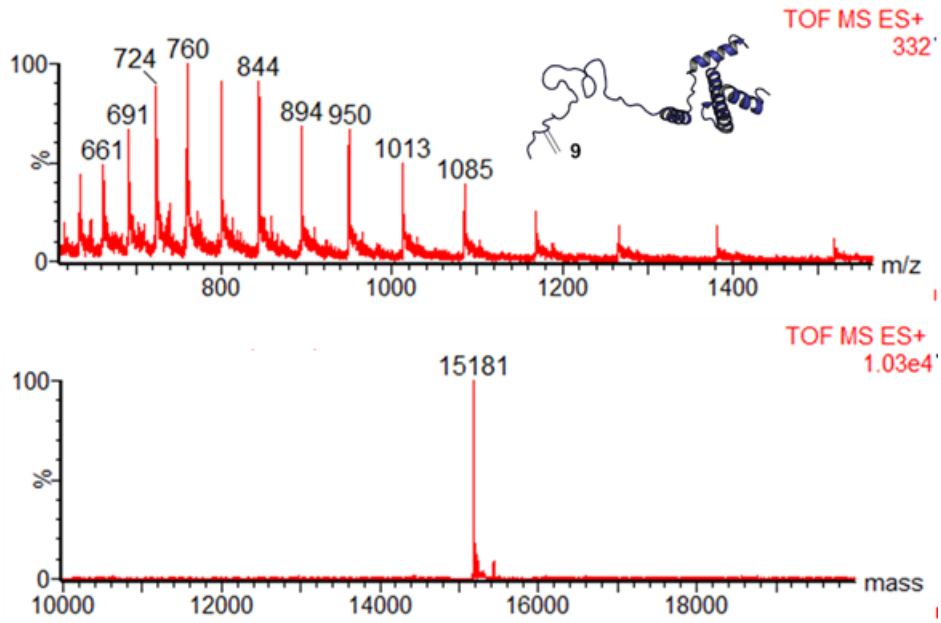


The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass = 15180.7

Observed mass = 15181



Histone H3 (S10C, C110A)

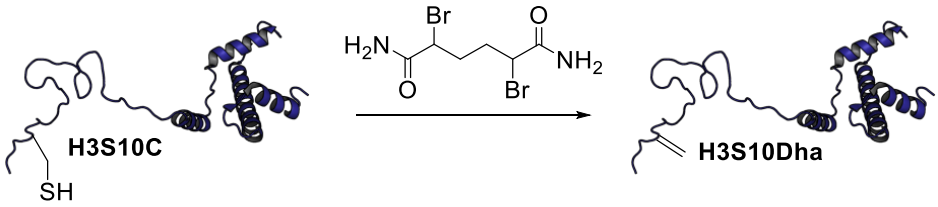
ARTKQTARK**C**TGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRR

YQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTN

LAAIHAKRVTIMPKDIQLARRIRGERA

Calculated mass = 15254.8
Observed mass = 15255

Dehydroalanine formation

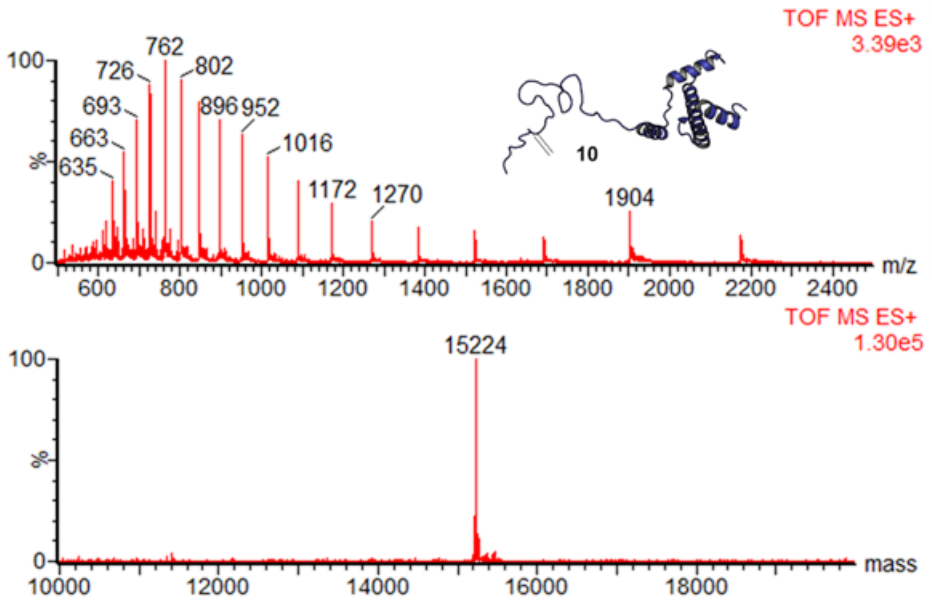


The dehydroalanine-containing histone was prepared according to general procedure **1.**

ESI-MS spectrum for the modified histone is shown below.

Calculated mass: 15220.7

Observed mass: 15224



Histone modification *via* oxidative generation of alkyl radicals from trifluoroborate salts (General procedure 7A)

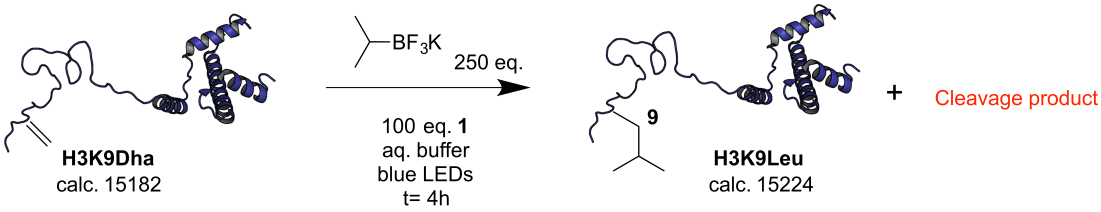
Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions for at least 6 hours (such a period can be reduced if the protein solution is sparged with argon before entry).

The trifluoroborate of interest (100-2000 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (Ir[dF(CF3)ppy]2(dtbbpy)PF6) was then added by pipette. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, the protein solution was removed from the glovebox and quenched via addition to, and elution (using the desired buffer) from a G25 size-exclusion spin column, as per

manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

Installation of leucine from dehydroalanine *via* oxidative generation of alkyl radicals from trifluoroborate salts

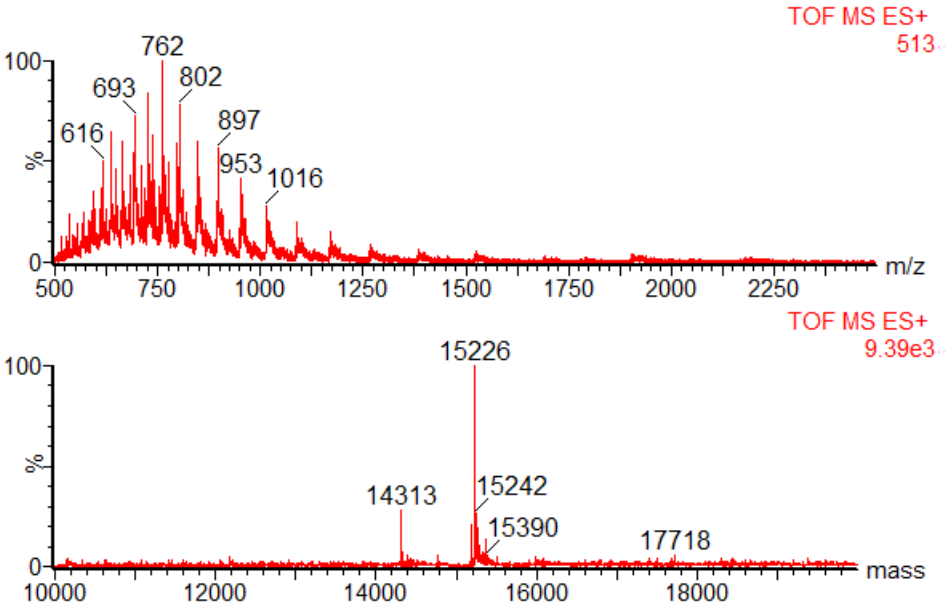
Initial experiments



Following general procedure **7A**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

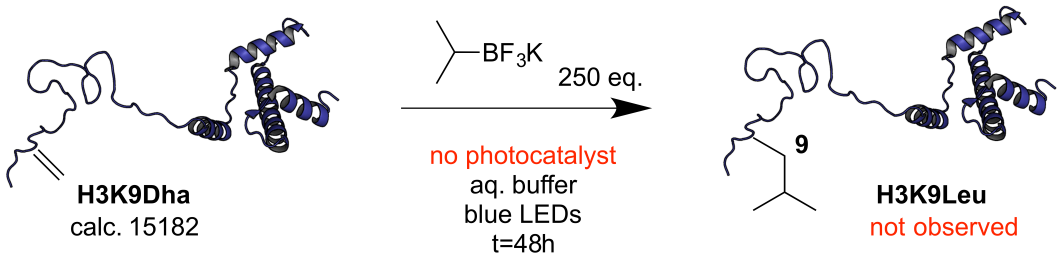
Potassium isopropyltrifluoroborate (0.5 mg, 250 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (4.3 µL of a 0.2M stock in DMF, 100 equiv. relative to protein) was then added by pipette. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, the protein solution was removed from the glovebox and quenched via addition to, and elution (using the desired buffer) from a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15224
Observed mass: 15226



Control reactions (Performed as above except for red)

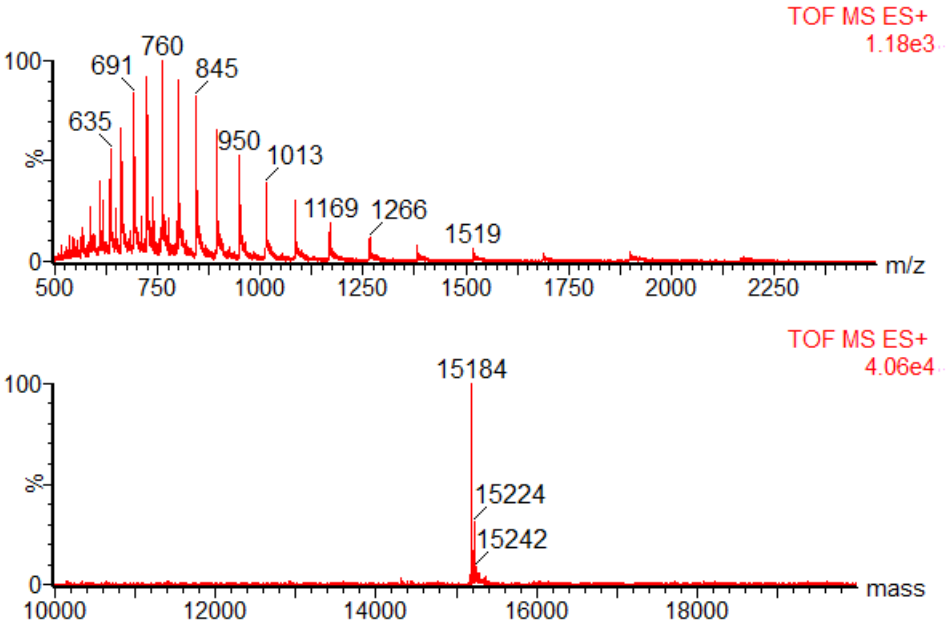
No photocatalyst



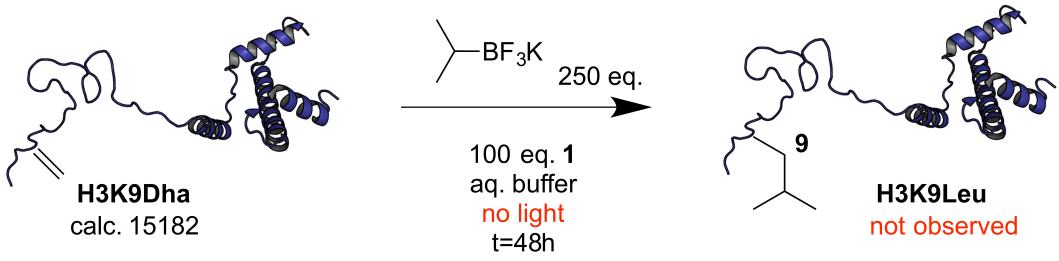
ESI-MS for the modified histone is shown below.

Calculated mass: 15224

Observed mass: 15184 (H3-K9Dha)

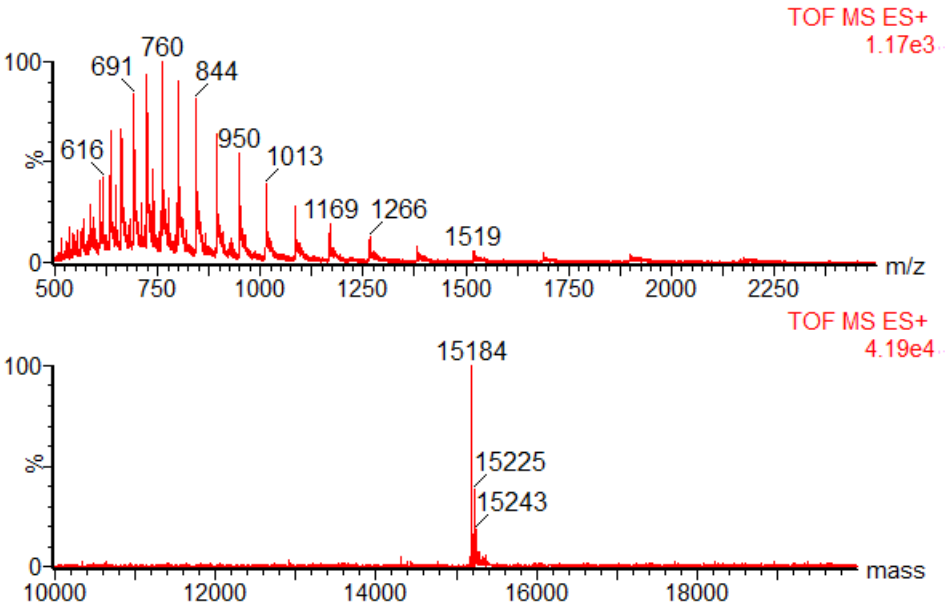


No light



ESI-MS for the modified histone is shown below.

Calculated mass: 15224
Observed mass: 15184 (H3-K9Dha)



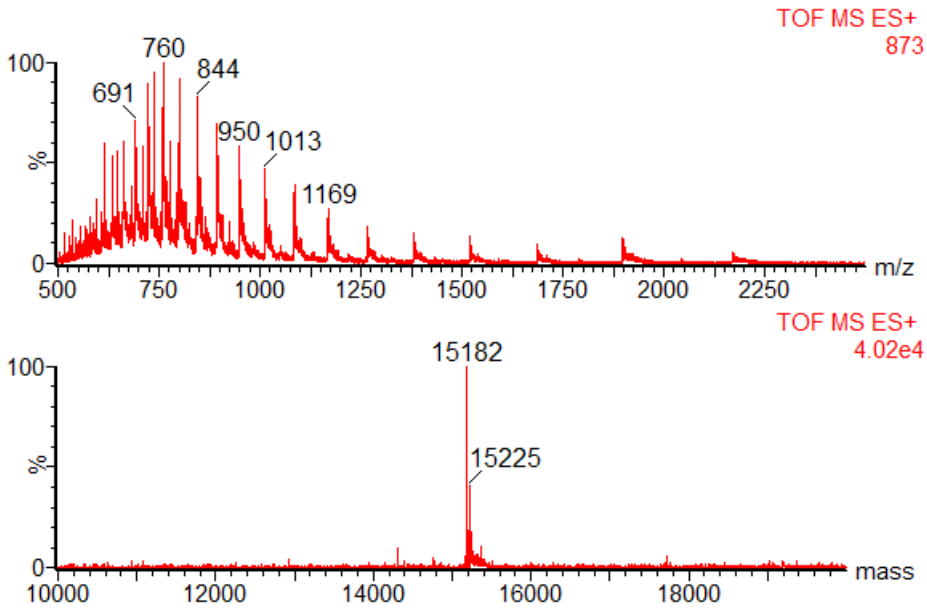
Chemical test for reduction to alanine and identification of site of reaction at dehydroalanine

Following general procedure **7A**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 μL) was added to an eppendorf tube, followed by 20 μL of 10 mg/mL Dha-protein stock solution (200 μg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 μM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Potassium isopropyltrifluoroborate (0.25 mg, 125 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (4.3 μL of a 0.2M stock in DMF, 100 equiv. relative to protein) was then added by pipette. The reaction was irradiated with blue LEDs for 2 hours and then removed from the glovebox and quenched via addition to, and elution (using pH 8.0 sodium phosphate buffer + 5M Gdn.HCl) from a G25 size-exclusion spin column, as per manufacturers instructions.
5 μL β-mercaptoethanol was added to the isolated protein and the resultant solution incubated for a further 1 hour at room temperature. The reaction was then analysed by LC-MS and stored at -20°C.

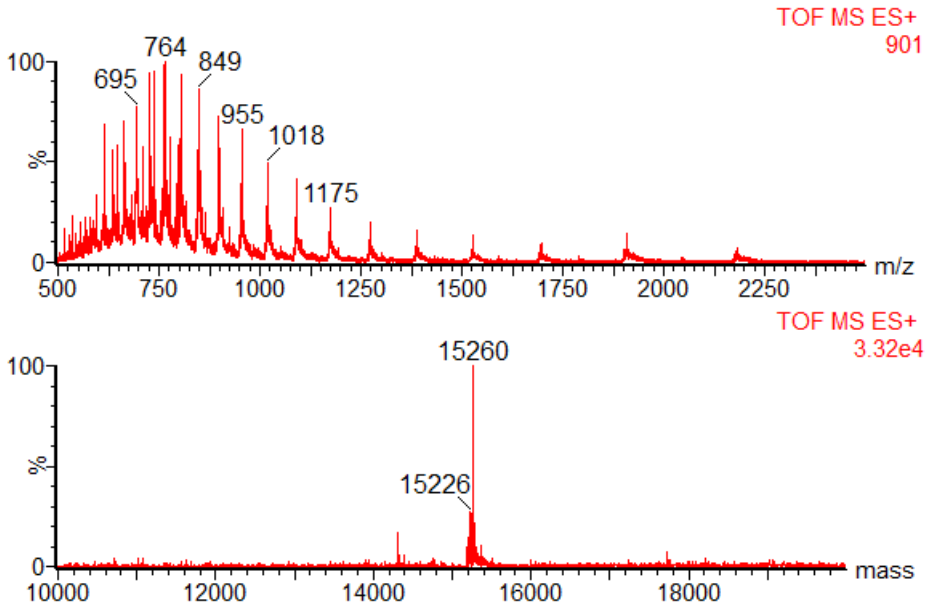
ESI-MS for the modified histone is shown below.
Calculated mass: 15224
Observed mass: 15182 (H3-K9Dha), 15225

t2h



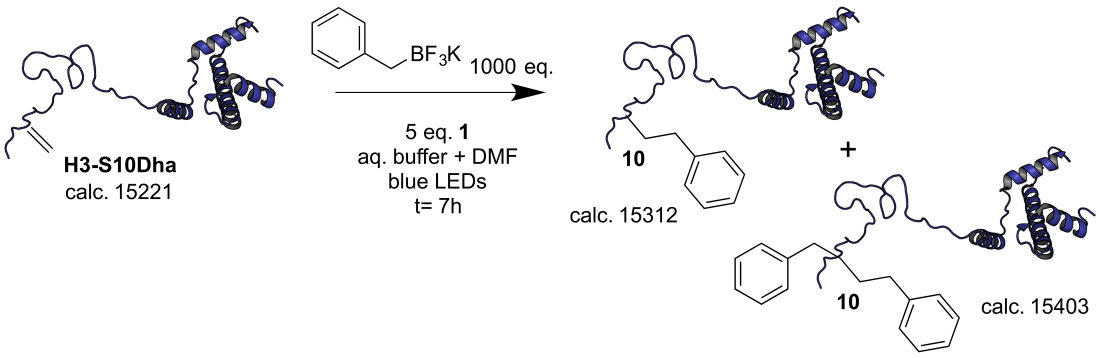
t=1h following addition of β-mercaptoethanol

Calculated mass: 15224 (H3-K9Leu), 15260 (+β mercaptoethanol addition)
Observed mass: 15225 (H3-K9Dha), 15260



Challenging reactions

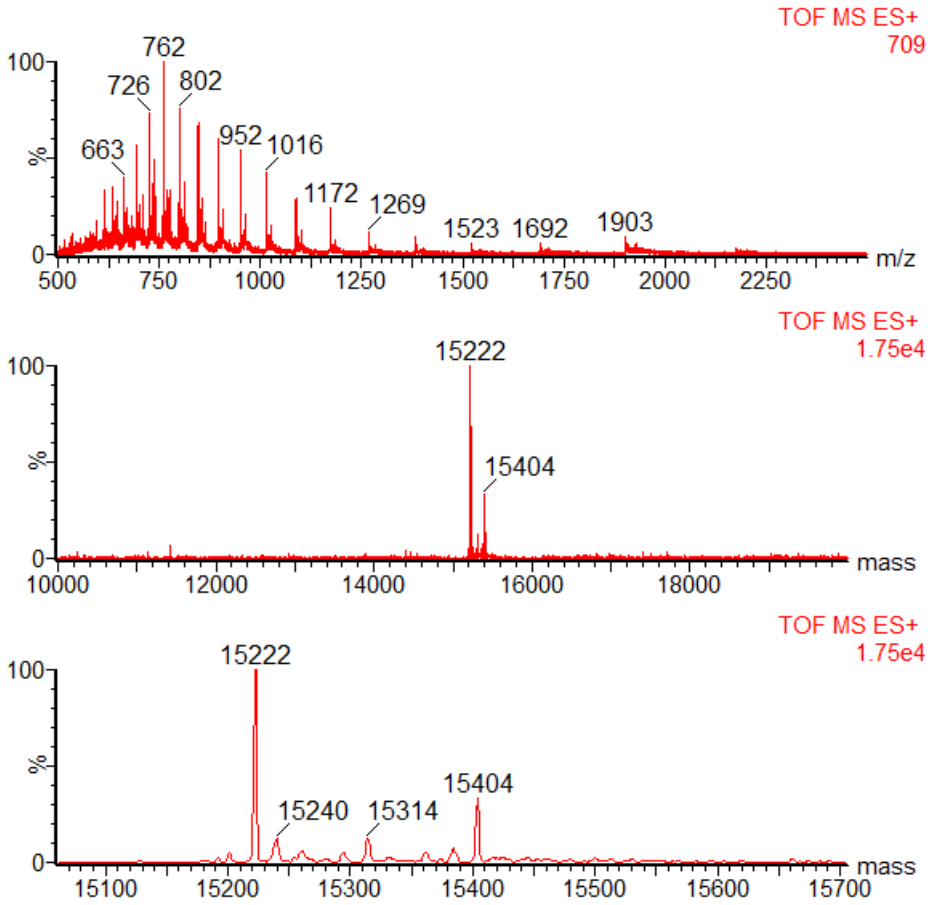
Reaction with potassium benzyltrifluoroborate



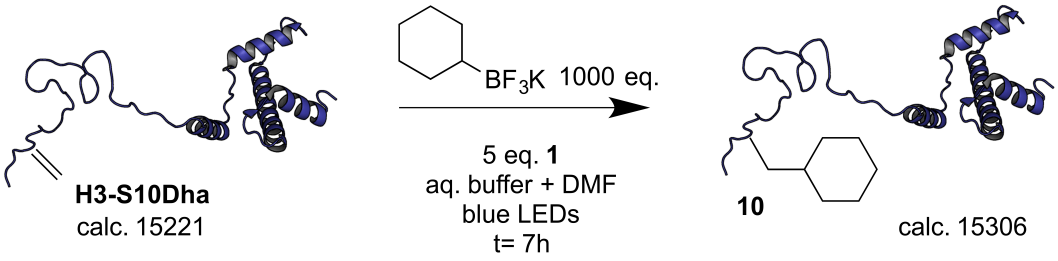
Following general procedure **7A**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Potassium benzylltrifluoroborate (2.6 mg, 1000 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (2.2 µL of a 20 mM stock in DMF, 5 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 7 hours, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15312 (H3-S10hPhe)
Observed mass: 15222 (H3-S10Dha), 15314, 15404


TOF MS ES+
709
100
%
0
762
726
802
663
952
1016
1172
1269
1523
1692
1903
500
750
1000
1250
1500
1750
2000
2250
m/z
TOF MS ES+
1.75e4
15222
15404
10000
12000
14000
16000
18000
mass
TOF MS ES+
1.75e4
15222
15240
15314
15404
15100
15200
15300
15400
15500
15600
15700
mass

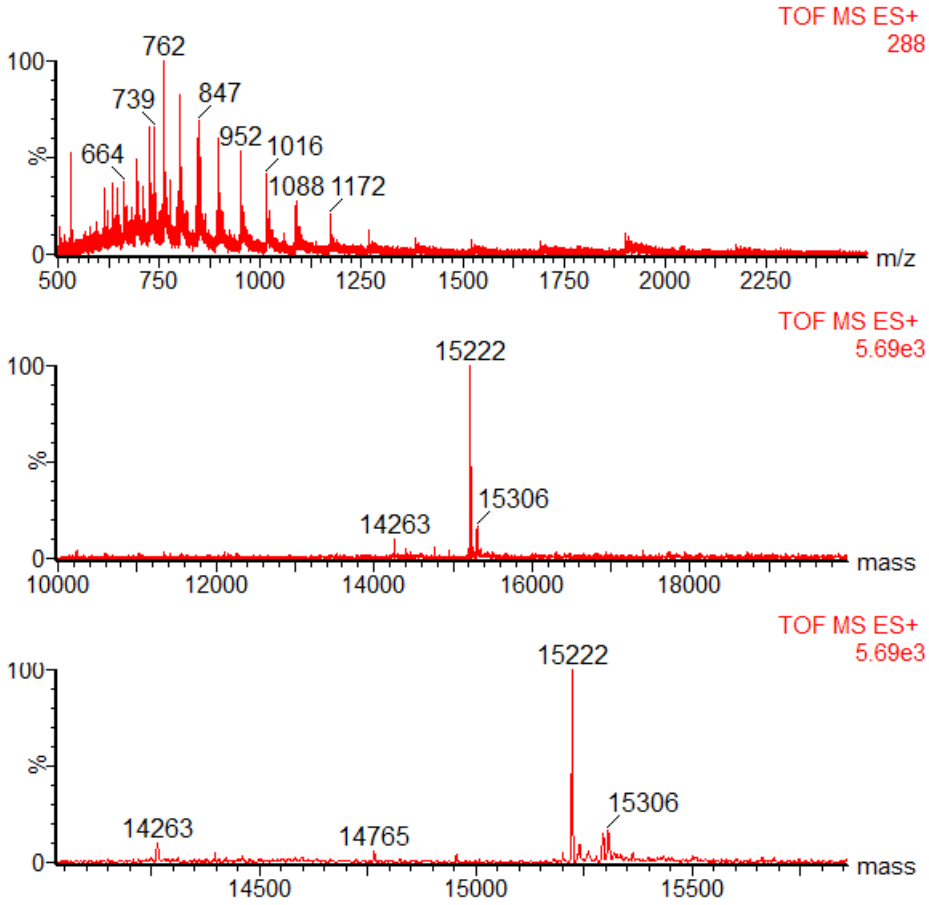
Reaction with potassium cyclohexyltrifluoroborate



Following general procedure **7A**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Potassium cyclohexyltrifluoroborate (2.5 mg, 1000 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (2.2 µL of a 20 mM stock in DMF, 5 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 7 hours, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15306 (H3-S10-cyclohexylAla)
Observed mass: 15222 (H3-S10Dha), 15306

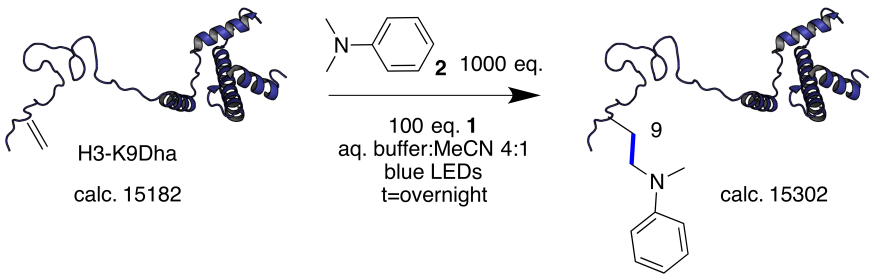

TOF MS ES+
288
100
%
0
762
739
847
952
1016
664
1088
1172
500
750
1000
1250
1500
1750
2000
2250
m/z
TOF MS ES+
5.69e3
100
%
0
15222
15306
14263
10000
12000
14000
16000
18000
mass
TOF MS ES+
5.69e3
100
%
0
15222
15306
14263
14765
14500
15000
15500
mass

Histone modification *via* alpha-amino radical conjugate addition (General procedure 7B)

Reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions for at least 6 hours (such a period can be reduced if the protein solution is sparged with argon before entry).

The dimethylamine of interest (500-2000 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** was then added by pipette. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, the protein solution was removed from the glovebox and quenched via addition to, and elution (using the desired buffer) from a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

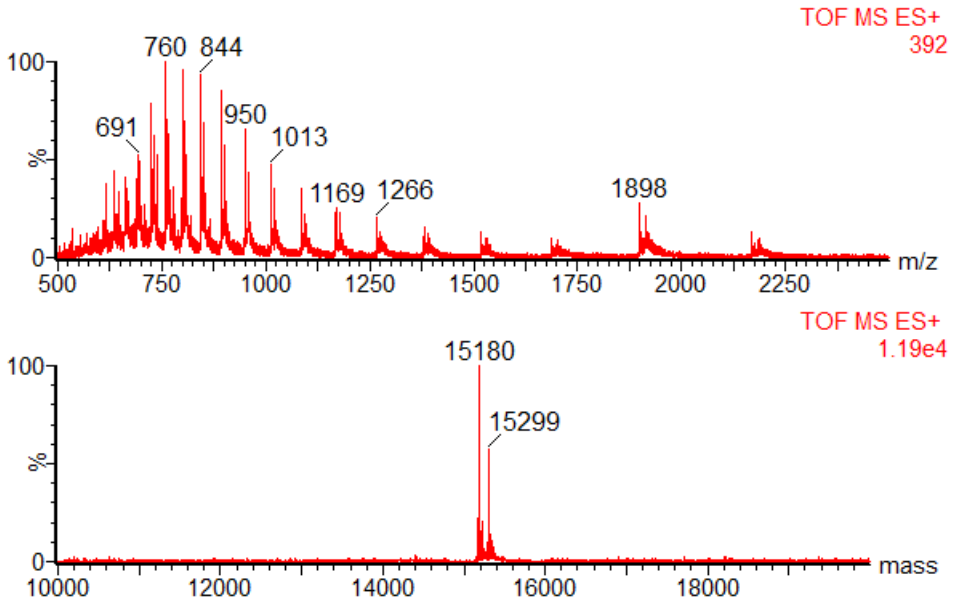
Initial investigations into the addition of dimethylaniline 2 to H3-K9Dha (Figure 5)



Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Dimethylaniline was ported into the glovebox and the appropriate volume (1.6 µL, 1000 molar equiv. relative to protein) dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (4.3 µL of a 0.2 M stock in DMF, 100 equiv. relative to protein) was then added by pipette. 50 µL degassed acetonitrile was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 24 hours, the protein solution was removed from the glovebox and quenched via addition to, and elution from (using pH 8.0 sodium phosphate buffer + 5M Gdn.HCl), a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

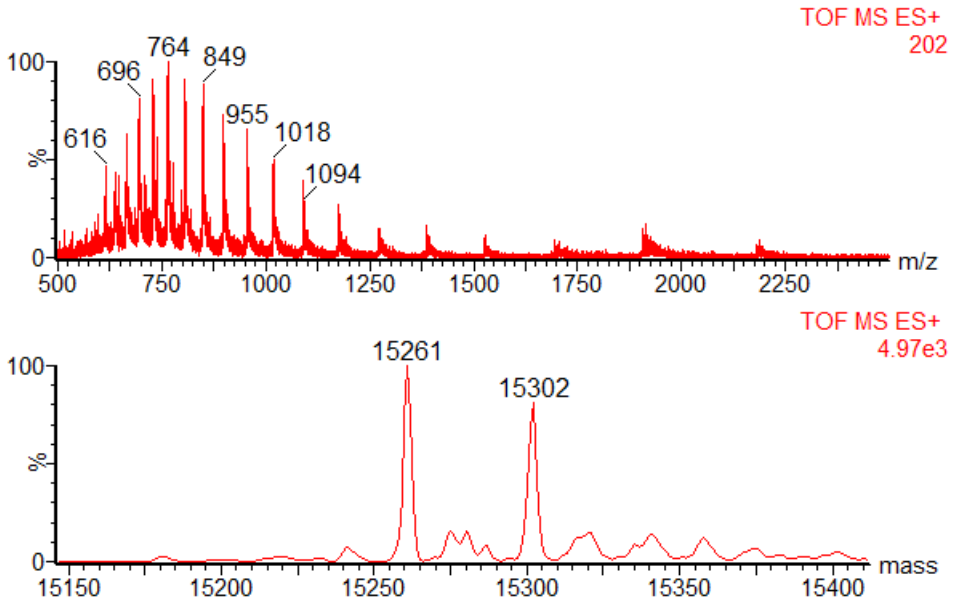
ESI-MS for the modified histone is shown below.
Calculated mass: 15302
Observed mass: 15180 (H3-K9Dha), 15299



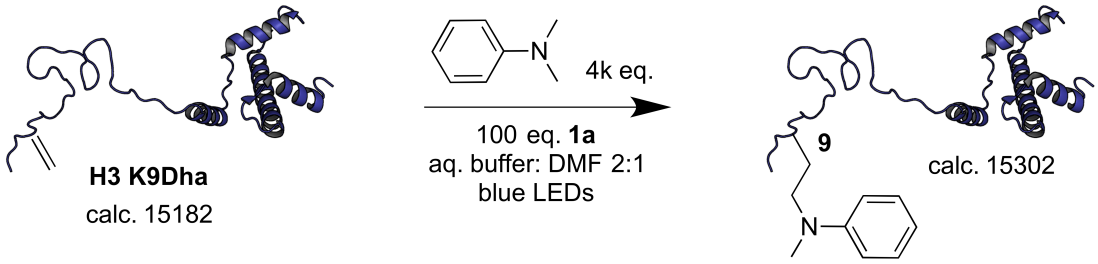
5 µL β-mercaptoethanol was added to the isolated protein and the resultant solution incubated for a further 1 hour at room temperature. The reaction was then analysed by LC-MS and stored at -20°C.

t=1h following addition of β-mercaptoethanol

Calculated mass: 15302 (product), 15260 (Dha+β mercaptoethanol addition)
Observed mass: 15261, 15302



Optimised conditions for the addition of dimethylaniline 2 to H3-K9Dha using 1

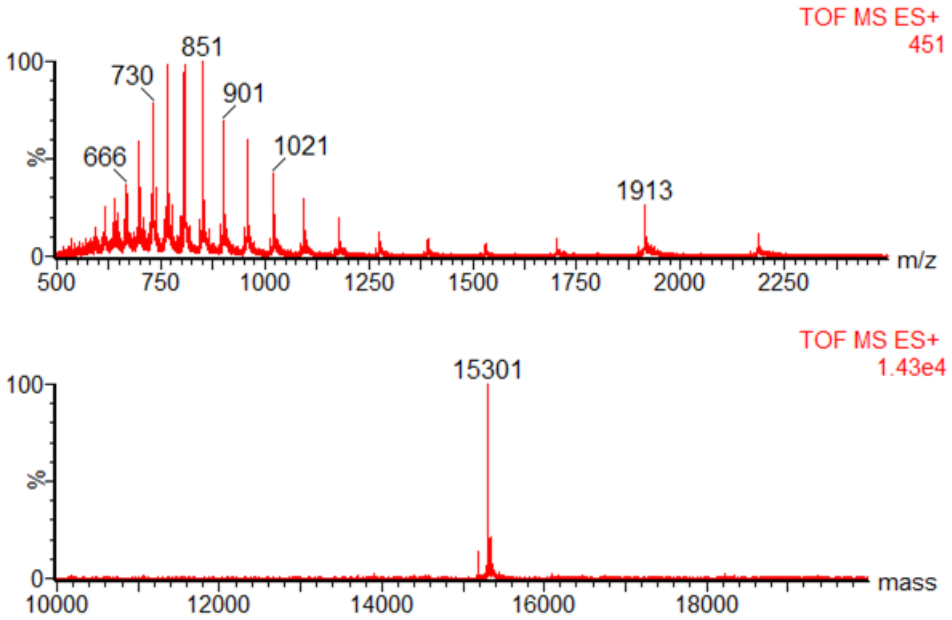


Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

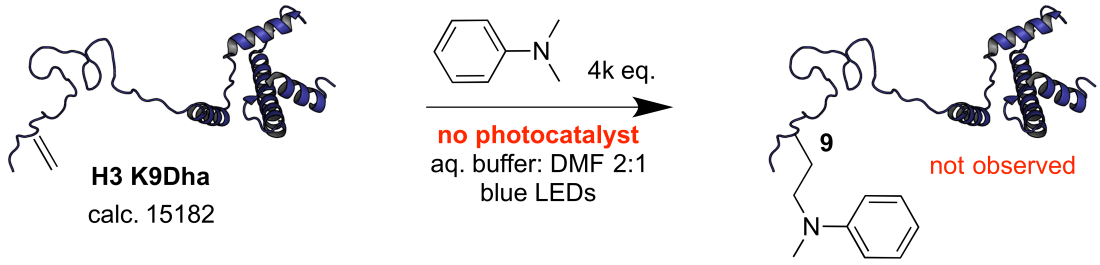
Dimethylaniline was ported into the glovebox and the appropriate volume (6.4 µL, 4000 molar equiv. relative to protein) dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (4.3 µL of a 0.2 M stock in DMF, 100 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 1 hour, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

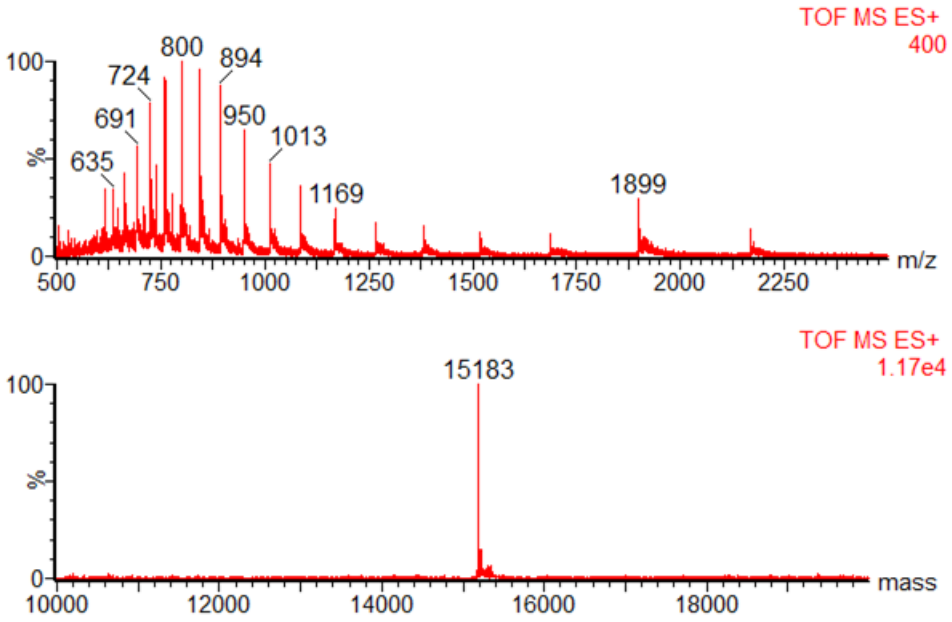
ESI-MS for the modified histone is shown below.
Calculated mass: 15302

Observed mass: 15301

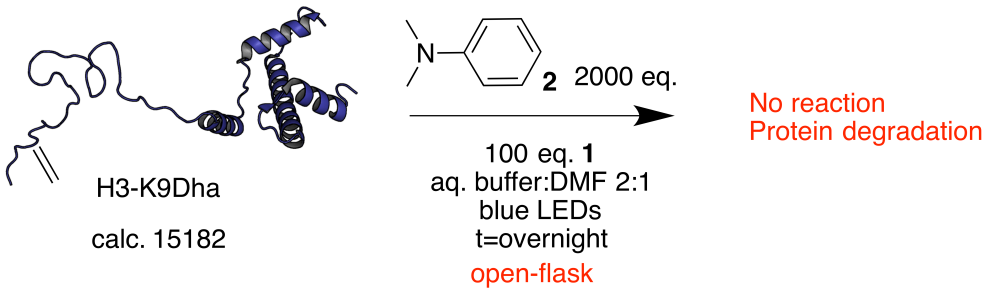


Control reaction no photocatalyst

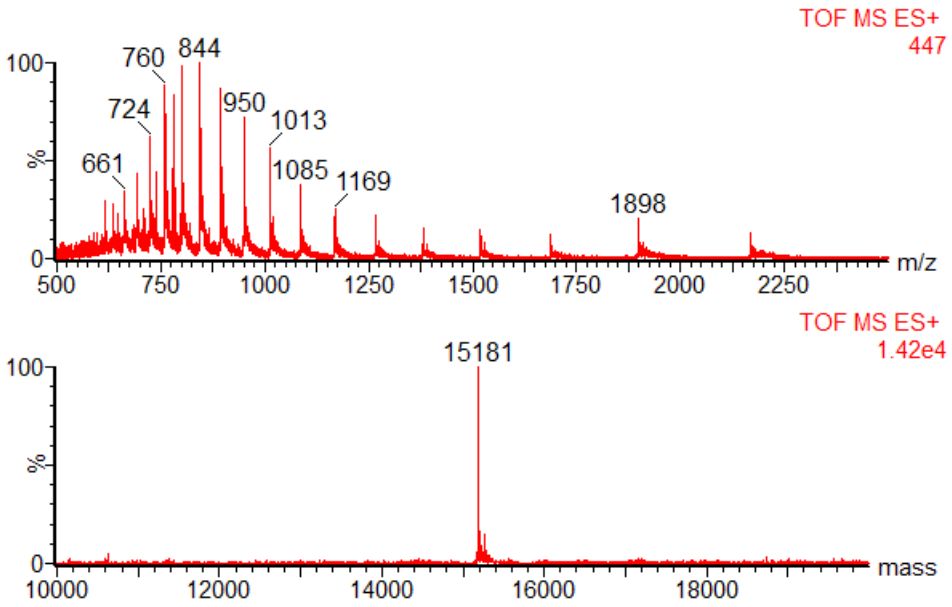




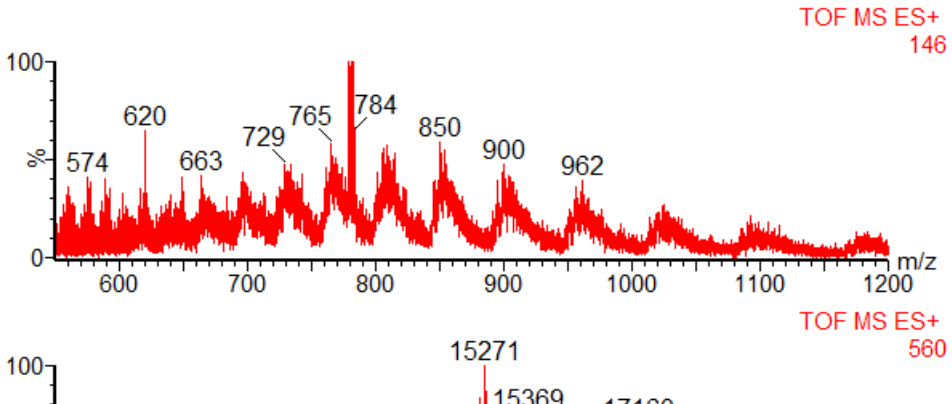
Control reaction under air

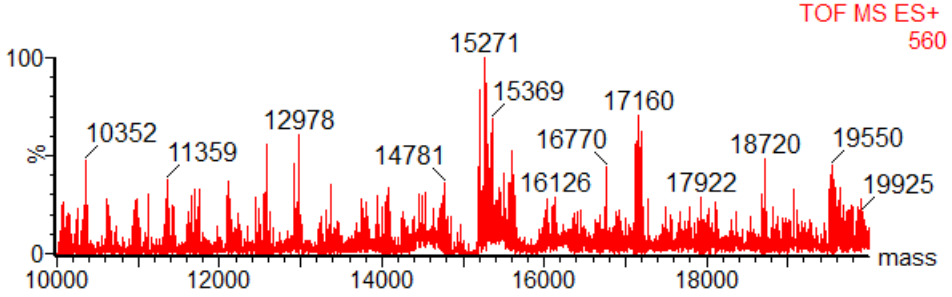


t=1h

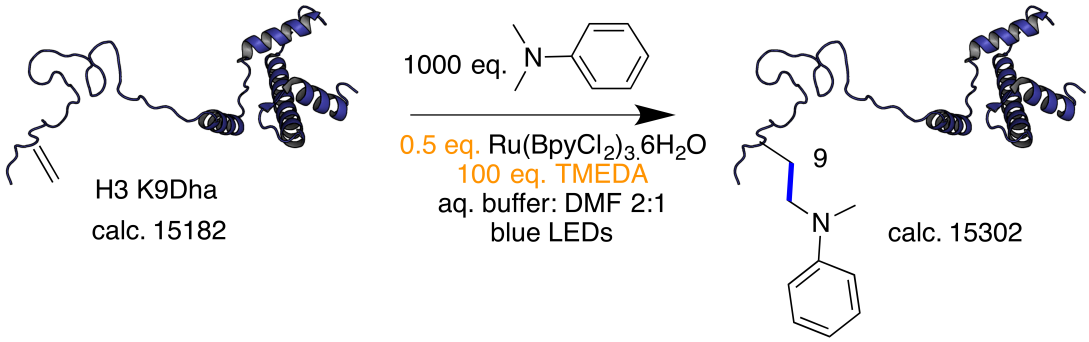


t=12h





Optimised conditions for the addition of dimethylaniline 2 to H3-K9Dha using catalytic $Ru(bpy)_3Cl_2.6H_2O$



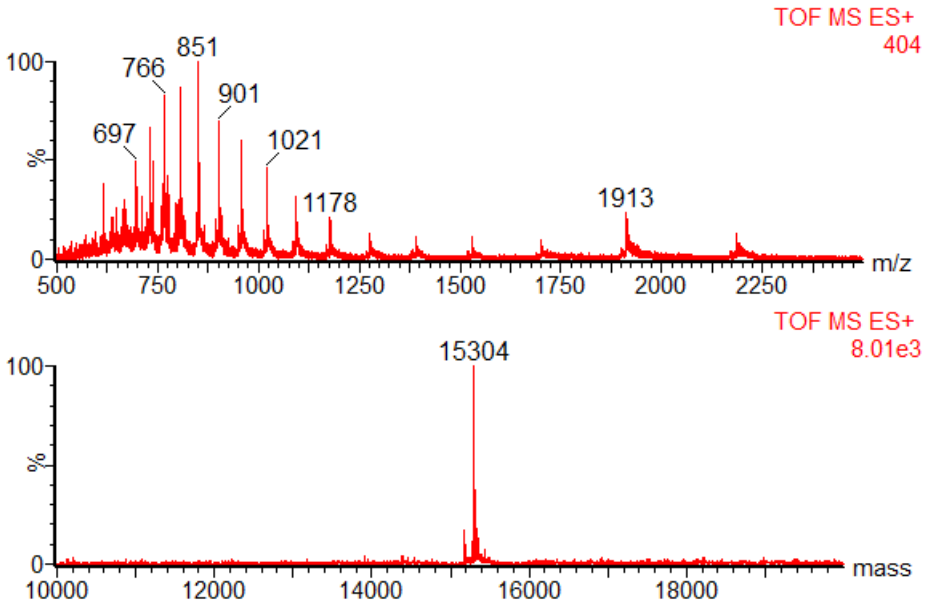
Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 170 µL) was added to an eppendorf tube, followed by 30 µL of 10 mg/mL Dha-protein stock solution (300 µg, ~20 nmol of protein, in the same buffer) to give a final protein concentration of 1.5 mg/mL (~100 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Dimethylaniline was ported into the glovebox and the appropriate volume (1.6µL, 1000 molar equiv. relative to protein) dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst $Ru(bpy)_3Cl_2.6H_2O$ (0.5 equiv. relative to protein, from a freshly prepared stock solution) was then added by pipette. Neat TMEDA (0.3 µL, 100 equiv. relative to protein), dissolved in 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 1 hour, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.

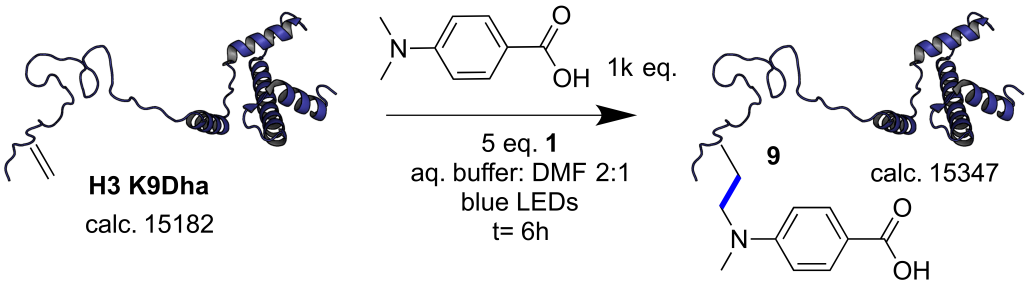
Calculated mass: 15302

Observed mass: 15304



Substrate scope

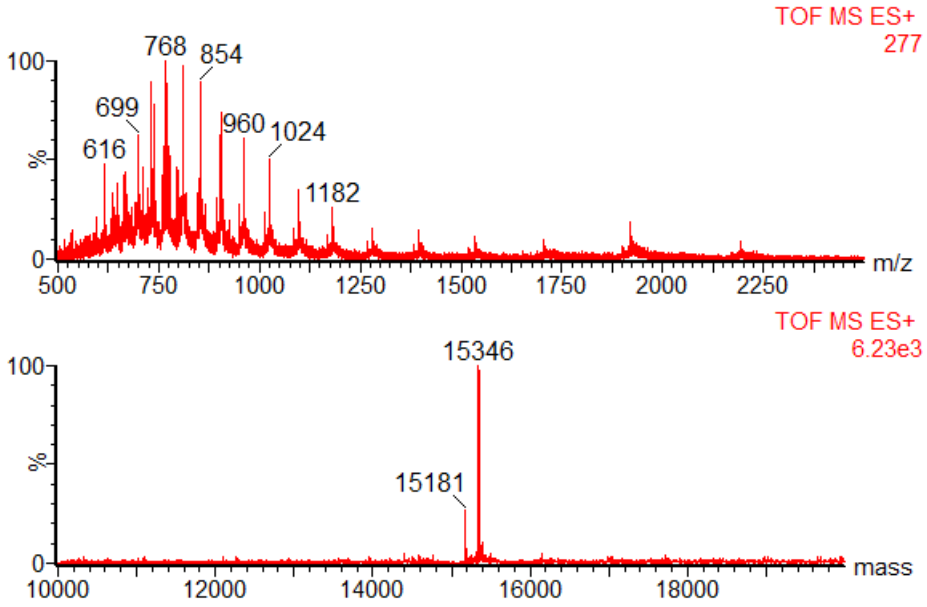
Reaction of 4-(dimethylamino)benzoic acid with H3-K9Dha



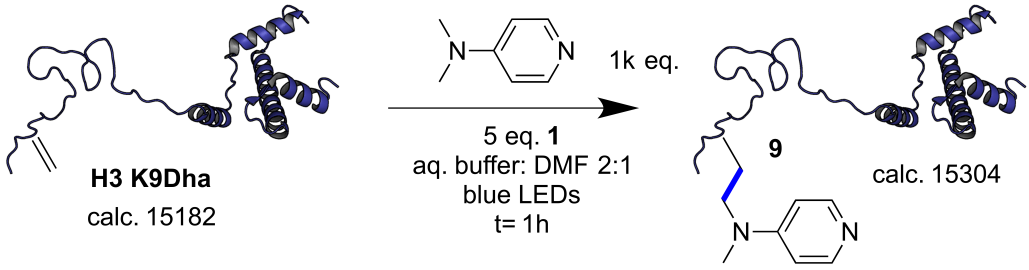
Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Dimethylaminobenzoic acid (2.2 mg, 1000 equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** ((2.2 µL of a 20 mM stock in DMF, 5 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 6 hours, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15347
Observed mass: 15346


TOF MS ES+
277
100
%
0
768
854
699
960
1024
616
1182
500
750
1000
1250
1500
1750
2000
2250
m/z
TOF MS ES+
6.23e3
100
%
0
15346
15181
10000
12000
14000
16000
18000
mass

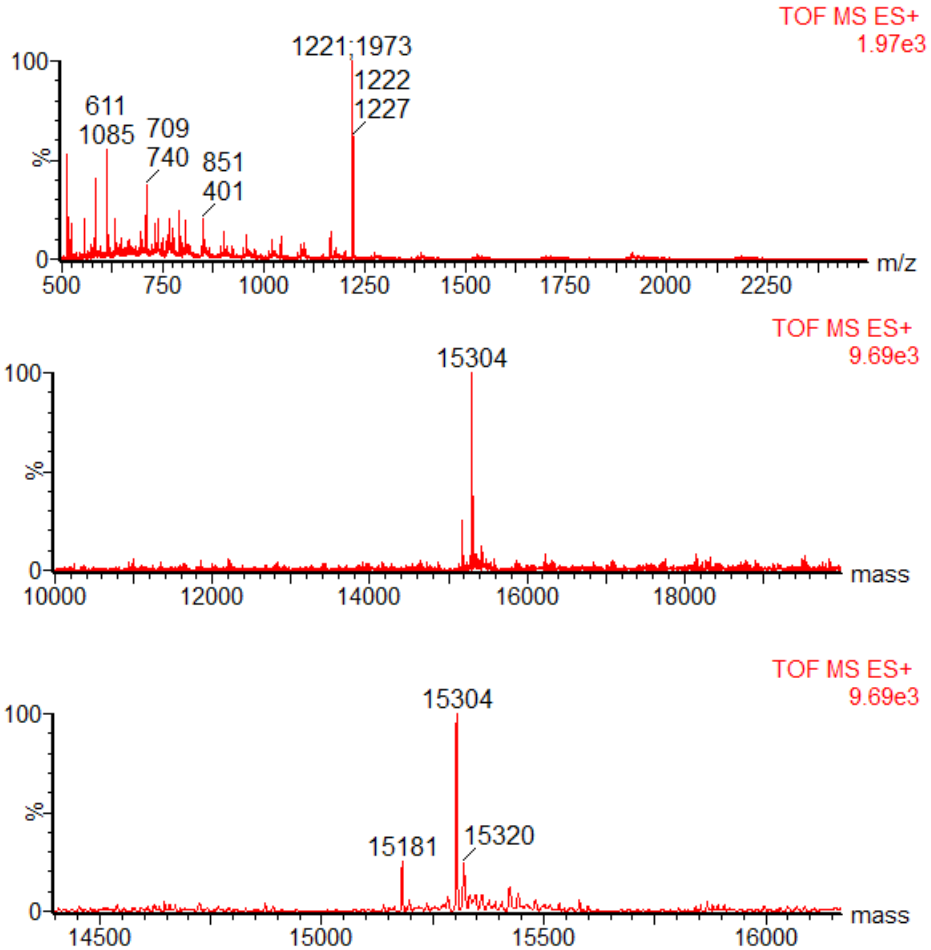
Reaction of 4-(dimethylamino)pyridine (DMAP) with H3-K9Dha



Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

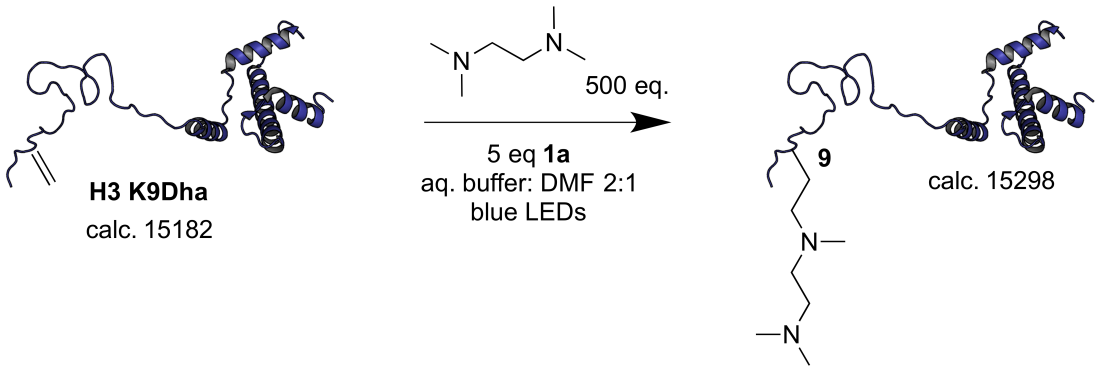
DMAP (1.6 mg, 1000 equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** ((2.2 µL of a 20 mM stock in DMF, 5 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 1 hour, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15304
Observed mass: 15304



Note: Spectra are of non-desalted reaction mixture. Dialysis may improve the quality of the ion series following reaction.

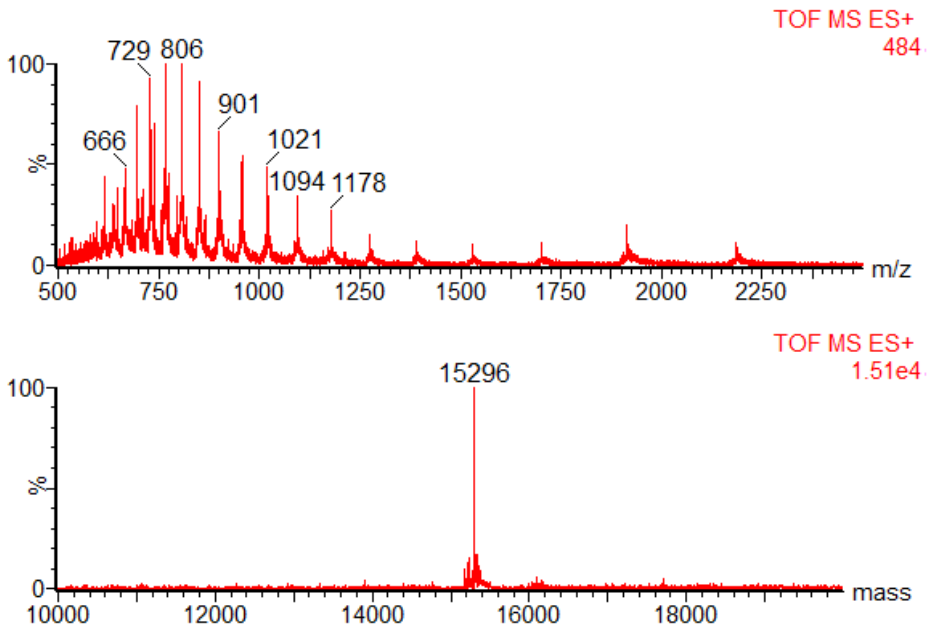
Reaction of tetramethylethylenediamine (TMEDA) with H3-K9Dha



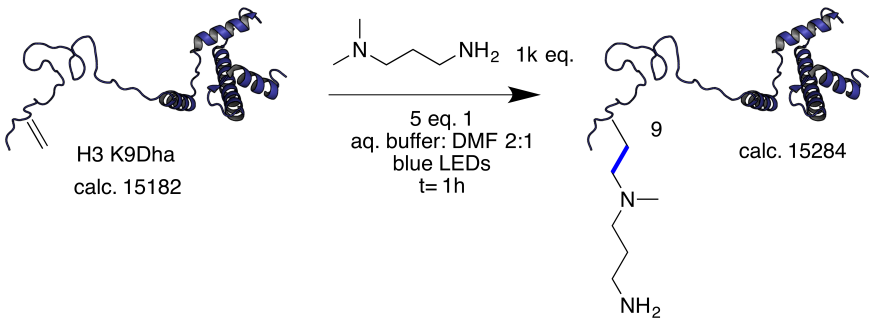
Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

TMEDA was ported into the glovebox and the appropriate volume (1.5 µL, 500 equiv. relative to protein) dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** ((2.2 µL of a 20 mM stock in DMF, 5 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 1 hour, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15298
Observed mass: 15296


TOF MS ES+
484
729
806
901
666
1021
1094
1178
100
%
0
500
750
1000
1250
1500
1750
2000
2250
m/z
TOF MS ES+
1.51e4
15296
100
%
0
10000
12000
14000
16000
18000
mass

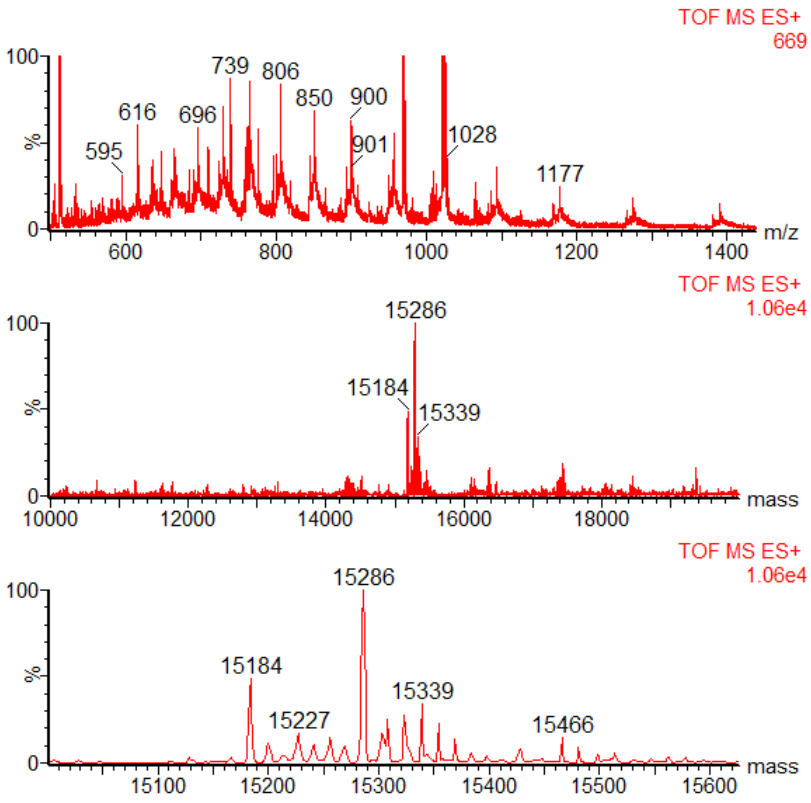
Reaction of 3-(dimethylamino)propylamine with H3-K9Dha



Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

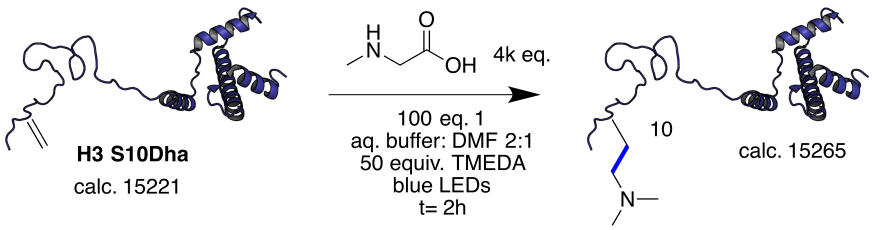
3-(dimethylamino)propylamine was ported into the glovebox and the appropriate volume (1.7 µL, 1000 equiv. relative to protein) dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** ((2.2 µL of a 20 mM stock in DMF, 5 equiv. relative to protein) was then added by pipette. 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 1 hour, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15284
Observed mass: 15184 (H3-K9Dha), 15286, 15339



Note: Spectra are of non-desalted reaction mixture. Dialysis may improve the quality of the spectra.

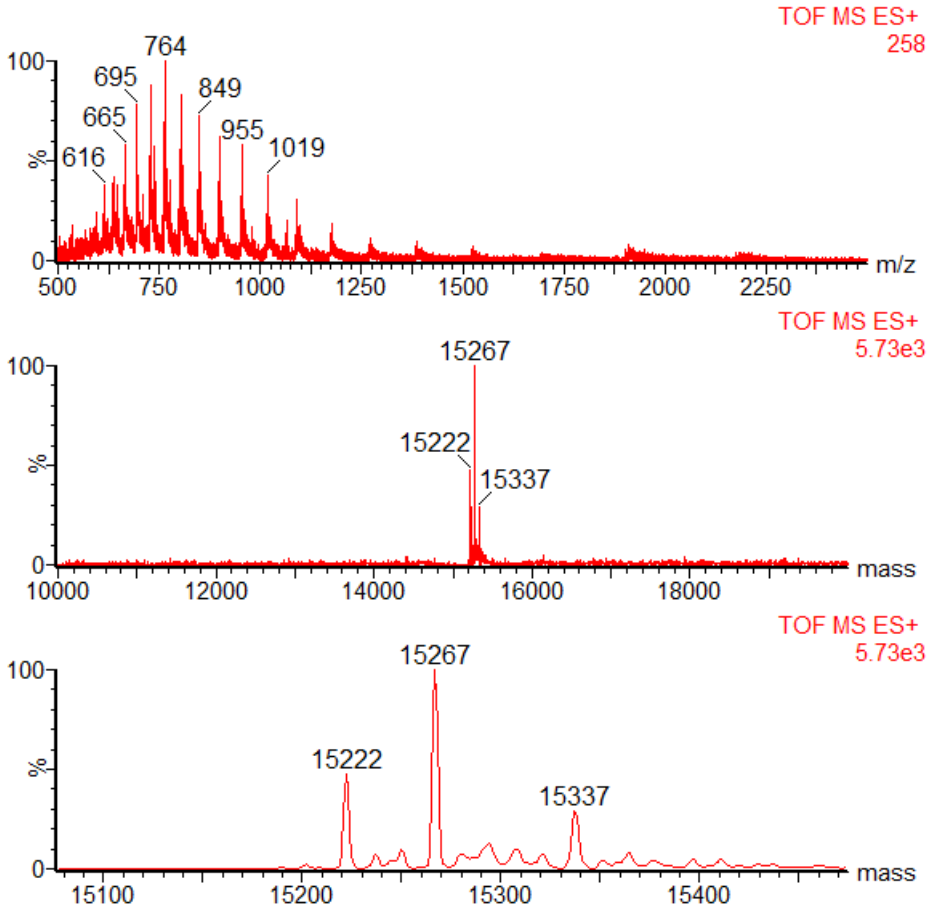
Reaction of sarcosine with H3-S10Dha



Following general procedure **7B**, reaction buffer (500 mM NH_4OAc, 5M Gdn.HCl, pH 6.0, 180 µL) was added to an eppendorf tube, followed by 20 µL of 10 mg/mL Dha-protein stock solution (200 µg, ~13 nmol of protein, in the same buffer) to give a final protein concentration of 1 mg/mL (~66 µM). The resultant protein solution was then added to an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions overnight.

Sarcosine (4.7mg, 4000 molar equiv. relative to protein) was ported into the glovebox and dissolved in the buffered protein solution, which was then pipetted into a glass vial. Photoredox catalyst **1** (4.3 µL of a 0.2 M stock in DMF, 100 equiv. relative to protein) was then added by pipette. TMEDA (0.2 µL, 50 equiv. relative to protein) dissolved in 100 µL degassed DMF was then added. The reaction was irradiated with blue LEDs and aliquots taken regularly for reaction monitoring by LC-MS. Upon reaction completion, after 1 hour, the protein solution was removed from the glovebox and quenched via addition to, and elution from, a G25 size-exclusion spin column, as per manufacturers instructions. The isolated protein was analysed by LC-MS and stored at -20°C.

ESI-MS for the modified histone is shown below.
Calculated mass: 15265
Observed mass: 15222 (H3-S10Dha), 15267, 15337 (H3-S10Dha+TMEDA)



References: Chapter 7

1 Kawamoto, T. & Ryu, I. Radical reactions of borohydrides. *Organic & biomolecular chemistry* **12**, 9733-9742 (2014).

2 Wickner, R. B. Dehydroalanine in histidine ammonia lyase. *Journal of Biological Chemistry* **244**, 6550-6552 (1969).

3 Hodgins, D. S. Yeast phenylalanine ammonia-lyase purification, properties, and the identification of catalytically essential dehydroalanine. *Journal of Biological Chemistry* **246**, 2977-2985 (1971).

4 Gavaret, J.-M., Nunez, J. & Cahnmann, H. Formation of dehydroalanine residues during thyroid hormone synthesis in thyroglobulin. *Journal of Biological Chemistry* **255**, 5281-5285 (1980).

5 Prier, C. K., Rankic, D. A. & MacMillan, D. W. Visible light photoredox catalysis with transition metal complexes: Applications in organic synthesis. *Chemical reviews* **113**, 5322 (2013).

6 Shaw, M. H., Twilton, J. & MacMillan, D. W. Photoredox catalysis in organic chemistry. *The Journal of organic chemistry* **81**, 6898 (2016).

7 Yan, M., Lo, J. C., Edwards, J. T. & Baran, P. S. Radicals: reactive intermediates with translational potential. *Journal of the American Chemical Society* **138**, 12692 (2016).

8 Stephenson, C. R., Studer, A. & Curran, D. P. (Beilstein-Institut, 2013).

9 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

10 Fancy, D. A. & Kodadek, T. Chemistry for the analysis of protein–protein interactions: rapid and efficient cross-linking triggered by long wavelength light. *Proceedings of the National Academy of Sciences* **96**, 6020-6024 (1999).

11 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).

12 Jay, D. G. Selective destruction of protein function by chromophore-assisted laser inactivation. *Proceedings of the National Academy of Sciences* **85**, 5454-5458 (1988).

13 Lee, J., Yu, P., Xiao, X. & Kodadek, T. A general system for evaluating the efficiency of chromophore-assisted light inactivation (CALI) of proteins reveals Ru (II) tris-bipyridyl as an unusually efficient "warhead". *Molecular BioSystems* **4**, 59-65 (2008).

14 McCarver, S. J. *et al.* Decarboxylative Peptide Macrocyclization through Photoredox Catalysis. *Angewandte Chemie International Edition* **56**, 728-732 (2017).

15 Crich, D., Davies, J., Negron, G. & Quintero, L. Preparation of α-amino acids by radical addition to a dehydroalanine derivative. *Journal of chemical research. Synopses*, 140-141 (1988).

16 Crich, D. & Davies, J. W. Free-radical addition to di-and tripeptides containing dehydroalanine residues. *Tetrahedron* **45**, 5641-5654 (1989).

17 Sorin, G. *et al.* Oxidation of Alkyl Trifluoroborates: An Opportunity for Tin‐Free Radical Chemistry. *Angewandte Chemie International Edition* **49**, 8721-8723 (2010).

18 Tellis, J. C., Primer, D. N. & Molander, G. A. Single-electron transmetalation in organoboron cross-coupling by photoredox/nickel dual catalysis. *Science* **345**, 433-436 (2014).

19 Miyake, Y., Nakajima, K. & Nishibayashi, Y. Visible-light-mediated utilization of α-aminoalkyl radicals: addition to electron-deficient alkenes using photoredox catalysts. *Journal of the American Chemical Society* **134**, 3338-3341 (2012).

20 Molander, G. A. Organotrifluoroborates: Another branch of the mighty oak. *The Journal of organic chemistry* **80**, 7837-7848 (2015).

21 Yasu, Y., Koike, T. & Akita, M. Visible Light‐Induced Selective Generation of Radicals from Organoborates by Photoredox Catalysis. *Advanced Synthesis & Catalysis* **354**, 3414-3420 (2012).

22 Miyazawa, K., Yasu, Y., Koike, T. & Akita, M. Visible-light-induced hydroalkoxymethylation of electron-deficient alkenes by photoredox catalysis. *Chemical Communications* **49**, 7249-7251 (2013).

23 Wallentin, C.-J., Nguyen, J. D., Finkbeiner, P. & Stephenson, C. R. Visible light-mediated atom transfer radical addition via oxidative and reductive quenching of photocatalysts. *Journal of the American Chemical Society* **134**, 8875-8884 (2012).

24 Narayanam, J. M. & Stephenson, C. R. Visible light photoredox catalysis: applications in organic synthesis. *Chemical Society Reviews* **40**, 102-113 (2011).

25 Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science* **2**, 1666-1676 (2011).

26 Hu, J., Wang, J., Nguyen, T. H. & Zheng, N. The chemistry of amine radical cations produced by visible light photoredox catalysis. *Beilstein journal of organic chemistry* **9**, 1977-2001 (2013).

27 Kohls, P., Jadhav, D., Pandey, G. & Reiser, O. Visible light photoredox catalysis: generation and addition of N-aryltetrahydroisoquinoline-derived α-amino radicals to Michael acceptors. *Organic letters* **14**, 672-675 (2012).

28 Lee, J., Udugamasooriya, D. G., Lim, H.-S. & Kodadek, T. Potent and selective photo-inactivation of proteins with peptoid-ruthenium conjugates. *Nature chemical biology* **6**, 258-260 (2010).

29 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 (2014).

30 Yu, C., Iqbal, N., Park, S. & Cho, E. J. Selective difluoroalkylation of alkenes by using visible light photoredox catalysis. *Chemical Communications* **50**, 12884-12887 (2014).

31 Koshi, Y. *et al.* Target-specific chemical acylation of lectins by ligand-tethered DMAP catalysts. *Journal of the American Chemical Society* **130**, 245-251 (2008).

32 El Khatib, M., Serafim, R. A. M. & Molander, G. A. α - Arylation/Heteroarylation of Chiral α - Aminomethyltrifluoroborates by Synergistic Iridium Photoredox/Nickel Cross - Coupling Catalysis. *Angewandte Chemie International Edition* **55**, 254-258 (2016).

33 Millet, A., Lefebvre, Q. & Rueping, M. Visible - Light Photoredox - Catalyzed Giese Reaction: Decarboxylative Addition of Amino Acid Derived α - Amino Radicals to Electron - Deficient Olefins. *Chemistry-A European Journal* **22**, 13464-13468 (2016).

Conclusion: Prospects for chemical mutation of proteins

Limitations of the method

This thesis will have made clear the challenges involved in deploying alkyl radicals for chemical modification of proteins at dehydroalanine residues. The chief chemical challenges I have described, competing di-alkylation at Dha and oxidative cleavage, are inherent to the reaction mechanism. The initial alkyl radical conjugate addition yields a Cα radical intermediate, which is sufficiently long-lived to allow the unwanted termination reactions to occur. The system developed in this thesis, employing borohydride salts as radical mediators, may begin to overcome the issues posed by the problematic Cα radical by engaging a suitable quenching cycle based on H-atom transfer (**Figure 4.1,** reproduced below). This hypothesis is in line with previous results from Kawamoto et al.[1], who proposed a polar effect wherein hydrogen transfer to the Cα-ester radical from borohydride reagents is preferred over direct transfer to alkyl-type radicals. Testing whether such a mechanism is operating in the protein case (by isotope labeling with $NaBD_4$ or D_2O) should guide future efforts to further optimize chemical mutation reactions.

As mentioned in **Chapter** 7, the 'dual' character of borohydride in aqueous solution (acting as single-electron initiator or mediator while also acting as a nucleophilic hydride source) can limit conversions in the case of bulky or electronically disfavoured alkyl halides, as reduction of Dha to alanine becomes a competing reaction. This observation motivated the development of alternative photoredox systems for generation and reaction of alkyl (and other classes of) radicals with Dha. The increased scope for temporal control of reaction offered by light-initiation, coupled with the range of available redox mediators and quenching agents for photoredox reaction manifolds, suggests that further investigation of these reactions offers great promise for future 'chemical mutations'.

Perhaps the most difficult challenge to overcome is the generation of an epimeric mixture at the site of modification as a direct consequence of radical conjugate addition. Notably, this will result in more complex mixtures of protein diastereomers for multiple mutation sites, which may hinder application of our 'chemical mutation' protocol to multiple sites within a single protein.

Our analysis of the stereoselectivity in the case of H3-S10Pfa (see **Chapter 6**) suggests roughly equal formation of both $_D$- and $_L$-configurations at the mutation site. It is important to note that this represents the ratio of epimers formed in the case of an unfolded, random coil state as excess denaturant was used in the reaction. It is likely that within the context of a protein fold, one epimer may be enriched by stereoselective hydrogen atom transfer or enolate quenching steps; testing this possibility rigorously will mandate generation of a panel of mutations across different protein environments and the development of a rapid assay for the epimeric ratio. An NMR-based approach would be ideal as it would not require any further processing of the protein post-reaction; the CF_2-Phosphate probe used in **Chapter 6** may be a useful starting point. Alternatively, Brik et al.[2] have recently reported a tryptic digest LC-MS/MS approach for measuring the ratio of epimeric products in the nucleophilic addition reaction of thiols to Dha-proteins. While this strategy requires additional enzymatic digestion and handling steps, it may be amenable to higher throughput than the NMR approach, enabling analysis of multiple proteins/sites rapidly.

Understanding the influence of protein structure may ultimately enable the design of conditions for stereoselective reactions on protein substrates, a grand challenge for future protein chemistry. This will require development of methods for either stereoselective delivery of a hydrogen atom (for direct quenching of the Cα radical) or for electron transfer and stereoselective protonation of the resultant protein Cα enolate (see Figure below). Guidance from the mechanisms of radical enzymes, such as the

recently described epimerases, and from recent developments in synthetic chemistry, may prove essential.

Alternatively, it may be possible to enrich the desired product by an affinity strategy. In **Chapter 5**, we demonstrated selective recognition of the natural $_{L}$-configured H3-K9$_{me3}$ histone by the antibody used for western blotting experiments. Attaching such a epimer-selective antibody to a resin and applying standard pull-down procedures may enable enrichment of the natural L-epimer of the modified protein. Unfortunately, this strategy will likely remain a bespoke endeavour, as antibodies are only commercially available for certain side-chains and sites of modification in a limited number of proteins.

In conjunction with other emerging approaches, we believe that protocols for 'chemical mutation', such as those developed here, will enable a new approach to 'protein chemical editing' that can supplement and extend genetic manipulations.

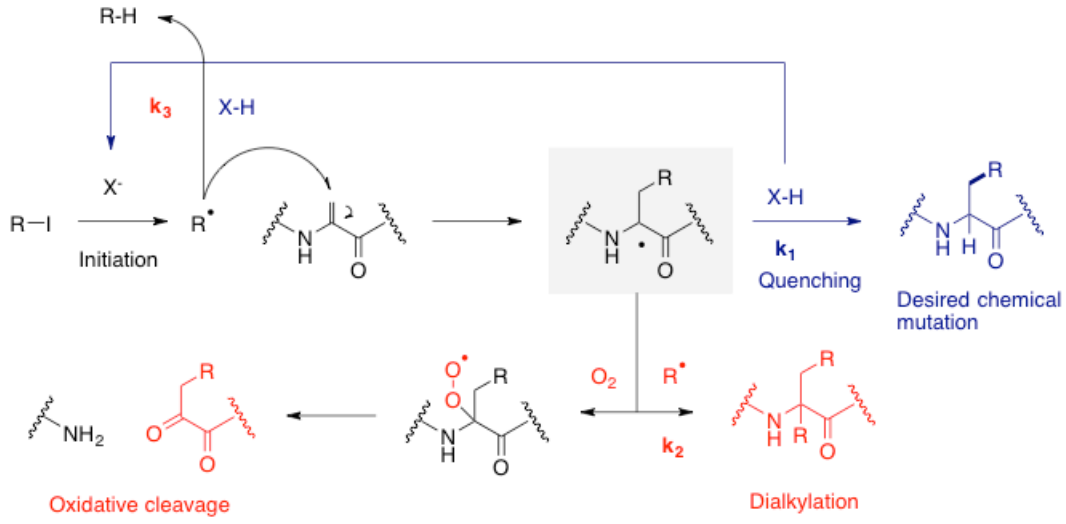


Figure (4.1): Proposed chain-cycle for quenching of the problematic Cα radical via hydrogen atom transfer

1 Kawamoto, T. *et al.* Borohydride-mediated radical addition reactions of organic iodides to electron-deficient alkenes. *The Journal of organic chemistry* **79**, 3999-4007 (2014).
2 Meledin, R., Mali, S. M., Singh, S. K. & Brik, A. Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1, 4-addition step. *Organic & biomolecular chemistry* **14**, 4817-4823 (2016).