UNIVERSIDADE DE LISBOA

Faculdade de Medicina

Development of a novel chemical biology strategy for the site-specific acetylation of histones

Cláudia Filipa Martins Afonso

Orientadores: Professor Doutor Gonçalo José Lopes Bernardes

Doutor Pedro Miguel Duarte Saraiva Cal

Tese especialmente elaborada para obtenção do grau de

Doutor em Ciências Biomédicas, especialidade em Biologia Celular e Molecular

2024

UNIVERSIDADE DE LISBOA

Faculdade de Medicina

Development of a novel chemical biology strategy for the site-specific acetylation of histones

Cláudia Filipa Martins Afonso

Orientadores: Professor Doutor Gonçalo José Lopes Bernardes

Doutor Pedro Miguel Duarte Saraiva Cal

Tese especialmente elaborada para obtenção do grau de

Doutor em Ciências Biomédicas, especialidade em Biologia Celular e Molecular

Júri:

Presidente: Doutor Mário Nuno Ramos de Almeida Ramirez, Professor Associado com Agregação e Vice-Presidente do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa

Vogais:

- PhD Christian Adam Olsen, Professor da University of Copenhagen (Denmark);
- Doutora Paula Alexandra de Carvalho Gomes, Professora Associada da Faculdade de Ciências da Universidade do Porto;
- Doutor Claus Maria Azzalin, Professor Associado Convidado da Faculdade de Medicina da Universidade de Lisboa;
- Doutor Gonçalo José Lopes Bernardes, Professor Auxiliar Convidado da Faculdade de Medicina da Universidade de Lisboa (Orientador).

Financiado pela Fundação para a Ciência e Tecnologia

(PD/BD/135512/2018 e Covid/BD/152516/2022)

2024

A impressão desta tese foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 19 de Março de 2024.

As opiniões expressas nesta publicação são da exclusiva responsabilidade do seu autor.

The opinions expressed in this publication are of the exclusive responsibility of its author.

Acknowledgements

I would like to thank my supervisors, Prof. Dr. Gonçalo Bernardes and Dr. Pedro Cal, for giving me the opportunity to conduct my doctoral thesis under their guidance and mentorship. I am grateful for their support which was fundamental for the completion of my thesis.

I would also like to extend my gratitude to everyone who contributed to my work, as well as to former and present members of the group for their help and support.

I would also like to express my appreciation to my thesis committee members, Prof. Dr. Pedro Góis, Prof. Dr. Sérgio Almeida, and Prof. Dra. Leonor Saúde for the productive discussions during our meetings.

I would like to acknowledge Fundação para a Ciência e Tecnologia (PD/BD/135512/2018 and Covid/BD/152516/2022) for funding my doctoral work in the context of the Lisbon Biomedical and Clinical Research PhD Program (LisbonBioMed).

I would also like to thank my LisbonBioMed PhD cohort for their friendship throughout this journey, in particular to Inês Faleiro for the many conversations over coffee we shared together during these years.

Finally, I would like to express my gratitude to my family, particularly to my parents and siblings for their unwavering support throughout these years.

Resumo

As proteínas são macromoléculas quimicamente diversas que estão envolvidas em quase todos os processos biológicos onde desempenham funções essenciais enquanto reguladoras da divisão celular, da diferenciação, do crescimento e do metabolismo, entre várias outras. A maioria destas funções depende da incorporação espacial e temporal de modificações pós-traducionais durante ou após a síntese proteica nos ribossomas. Ao alterarem propriedades físico-químicas, tais como a carga e a solubilidade, estas modificações podem influenciar de forma significativa a estrutura e dinâmica das próprias proteínas. Com mais de 400 atualmente descritas, as modificações de proteínas são características ubíquas e indispensáveis de qualquer sistema biológico, pois permitem expandir a diversidade estrutural e funcional destas macromoléculas. Um exemplo claro da importância biológica que as modificações pós-traducionais exercem em organismos vivos é dado pelas histonas.

As histonas representam uma família de proteínas de baixo peso molecular, ricas em aminoácidos básicos e altamente conservadas durante a evolução. Estas proteínas são responsáveis por organizar o ADN negativamente carregado em unidades fundamentais e repetitivas da cromatina, que se designam por nucleossomas. Compostos por um octâmero de histonas e aproximadamente 146 pares de bases de ADN, os nucleossomas representam o primeiro nível de organização da cromatina. As modificações pós-traducionais de histonas constituem um mecanismo epigenético essencial que modula a estrutura e função da cromatina, quer diretamente por influenciarem as interações eletrostáticas que se estabelecem com o ADN no nucleossoma, quer indiretamente por atuarem como marcadores para a associação de outras proteínas.

Descoberta pela primeira vez há mais de 50 anos, a acetilação é uma das modificações pós-traducionais mais comuns. Em histonas, esta modificação ocorre maioritariamente na amina da cadeia lateral das lisinas, onde a transferência de um grupo acetilo neutraliza a carga positiva destes resíduos. Como tal, a acetilação de histonas resulta numa diminuição da interação destas proteínas com o ADN, facilitando assim a acessibilidade deste último à maquinaria de replicação, transcrição e reparação. A acetilação de lisinas é um processo dinâmico e reversível, sendo regulado por dois grupos de enzimas que se designam por acetiltransferases e desacetilases. Enquanto as acetiltransferases catalisam a transferência de grupos acetilo para os resíduos de lisina, as desacetilases atuam no sentido oposto. A acetilação da lisina 9 na histona H3 é uma modificação pós-traducional que frequentemente ocorre nas regiões promotoras de genes ativamente transcritos e que pode levar ao recrutamento do fator de transcrição TFIID. Por sua vez, a acetilação da lisina 56 na histona H3 desempenha um papel importante para a estabilidade genómica, pois promove uma montagem eficiente da cromatina após a replicação e reparação do ADN.

Apesar da acetilação estar amplamente caracterizada em diversas posições da sequência de aminoácidos das histonas, maioritariamente por técnicas de espectrometria de massa, é fundamental existir uma interpretação mecanística de como esta modificação se combina com outras em locais específicos para regular a estrutura e função da cromatina. Assim, a capacidade

de incorporar a acetilação localmente, de uma forma específica, em histonas é essencial para avaliar o papel funcional desta modificação pós-traducional nestas proteínas. No entanto, é difícil obter amostras homogéneas com uma única acetilação, nas quantidades necessárias para estes estudos, a partir de fontes biológicas ou métodos enzimáticos.

Neste contexto, várias estratégias têm sido reportadas para acetilar histonas de forma específica, tais como a mutagénese química, a expansão do código genético, a ligação de proteínas, e o uso de ligandos de afinidade acoplados a catalisadores de transferência de grupos acetilo. Contudo, estas apresentam diversas desvantagens, nomeadamente a necessidade de mutação dos aminoácidos de interesse para gerar mímicos de acetil-lisina, o que pode resultar em interações sub-ótimas destas histonas modificadas com os seus parceiros naturais de ligação no caso da mutagénese química, o menor rendimento de histonas modificadas obtido através da expansão do código genético, a natureza desafiante dos métodos de ligação de proteínas, e a dependência da existência de alvos de interação com ligandos de modo a promover a acetilação de lisinas proximais na sequência proteica das histonas, no caso do uso de ligandos de afinidade acoplados a catalisadores de transferência de grupos acetilo.

Esta dissertação descreve o desenvolvimento de uma nova estratégia para a acetilação específica de histonas, que visa superar muitas das desvantagens dos métodos atuais. Para tal, utiliza de forma sequencial a conjugação via cisteína com química do “clique” para controlar de forma espacial e temporal a introdução de um único grupo acetilo na lisina mais próxima. Concretamente, nesta estratégia introduz-se uma única cisteína a uma distância de cinco resíduos da lisina de interesse. De seguida, uma maleimida que contém um fragmento de dibenzociclo-octino reage com o grupo tiol da cisteína através de uma reação de Michael. Este reagente bifuncional serve para ligar a histona a um modificador de lisinas que, por sua vez, contém uma azida e um doador estável de grupos acetilo. Na presença do reagente modificador de lisinas, ocorre uma reação de cicloadição azida-alcino promovida por tensão de anel que fixa o doador de grupos acetilo na vizinhança da cisteína modificada, onde se encontra também a lisina de interesse. Posteriormente, a acetilação é desencadeada pelo ataque nucleofílico da amina da cadeia lateral da lisina num dos carbonilos do doador de grupos acetilo.

Esta estratégia foi inicialmente aplicada para acetilar a lisina 9 da histona H3, tendo-se produzido para este efeito a correspondente proteína recombinante com uma cisteína na posição 4 (H3K4C). A escolha pela acetilação na posição 9 da H3 deveu-se a três fatores principais, em particular a importância biológica desta marca epigenética, a inexistência na sequência primária de outros resíduos de lisina a uma distância da cisteína igual ou semelhante à da lisina 9, e a acessibilidade deste resíduo que lhe é conferida por estar presente na cauda N-terminal da histona. Tendo sido aplicada com sucesso para gerar a respetiva histona mono-acetilada na lisina 9, esta estratégia foi subsequentemente testada para acetilar uma posição mais interna na sequência da H3. Assim, a lisina 56 foi escolhida como o segundo alvo da estratégia de acetilação, tendo-se para tal gerado a correspondente proteína recombinante com uma cisteína na posição 52 (H3R52C). Após várias tentativas de otimização, foi possível obter a respetiva proteína mono-acetilada na lisina 56. De seguida, avaliou-se a possibilidade de aplicar esta

estratégia para criar uma H3 duplamente acetilada nas posições 9 e 56, tendo o resultado sido igualmente favorável.

Com a estratégia de acetilação específica otimizada, as histonas H3K4C e H3R52C foram mono-acetiladas na posição 9 ou 56, respetivamente, enquanto que a proteína recombinante H3K4CR52C foi duplamente acetilada nestas duas lisinas, sem necessidade de recorrer a técnicas de biologia ou purificação complexas. Deste modo, a transferência inespecífica de grupos acetilo para outras lisinas nucleofílicas mais distantes não foi observada. É importante salientar que estas proteínas quimicamente acetiladas interagiram com os seus parceiros biológicos relevantes, mesmo na presença do produto secundário que permanece ligado à cisteína após a reação "clique". Os anticorpos comerciais gerados contra a lisina acetilada nas posições 9 e 56 da histona H3 reconheceram as proteínas mono- e di-acetiladas, conforme avaliado por experiências de Western blot e ELISA. Além disso, a sirtuína-3, que possui atividade de desacetilase, foi capaz de remover o grupo acetilo das histonas mono-acetiladas nas posições 9 e 56, demonstrando assim que estas proteínas podem ser utilizadas no estudo de enzimas modificadoras de histonas.

Como tal, estes resultados permitem afirmar que a estratégia desenvolvida no âmbito desta dissertação facilita o acesso a amostras de histonas acetiladas de forma específica. Estas proteínas podem ser futuramente usadas para estudar o papel de eventos de acetilação simples e duplos em determinados resíduos de lisina, o que poderá ser útil para decifrar a contribuição da acetilação para a função biológica das histonas.

PALAVRAS-CHAVE: acetilação específica, histonas, química do "clique", conjugação de cisteínas

Abstract

Proteins are diverse macromolecules involved in almost every biological process, serving crucial functions as regulators of cell division, differentiation, growth and metabolism, among others. Most of these functions depend on the spatiotemporal incorporation of post-translational modifications during or following protein synthesis and assembly in ribosomes. By altering physiochemical properties, such as charge and solubility, these modifications can dramatically influence protein structure and dynamics. With more than 400 reported so far, protein modifications are ubiquitous features of natural systems that expand the structural and functional diversity of these macromolecules, adding an extra layer of complexity to the proteome beyond the information encoded in the genome. A clear example of the biological significance that post-translational modifications exert in living organisms is afforded by histones.

Histones represent a family of small and highly conserved basic proteins that organize the negatively charged DNA into the repeating cored structures, termed nucleosomes, that form chromatin. Consisting of approximately 146 base pairs of DNA tightly wrapped around a histone octamer, nucleosomes represent the first level of chromatin compaction. Post-translational modifications in histones constitute an essential epigenetic mechanism that modulates chromatin structure and function, both directly by influencing the strength of the electrostatic interactions within the nucleosome and indirectly by acting as markers for the binding of chromatin-associated proteins.

First discovered more than 50 years ago, acetylation is one of the most common post-translational modifications. In histones, this modification mostly occurs at the side chain amine of lysine residues, where the transfer of an acetyl group neutralizes their positive charge. As a result, acetylation decreases the interaction of histones with the negatively charged DNA, thereby facilitating its accessibility to replication, transcription and repair machineries. Lysine acetylation is a dynamic and reversible process which is under the regulation of two groups of enzymes known as acetyltransferases and deacetylases. Whereas acetyltransferases catalyze the transfer of acetyl groups to lysine residues, deacetylases act in the opposite direction. Acetylation of lysine 9 in histone H3 is a frequent post-translational modification found in the promoter regions of actively transcribed genes and can recruit transcription factor TFIID, while acetylation of lysine 56 in this protein is critical for genomic stability by promoting efficient chromatin assembly following DNA replication and repair.

Although many acetylation marks have been characterized in histones, particularly by mass spectrometry-based approaches, a mechanistic understanding of how this modification combines with others at specific sites to regulate chromatin structure and function is essential. Thus, the ability to incorporate this modification in a site-specific manner is critical to evaluate its functional role. However, access to homogeneous samples of histones bearing a single acetyl group in the quantities required for these studies is difficult to achieve from biological sources or enzymatic methods.

In this context, several strategies have been reported to achieve site-specific acetylation of histones, namely chemical or post-expression mutagenesis, genetic code expansion, protein

ligation and the use of affinity ligands tethered to acetyl transfer catalysts. Nevertheless, these present many disadvantages, namely the mutation of the amino acids of interest to generate acetyl-lysine mimics, which may lead to sub-optimal interactions with their natural binding partners in the case of chemical or post-expression mutagenesis, the lower yield of modified histones with genetic code expansion, the more challenging nature of protein ligation, and the dependency on the existence of corresponding interacting ligands that can promote acetylation of proximal lysine residues in the protein sequence, when using affinity ligands tethered to acetyl transfer catalysts.

This thesis describes the development of a novel strategy for acetylating histones with site-specificity that seeks to overcome many of the disadvantages presented by the current methods. To achieve this, it sequentially uses cysteine conjugation with click chemistry to control the single addition of an acetyl group to the proximal lysine residue in a spatial and temporal manner. Specifically, in this strategy a unique cysteine residue is first engineered at a position that is five residues away from the intended lysine to be acetylated. Then, a maleimide reagent containing a dibenzocyclooctyne moiety is installed at the cysteine via a Michael addition reaction. This heterobifunctional crosslinker, with its sulfhydryl reactive moiety and clickable handle, serves to connect the protein to a lysine modifying reagent containing a click counterpart and a stable acetyl donor with a fixed spacer. In the presence of the lysine modifying reagent, a strain promoted azide-alkyne cycloaddition reaction occurs which locks the acetyl donor in proximity to the lysine of interest. Acetylation is subsequently triggered by the nucleophilic attack of the side chain amine of the lysine to one of the carbonyl groups of the acetyl donor.

This strategy was first tested to acetylate lysine 9 of histone H3, with the corresponding recombinant protein containing a cysteine at position 4 (H3K4C) produced for this purpose. The choice of acetylating position 9 of H3 was based on three main factors, namely the biological significance of this epigenetic mark, the absence of other lysine residues in the primary sequence at equal or similar distance from the cysteine, and the accessibility of K9 conferred by being in the N-terminal tail of the histone. Having successfully applied this strategy to generate the respective mono-acetylated protein at lysine 9, its broader applicability to acetylate a more internal position in the H3 sequence was evaluated. To this end, lysine 56 was chosen as the second target of the directed acetylation strategy and the corresponding recombinant protein containing a cysteine at position 52 (H3R52C) was produced. After several rounds of optimization, the respective mono-acetylated protein at lysine 56 was obtained. The possibility of applying this strategy to create a dual acetylated H3 at positions 9 and 56 was subsequently evaluated, for which the outcome was favorable as well.

With this cysteine-assisted click-chemistry strategy completely optimized, the recombinant histones H3K4C and H3R52C were mono-acetylated at position 9 or 56, respectively, while H3K4CR52C was di-acetylated at these two lysine residues without resorting to any complex biological or purification techniques. The unspecific transfer of acetyl groups to other, more distant nucleophilic lysine residues was not observed. Importantly, even in the presence of the click reaction by-product that remains attached at the cysteine, the chemically acetylated proteins were able to interact with their relevant biological partners. Commercial antibodies generated against

the natural acetyl-lysine at positions 9 and 56 of histone H3 recognized the mono- and di-acetylated proteins, as evaluated by Western Blot and ELISA experiments. Furthermore, the sirtuin-3 deacetylase was able to remove the acetyl group from the mono-acetylated histones at positions 9 and 56, thereby demonstrating that these proteins can be used in the study of histone-modifying enzymes.

Therefore, the strategy outlined in this dissertation facilitates access to samples with site-specifically acetylated histones. These proteins can be used in future studies to probe the role of single and double acetylation events at precise lysine residues, which may help to decipher the contribution of acetylation towards the biological function of histones.

KEYWORDS: site-specific acetylation, histones, click chemistry, cysteine conjugation

Contents

List of Figures

List of Schemes

List of Tables

Abbreviations

5MP	5-Methylene pyrrolone
6xHis	Six-residue polyhistidine
Acetyl-CoA	Acetyl-coenzyme A
ADMA	Asymmetric di-methyl arginine
ADP	Adenosine diphosphate
ASF1	Anti-silencing function 1
ATP	Adenosine triphosphate
BET	Bromodomain and extra-terminal domain
BME	β-mercaptoethanol
Boc	*tert*-Butyloxycarbonyl
CD	Circular dichroism
CHD	Chromodomain Helicase DNA-binding
CID	Collision-induced dissociation
CoA	Coenzyme A
DBCO	Dibenzocyclooctyne
DBHDA	2,5-Dibromohexanediamide
Dha	Dehydroalanine
DMAP	4-dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSH	DMAP with a mercaptomethyl group in the second position
DTT	Dithiothreitol
ECD	Electron-capture dissociation
EDTA	Ethylenediaminetetraacetic acid
EK	Enterokinase
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
EPL	Expressed protein ligation
ESI	Electrospray ionization
ESI-MS	Electrospray ionization mass spectrometry
E. coli	*Escherichia coli*
FAD	Flavin adenine dinucleotide
FeBABE	Iron-(S)-1-[p-(bromoacetamido)benzyl] ethylenediaminetetraacetate
Fmoc	9-Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
FT-ICR	Fourier-transform ion cyclotron resonance
H3K9Ac	Histone H3 acetylated at lysine 9

H3K56Ac	Histone H3 acetylated at lysine 56
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HF	Hydrofluoric acid
HPLC	High-performance liquid chromatography
iMet	Initiator methionine
INO80	Inositol requiring 80
IPTG	Isopropyl β-D-thiogalactopyranoside
ISWI	Imitation switch
JmjC	Jumonji C
KAT	Lysine acetyltransferase
KDAC	Lysine deacetylase
LANA	Latency-associated nuclear antigen
LB	Luria broth
LC-MS	Liquid chromatography coupled to mass spectrometry
MBP	Maltose-binding protein
MMA	Mono-methyl arginine
mRNA	Messenger RNA
MS	Mass spectrometry
MTCA	Methylthiocarbonyl-aziridine
MTCTK	Methylthiocarbonyl-thialysine
m/z	Mass-to-charge ratio
NAD	Nicotinamide adenine dinucleotide
NAM	Nicotinamide
NAT	N-terminal acetyltransferase
NBS	N-bromosuccinimide
ncAA	Non-canonical amino acid
NCL	Native chemical ligation
NME	N-terminal methionine excision
NMR	Nuclear magnetic resonance
NRMT1	N-terminal RCC1 methyltransferase 1
NRMT2	N-terminal RCC1 methyltransferase 2
Nt-acetylation	N^{α}-terminal acetylation
Nt-methylation	N^{α}-terminal methylation
NTMT	N-terminal methyltransferase
Nt-ubiquitination	N^{α}-terminal ubiquitination
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBS-0.05	PBS with 0.05% Tween-20
PKDM	Protein lysine demethylase

PKMT	Protein lysine methyltransferase
PMT	Protein methyltransferase
ppm	Parts per million
PRMT	Protein arginine methyltransferase
PTM	Post-translational modification
RCC1	Regulator of chromatin condensation 1
RNA	Ribonucleic acid
RSC	Remodeling the structure of chromatin
RSF	Remodeling and spacing factor
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SAU	Sodium-acetate-urea
SDMA	Symmetric di-methyl arginine
SDS	Sodium dodecyl-sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
Sirt3	Sirtuin-3
Sirt6	Sirtuin-6
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPPS	Solid-phase peptide synthesis
SWI/SNF	Switch/sucrose non-fermentable
TBS	Tris-buffered saline
TBST-0.05	TBS with 0.05% Tween-20
TBST-0.1	TBS with 0.1% Tween-20
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TFIID	Transcription factor IID
Thz	Thiazolidine
tRNA	Transfer ribonucleic acid
UPLC	Ultra-performance liquid chromatography
UPS	Ubiquitin proteasome system
UV	Ultraviolet
X. laevis	*Xenopus laevis*

1. Introduction

Together with nucleic acids, polysaccharides and lipids, proteins constitute one of the main classes of macromolecules that orchestrate life at the cellular level. Of these, proteins are the most abundant and versatile, occurring in every cell and with enormous variety in size, shape, composition and function.[1] Bridging the complex space between the genetic information encoded within DNA and the various phenotypes in health and disease, proteins are involved in almost every biological process in the crowded molecular environment of cells, serving crucial roles as regulators of cellular division, differentiation, growth and metabolism, among others.[2]

Their functional diversity is particularly remarkable, considering that proteins as biologically distinct as enzymes, hormones, antibodies, signal receptors or transporters are composed of the same genetically encoded 20 amino acids.[1] Although the side chains of these canonical amino acids are sufficiently diverse from a chemical perspective to enable some of the processes associated with proteins, such as proper folding into their native three-dimensional structures, protein function relies tremendously on the *in vivo* spatiotemporal incorporation of modifications during or following their synthesis and assembly in ribosomes.[3] Some of these modifications, most notably N-terminal methionine excision and N-terminal acetylation, occur during protein synthesis.[4] However, for simplicity, both co- and post-translational events will be designated jointly as post-translational modifications (PTMs).[3]

1.1 Post-translational modifications as a source of protein diversity

Almost every eukaryotic protein is chemically modified at the post-translational level. With more than 400 different PTMs reported so far, protein modification is a ubiquitous feature of natural systems that expands the structural and functional diversity of these molecules, adding an extra layer of complexity to the proteome beyond the information encoded in the genome.[5] By altering their physicochemical properties, such as charge and solubility, PTMs can influence the overall structure and dynamics of proteins, thereby affecting their stability, localization and function.[6]

The scope and diversity of PTMs is so substantial that these modifications can be classified in many ways. For instance, PTMs can be divided into subtypes according to the location in the protein where the modification occurs. In this context, PTMs can occur at the N- or C-termini of proteins, the polypeptide backbone or on the residue side chains. From a structural perspective, PTMs can be classified based on whether proteins are cleaved or modified at specific sites. The latter can consist of the covalent addition of simple chemical groups, such as phosphate, acetyl and methyl, or complex structures, such as polypeptides, glycans and lipids.[2] Furthermore, some PTMs can result from the direct modification of amino acids, as in the case of disulfide bond formation and deamidation.[7] A schematic illustration of PTM classification according to the type of modification introduced, together with a subset of representative examples, is shown in **Figure 1**. PTMs can also be grouped as enzymatic or non-enzymatic. Many modifications are mediated by enzymes through highly regulated pathways, whereas some are promoted by spontaneous

chemical reactions that reflect the integration of various environmental stimuli in a particular genomic, transcriptomic and metabolomic setting.[8] Lastly, PTMs can be reversible or irreversible, depending on whether the modifications proceed in one or both directions.

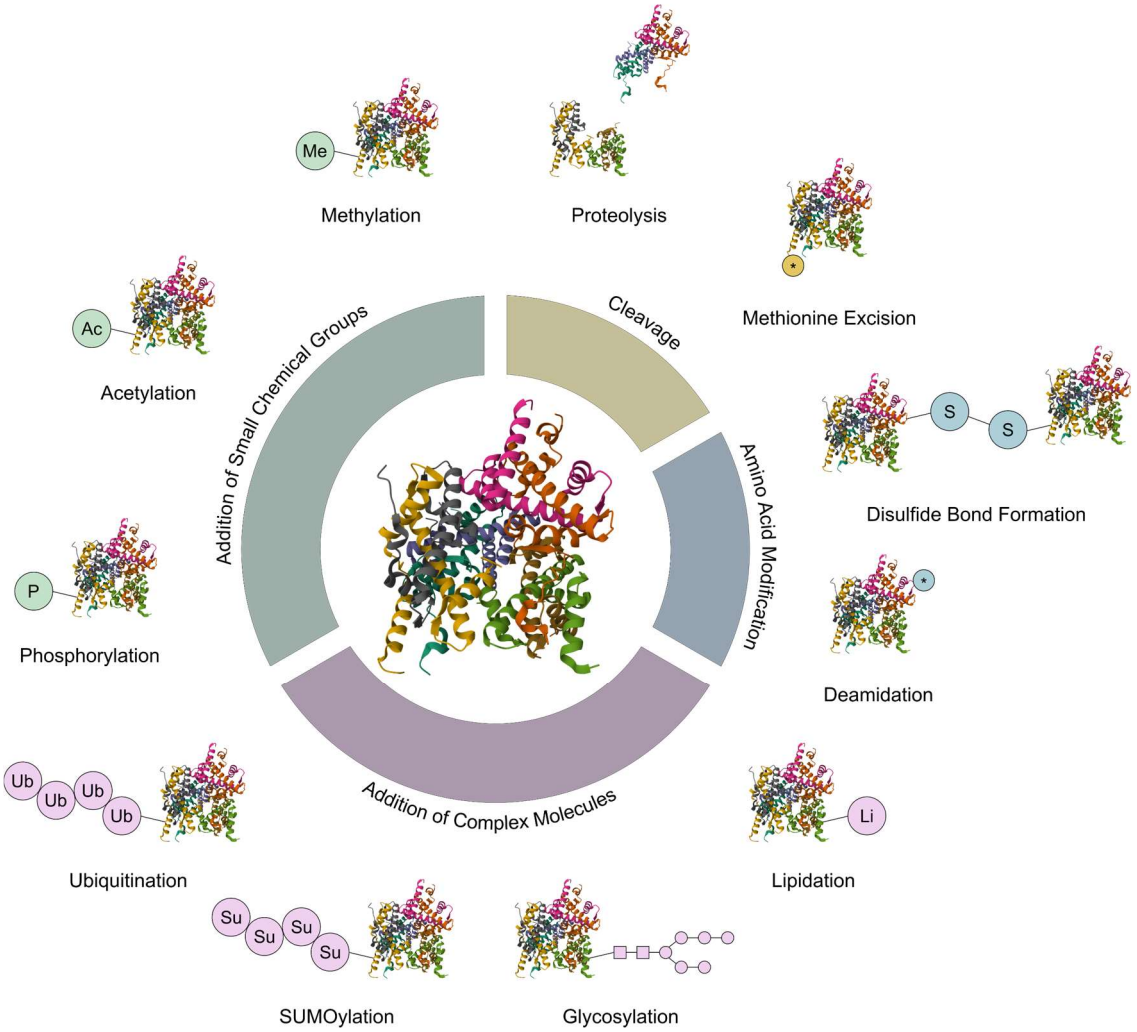


Figure 1. Schematic illustration of PTM classification according to the type of modification introduced, together with a subset of representative examples. Proteins can be cleaved or modified at specific sites. The latter can consist of the addition of simple chemical groups or complex structures. In addition, some PTMs result from the direct modification of amino acids. This figure was adapted from [7,9,10]. The represented structure corresponds to the native histone octamer derived from chicken erythrocytes and was retrieved from the Protein Data Bank (PDB) archive (PDB ID: 2ARO).[11,12,13]

Terminal PTMs target more than 90% of the mammalian proteome and are important for a variety of biological processes, such as protein maturation, sorting, membrane integration and cellular signaling.[14] The protein termini are particularly amenable to modification due to the presence of reactive amino and carboxyl groups of the polypeptide chain, which can be further primed by the enzymatic machinery of cells to expose highly nucleophilic residues.[14] For instance, removal of the signal peptide or the pro-domain by proteases generates two new polypeptides, one with a neo C-terminus and another with a neo N-terminus, both of which can be subsequently targeted by various modifications.[15] Due to the lower chemical reactivity of the C-terminus compared to the N-terminus, the identification and study of C-terminal modifications is more challenging and these PTMs are thus assumed to be less prevalent.[16]

Most PTMs targeting the amino and carboxyl groups of the N- and C-termini of proteins, respectively, can also occur at the side chains of residues. In addition, many others are found exclusively on these side chains due to the presence of a variety of nucleophilic groups that can easily target surrounding electrophiles.[17] Modifications occurring at the side chains of proteins are mostly enzymatic and reversible.[18] According to dbPTM, a comprehensive database that provides structural and functional analysis of PTM sites, the most common modifications are phosphorylation, ubiquitination and acetylation, which together comprise more than 90% of all modified substrate sites. Furthermore, serine is the most frequently modified residue, mainly by phosphorylation, while lysine is the amino acid that can be modified by the widest range of PTM types.[19] This section will provide a brief description of the three most prevalent PTMs mentioned above, as well as methylation and methionine excision.

1.1.1 Phosphorylation

One of the most significant, well-characterized and widely observed PTMs in proteins is phosphorylation, with more than two-thirds of proteins encoded by the human genome subjected to this modification.[20] This dynamic and reversible PTM is governed by the opposing activities of enzymes known as protein kinases and phosphatases. More than 500 protein kinases and approximately 180 protein phosphatases are encoded in the human genome.[21] Protein kinases catalyze the transfer of a phosphoryl group from the γ position of adenosine triphosphate (ATP) to specific residues in proteins substrates. This catalytic process requires at least one divalent cation within the active site of kinases, typically Mg^{2+}, to lower the transition state barrier of the phosphoryl transfer reaction and enhance the binding affinity of ATP to protein kinases through electrostatic interactions.[22,23] In contrast, phosphatases catalyze the transfer of a phosphoryl group from a phosphoprotein to a water molecule. The latter initiates a nucleophilic attack on the covalently bound phosphorous which results in the release of an inorganic phosphate. Metal cations, such as Mg^{2+} and Mn^{2+}, play a catalytic role for many phosphatases through the activation of a water molecule.[24]

Phosphorylation events commonly occur on one of the three amino acids bearing a side chain hydroxyl group, among which serine is the most phosphorylated residue (86%), followed by threonine (12%) and tyrosine (2%).[25] Based on their ability to phosphorylate these residues, most kinases are classified into two families. Protein kinases acting on serine or threonine residues of protein substrates are referred to as serine/threonine kinases, whereas those that phosphorylate tyrosine residues are termed tyrosine protein kinases. A third class of kinases, known as dual specificity protein kinases, can phosphorylate protein substrates on serine, threonine and tyrosine residues. While kinases can also act on other residues, namely histidine and aspartate, these modifications are rare in eukaryotes and less stable.[26] Protein kinases rely on the local region surrounding the phosphorylation site for substrate recognition. These enzymes predominantly target residues within intrinsically disordered regions of proteins, which lack a well-defined three-dimensional structure and typically present a high content of polar and charged amino acids.[27] This preference enables kinases to shape the region of interest to a conformation that fits their

catalytic pocket. In addition, the presence of phosphorylated residues in disordered regions of proteins may facilitate their interaction with functional partners.[28] A schematic illustration of the phosphorylation process is shown in **Figure 2**.

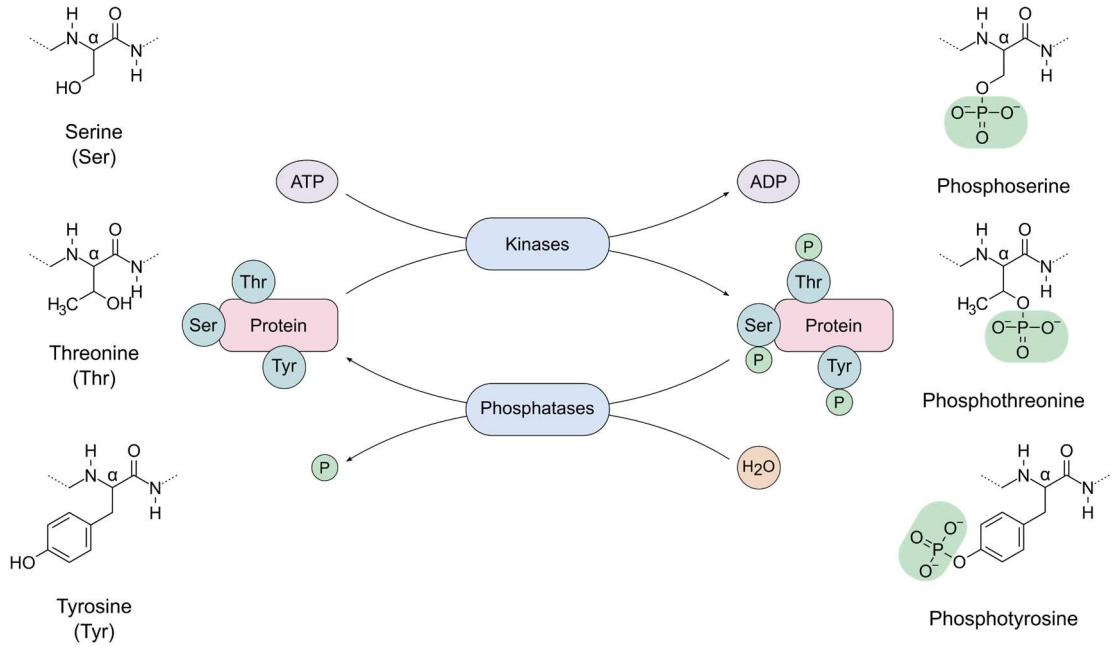


Figure 2. Schematic illustration of the phosphorylation process. The addition of a phosphoryl group mainly occurs on serine, threonine and tyrosine residues and is catalyzed by enzymes known as kinases. In the opposite direction, dephosphorylation is mediated by phosphatases. This figure was adapted from [9,26].

Phosphorylation can significantly impact the structure of target proteins. The presence of a bulky group carrying a double negative charge at physiological pH, a characteristic not shared by any other naturally occurring amino acid, enables phosphorylated residues to establish intra- and intermolecular hydrogen bonds and salt bridges. Two types of interactions between proteins and phosphorylated residues are prevalent. First, at high-affinity binding sites which are required to stabilize a conformational state, the phosphoryl group can interact with one or more arginine residues and form directed hydrogen bonds due to the rigid and planar structure of the guanidinium group. The resulting bonds are stronger and more stable than those occurring between arginine and either aspartate or glutamate.[29,30] Secondly, at less tightly bound sites, the phosphoryl group can leverage the positive charge of the α-helix dipole to interact with the backbone nitrogen atoms at the beginning of this secondary structure.[30]

The structural changes bestowed by phosphorylation and the reversible nature of this modification provide a regulatory mechanism that is crucial to control the function, stability and subcellular localization of proteins. The size and charge properties of covalently linked phosphoryl groups can introduce steric and electrostatic effects that prevent substrate access to catalytic sites of enzymes or disrupt interactions between proteins. Typically, phosphorylation also induces conformational rearrangements or allosteric transitions that enhance the catalytic activity of enzymes, as is the case of many protein kinases that depend on this PTM by upstream kinases for activation. In addition, phosphorylated substrates can be recognized by specific binding

domains of proteins, thereby acting as a molecular switch to promote inducible interactions between proteins in response to specific environmental cues. These phosphorylation-dependent interactions enable cells to dynamically integrate and respond to transient changes in their environment and are thus a central component of signal transduction pathways, mediating processes as diverse as cell proliferation, metabolism, and apoptosis.[31,32,33]

1.1.2 Ubiquitination

Ubiquitination is another important PTM that also occurs at both the N-terminus of proteins and at the ε-amino group of lysine residues. This modification signals substrate proteins for degradation by a regulated pathway known as the ubiquitin proteasome system (UPS).[14] At its center is an ATP-dependent macromolecular complex known as the 26S proteasome, which recognizes and catalyzes the degradation of ubiquitinated proteins.[34] The addition of ubiquitin, a small and highly conserved protein of 76 amino acids, to the free N-terminus of a target substrate proceeds through the same three-step cascade mechanism as ubiquitination of ε-amino groups of internal lysine residues in the sequence. The initial step, catalyzed by the ubiquitin-activating enzyme (E1), consists of the activation of the C-terminus of ubiquitin for its eventual nucleophilic attack by the amino group of the target substrate. This activation step results in the formation of a thioester bond between the carboxyl group of the C-terminal glycine of ubiquitin and a cysteine residue of E1. The intermediate step consists of the transfer of activated ubiquitin from E1 to another cysteine residue in the ubiquitin-conjugating enzyme (E2). In the final step, ubiquitin is covalently linked to the amino group of the substrate protein through the action of a ubiquitin-protein ligase (E3).[35]

The sequential activity of E1, E2 and E3 enzymes results in the conjugation of one ubiquitin molecule or mono-ubiquitination at the N-terminus or ε-amino group of a target protein. Additional molecules can be attached at their C-terminus by conjugation to any of the seven lysine residues (K6, K11, K27, K29, K33, K48, K63) or the N-terminus of another ubiquitin. The resulting polyubiquitin chain at the N-terminus or at the ε-amino group of the substrate protein serves as the recognition signal for its proteasome-mediated degradation.[34] A schematic illustration of the ubiquitination process at the N-terminal α-amino group of proteins and the ε-amino group of lysine residues is shown in **Figure 3**.

Lysine ubiquitination was originally described in a series of seminal articles between the late 1970s and early 1980s. Through the combined application of biochemical fractionation and enzymology techniques, it was found that certain proteins were covalently attached to ubiquitin when added to reticulocyte lysates, which are devoid of lysosomes. Furthermore, it was also observed that these ubiquitinated proteins were degraded by an ATP-dependent protease present in the extract.[36,57,37,38] Subsequently, the E1, E2 and E3 enzymes involved in the conjugation process of ubiquitin to target proteins were identified in the 1980s,[39,40] with the 26S proteasome characterized in the 1990s.[41]

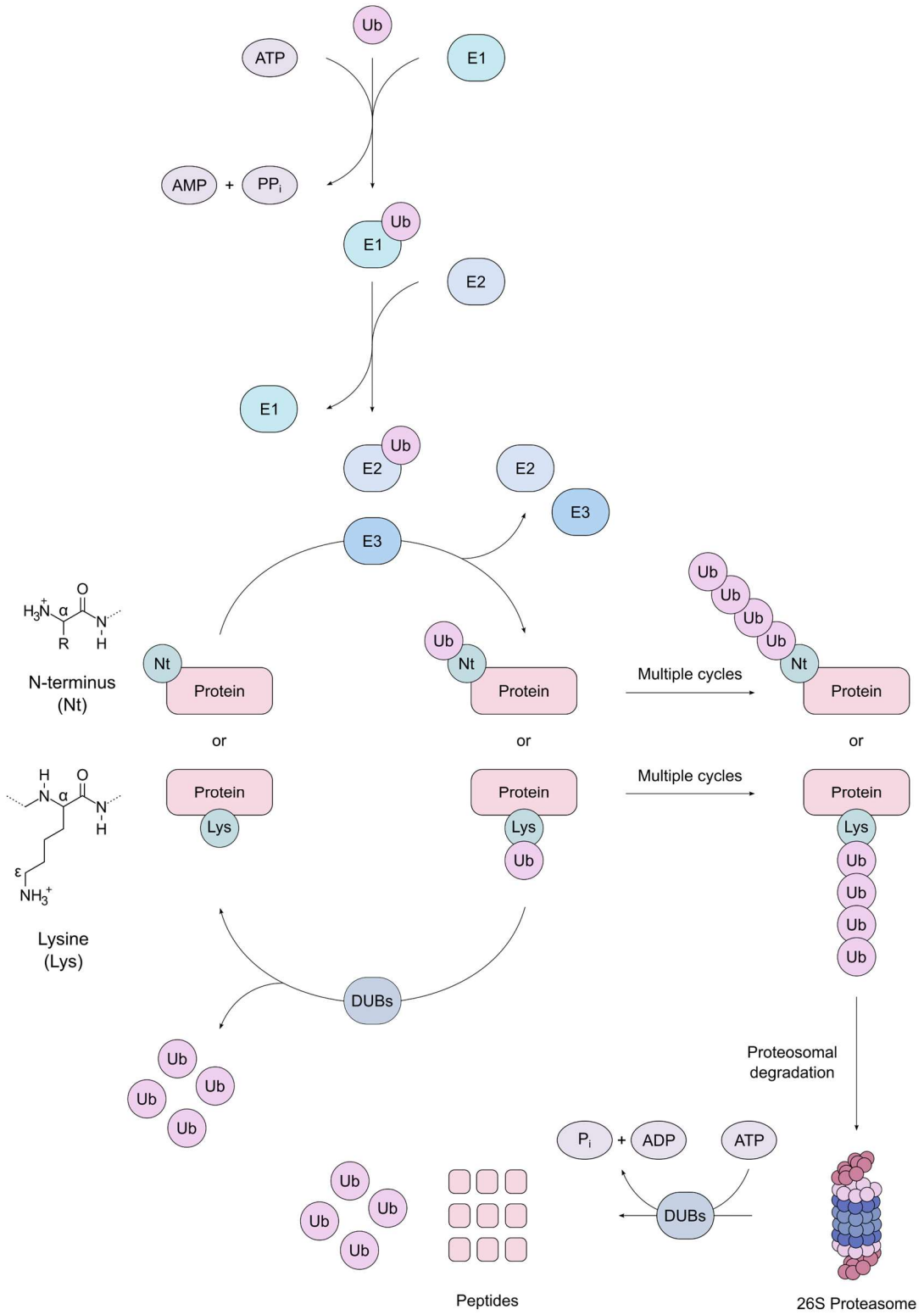


Figure 3. Schematic illustration of the ubiquitination process. The introduction of this PTM begins with the attachment of ubiquitin to the ubiquitin-activating enzyme (E1) in an ATP-dependent manner. The subsequent binding of ubiquitin to the ubiquitin-conjugating enzyme (E2) ultimately results in the transfer of the former to the N-terminal α-amino group or to the ε-amino group of lysine residues in target proteins through the action of a ubiquitin-protein ligase (E3). Multiple cycles of this process result in the polyubiquitination of the target protein, which serves as a substrate for its degradation by the 26S proteasome. This figure was adapted from [42,43,44].

The discovery of N^{α}-terminal ubiquitination (Nt-ubiquitination) occurred at a much later stage than its lysine counterpart. Specifically, Nt-ubiquitination was initially identified in 1998 in the MyoD protein, a transcriptional activator of myogenic gene expression during development. This protein displays a short half-life of approximately 45 minutes and its degradation is mediated by the UPS. Substitution of all its nine lysine residues did not significantly alter the degradation of MyoD, but modification of its α-amino group stabilized the protein, suggesting the existence of an alternative binding site for the ubiquitin protein.[45] Direct evidence that UPS-mediated degradation can be entirely dependent on Nt-ubiquitination was obtained by mass spectrometry (MS) of the human papillomavirus 58 E7 oncoprotein, a naturally-occurring lysine-less protein.[46]

1.1.3 Acetylation

Acetylation is an evolutionary conserved and highly regulated PTM in which the acetyl group from a donor molecule is transferred to a target protein. In eukaryotes, the key metabolic intermediate acetyl-coenzyme A (acetyl-CoA) serves as the sole donor of acetyl groups. This molecule is composed of an acetyl moiety attached through a thioester bond to coenzyme A (CoA), which in turn is a derivative of vitamin B_5 and cysteine. Since a thioester is a high-energy bond, the transfer of an acetyl moiety to amino groups in proteins is facilitated by the chemical structure of acetyl-CoA.[47] Protein acetylation typically occurs at the α-amino group of the N-terminus or at the ε-amino group of lysine residues and together these two forms represent the cell-wide acetylome.[48] A schematic illustration of this process is shown in **Figure 4**.

N^{α}-terminal acetylation (Nt-acetylation) is present in more than 50% and 80% of cytosolic yeast and human proteins, respectively, making it one of the most abundant modifications in eukaryotes.[49] N-terminal acetyltransferases (NATs) catalyze the transfer of an acetyl moiety from acetyl-CoA to the α-amino group of the target protein. By neutralizing the positive charge of the free α-amino group and increasing the hydrophobicity of the protein, Nt-acetylation alters the electrostatic and steric properties of the N-terminus while also blocking it to future modifications. Since no N-terminal deacetylases have been discovered, this PTM is considered irreversible.[50] Most NATs acetylate a variety of proteins and thus establishing a causal relationship between Nt-acetylation and specific cellular phenotypes has remained a challenge.[51] While the biological role of Nt-acetylation is not yet fully elucidated, studies on individual proteins as well as knockouts of single NATs have demonstrated that this modification regulates protein turnover, aggregation and targeting.[14,51]

Historically, Nt-acetylation was regarded as a PTM that protects proteins against ubiquitin-mediated degradation.[52] However, this view was recently challenged by the discovery that this modification can trigger protein degradation by the ubiquitin-dependent N-end rule pathway.[53] The N-end rule states that the *in vivo* half-life of a protein is dictated by the identity of its N-terminal residue. A degradation signal that is recognized by the N-end rule pathway and whose main determinant is a modified or unmodified destabilizing N-terminal residue is known as an N-degron. In eukaryotes, the N-end rule pathway comprises two branches. The first, termed Arg/N-end rule pathway, recognizes unacetylated N-terminal residues.[54] The second, termed Ac/N-end rule

pathway, targets for degradation Nt-acetylated proteins with methionine or the small uncharged alanine, valine, serine, threonine or cysteine residues at their N-termini. These acetylated N-terminal residues act as a specific degradation known as ^{Ac}N-degron.[53]

Since most eukaryotic proteins are Nt-acetylated and this PTM is irreversible and primarily co-translational, then many also possess an ^{Ac}N-degron from the moment they emerged from the ribosome. Simultaneously, however, many proteins are destined to have relatively long half-lives. To reconcile these seemingly conflicting observations, it has been proposed that Nt-acetylation is part of a quality control mechanism in cells that targets incorrectly folded proteins for degradation and regulates subunit stoichiometry in oligomeric complexes. In this context, if a specific protein protects its N-terminal domain in time, whether by acquiring a particular folding, interacting with a cognate binding partner or becoming assembled into a multi-subunit complex, then its ^{Ac}N-degron would be inaccessible for targeted degradation.[53,55]

It should be noted that the functional role of Nt-acetylation does not appear to be absolute for most proteins. It is possible that this modification is only essential to modulate the activity or stability of some proteins, while the remainder are acetylated by chance due to having the same consensus sequences.[56] This might explain why Nt-acetylation can be either complete or partial, with acetylated and unacetylated forms of the same protein existing in the latter situation.[4] Furthermore, the specificity of this Nt-acetylation can be influenced by other determinants. For instance, it has been shown that a proline residue in the first or second position can prevent Nt-acetylation in *Drosophila melanogaster*, in what was named as the (X)PX rule.[57]

The other main type of protein acetylation occurs at the ε-amino group of lysine residues. Lysine acetylation was first described in histones, a family of small, basic and highly conserved proteins that are essential to organize the negatively charged DNA in the nucleus of eukaryotic cells.[58] In 1964, Vincent Allfrey and colleagues made a significant breakthrough by showing that acetylation of histones reduces the ability of these proteins to inhibit RNA synthesis by the DNA-dependent RNA polymerases. This key finding led to their prescient hypothesis that histone acetylation, among other modifications, can serve as a molecular switch to regulate gene transcription.[59,60] Due to its initial discovery in histones, the enzymes that catalyze protein acetylation were named histone acetyltransferases (HATs). However, since acetylation events are not exclusive to histones, these enzymes have since been renamed lysine acetyltransferases (KATs).[61] In addition to histones, other known substrates of KATs include transcription factors, cellular signaling regulators, cytoskeleton proteins and metabolic enzymes.[62,63,64] For instance, nonhistone lysine acetylation occurs in the tumor suppressor protein p53, a transcription factor that regulates the activity of genes involved in several biological processes, such as cell cycle arrest, DNA repair, autophagy, metabolism and apoptosis.[65,66]

The status of acetylated lysine residues can be reversed by enzymes known as lysine deacetylases (KDACs). Thus, in contrast to the Nt-acetylation of proteins, lysine acetylation is a dynamic and reversible chemical modification. KDACs are primarily categorized into four distinct classes based on their sequence homology. Classes I, II and IV are commonly referred to as classical histone deacetylases (HDACs), even though most of these exhibit activity against other

substrates. Their catalytic mechanism depends on the presence of a Zn^{2+} ion to weaken the acetamide bond and promote the nucleophilic attack by a water molecule. In contrast, class III KDACs are known as sirtuins and these rely on nicotinamide adenine dinucleotide (NAD^+) as a cofactor to catalyze the cleavage of acetyl groups from lysine residues. While the classical HDACs generate an acetate molecule after hydrolysis of an acetyl moiety from a lysine residue, sirtuins break the nicotinamide ribosyl bond of NAD^+ and transfer the acetyl group from proteins to adenosine diphosphate (ADP) ribose.[67] The dynamic interplay between KATs and KDACs is a complex and finely tuned process that is essential for many cellular functions. Due to the consumption of acetyl-CoA and NAD^+ by KATs and sirtuins, respectively, protein acetylation is closely connected to metabolic processes and energy homeostasis. As a result, dysfunction of acetylation homeostasis is associated with pathological conditions, particularly cancer and neurodegenerative diseases.[68,69]

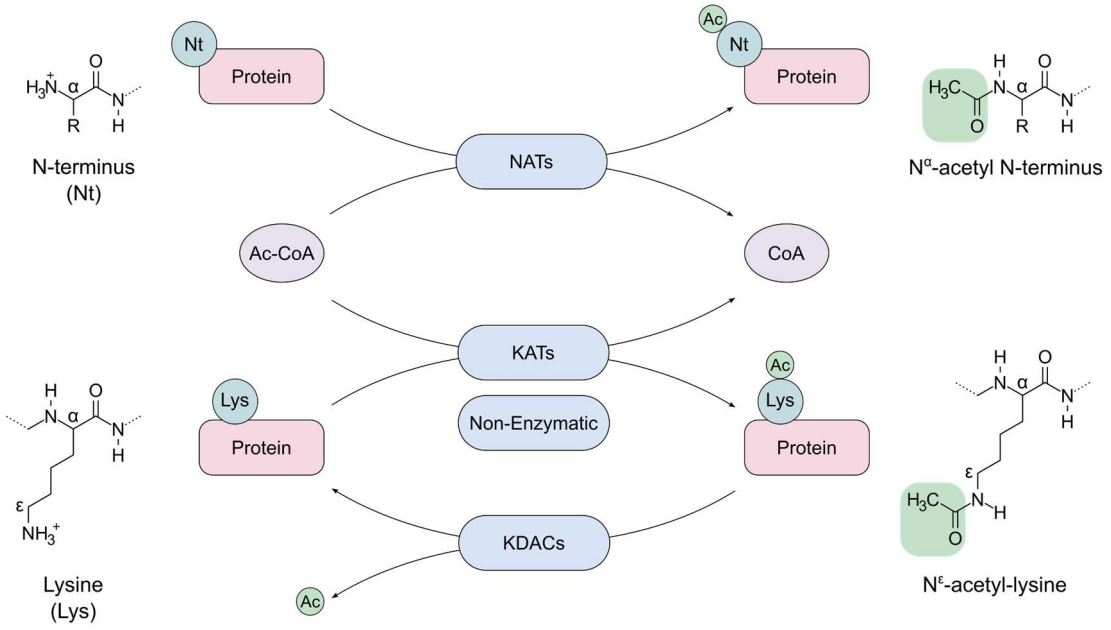


Figure 4. Schematic illustration of the acetylation process. This PTM mainly occurs at the α-amino group of the N-termini of proteins or at the ε-amino group of lysine residues. Acetylation can occur enzymatically (through the action of NATs at the N-terminus or KATs at lysine residues) or spontaneously, with acetyl-CoA acting as the sole donor for both cases in eukaryotes. Removal of the acetyl moiety at lysine residues occurs via action of KDACs. This figure was adapted from [9,70].

In addition to the acetylation promoted by KATs, lysine residues can also be spontaneously acetylated via an acetyl transfer from acetyl-CoA to the protein. This non-enzymatic reaction is favored by the abundance of acetyl-CoA in alkaline environments, such as the mitochondria. Since lysine residues can be naturally deprotonated at a basic pH, which increases their nucleophilicity, the mechanism initially proposed for this spontaneous reaction was a nucleophilic attack by the side chain amine on the acetyl carbonyl.[71] However, this direct acetylation reaction will be slow even in the slightly alkaline mitochondrial matrix (pH ~ 7.8) because cysteine (pK_a ~ 8.5) is typically a better nucleophile than lysine (pK_a ~ 10.5). Therefore, a second mechanism was

proposed in which acetylation occurs initially at a cysteine residue to form an acetylated thiol intermediate, followed by the subsequent transfer of the acetyl moiety to a proximal lysine.[72]

1.1.4 Methylation

Methylation is another widespread PTM that consists of the addition of one or more methyl groups to a target protein. This modification is one of the smallest in proteins and therefore does not significantly increase the "steric bulk" of the modified residues. Methylation is catalyzed by enzymes generally known as protein methyltransferases (PMTs), which can modify a variety of chemical moieties including the N-terminal α-amino and C-terminal carboxyl groups, as well as the side chains of specific residues in proteins. While the most frequently methylated side chains are those of lysine and arginine residues, protein methylation has also been described in other amino acids (methionine, cysteine, histidine, glutamine, asparagine, glutamate and aspartate).[73] PMTs catalyze the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to specific protein substrates, generating methylated proteins and S-adenosyl-L-homocysteine (SAH) as a by-product. [74,75] A schematic illustration of the methylation process at the N-terminal α-amino group of proteins and the side chains of lysine and arginine residues is shown in **Figure 5**.

Protein methylation was originally described by Richard Ambler and Maurice Rees in 1959 on lysine residues during the analysis of bacterial flagellar proteins.[76] Soon after in 1964, the occurrence of N^{ε}-methyl-lysine was observed in low and heterogeneous amounts in different histone fractions.[77] Lysine methylation is a reversible and dynamic process that is regulated by the balanced activities of opposing enzymes. Protein lysine methyltransferases (PKMTs) catalyze the transfer of up to three methyl groups to the ε-amino group of lysine residues in proteins, resulting in the formation of mono-, di- and trimethylated lysine. Within the human proteome, most PKMTs contain a "SET domain" that primarily targets lysine residues in the flexible N-terminal tails of histones. Nonetheless, most methyltransferases belong to the seven-β-strand (7βS) family, which is larger and structurally more diverse and can methylate both lysine and arginine residues in proteins, as well as small metabolites, lipids and nucleic acids.[78,79]

Protein lysine demethylases (PKDMs) drive the opposite reaction of PKMTs and are thus responsible for the removal of methyl groups from the ε-amino group of lysine residues. Based on their mechanism of action, PKDMs are typically classified into two major groups, namely the smaller lysine-specific demethylase (LSD) family and the larger Jumonji C (JmjC) domain family.[80] The first consists of flavin-adenine dinucleotide (FAD)-dependent amine oxidases which use FAD as a cofactor to catalyze the removal of a methyl group. Since these enzymes rely on the presence of a lone pair of electrons at the methylated amine, their demethylation activity is limited to mono- and di-methylated lysine residues. The second family comprises JmjC domain-containing proteins which employ an oxygenase mechanism to demethylate all three possible methylated lysine states.[81]

Methylation also regularly occurs at the side chains of arginine residues, which can receive up to two methyl groups on its guanidino group. Three distinct states of methylated arginine residues have been described in eukaryotes, with the most common being the asymmetric di-

methyl arginine (ADMA) in which two methyl groups are added to the same terminal (ω) guanidino nitrogen atom. The other two states correspond to the mono-methyl arginine (MMA) in which only one methyl group is added and the symmetric di-methyl arginine (SDMA) in which two methyl groups are attached to both ω-nitrogen atoms of the guanidino group.[82]

Arginine methylation is catalyzed by PMTs known as protein arginine methyltransferases (PRMTs), which can be classified based on which methylarginine product is formed. Type I, II and III PRMTs catalyze the formation of MMA. In addition, type I can catalyze the generation of ADMA from MMA, whereas type II can produce SDMA from MMA. Type IV enzymes have only been described in yeast and plants and catalyze the formation of a less common state of methylarginine in which the methyl group is added to the δ nitrogen atom of arginine.[83,84] Although initially considered to be an irreversible modification, it has been demonstrated that arginine methylation can be reversed by a subset of JmjC PKDMs.[85,86]

Although lysine and arginine methylation can present themselves in various distinctive structural forms, the overall +1 formal charge of both residues is maintained at physiological pH values throughout each methyl addition. Instead, lysine and arginine methylation generates an epitope for molecular recognition by sequentially reducing the hydrogen bonding capacity of these residues, thereby increasing the hydrophobicity of the ε-amino and guanidino groups with each proton removal and corresponding methyl addition.[87]

Similar to other PTMs such as phosphorylation and acetylation, methylation at lysine and arginine residues plays an important regulatory role in many fundamental cellular processes, ranging from gene transcription to signal transduction. In contrast to histone acetylation which is generally associated with transcriptional activation, methylation can promote both activation or repression of gene expression, depending on the specific methylated sites and the number of methyl groups added. For instance, each of the three possible methylation states at position K4 of histone H3 is correlated with transcriptional activation, whereas methylation at K9 of H3 is associated with transcriptional silencing.[88,89] In addition, asymmetric di-methylation of H3 at R2 constitutes an epigenetic mark for transcriptional repression, whereas mono-methylation and symmetric di-methylation at this position are associated with active transcription.[90,91]

As mentioned above, methylation also occurs at the N-termini of proteins (Nt-methylation) and this process is catalyzed by N-terminal methyltransferases (NTMTs). The existence of Nt-methylation has been known since 1976, when it was first reported in two *Escherichia coli* (*E. coli*) ribosomal proteins.[92] However, its functional importance has only recently unfolded following the discovery of the first eukaryotic N-terminal methyltransferases in 2010.[93,94] The chemical properties of Nt-methylated proteins depend on the degree of methylation at the nitrogen atom. Mono-methylation has a minor impact on the basicity of the α-amino group, by slightly increasing its pK_a and introducing some steric hindrance that can reduce its reactivity. In contrast, nitrogen quaternization introduced by trimethylation (or demethylation in the case of proline) generates a permanent positive charge, thereby abolishing the nucleophilicity of the N-terminal α-amino group.[92] Since no N-terminal demethylase has been found to date, this PTM is believed to be irreversible.[95]

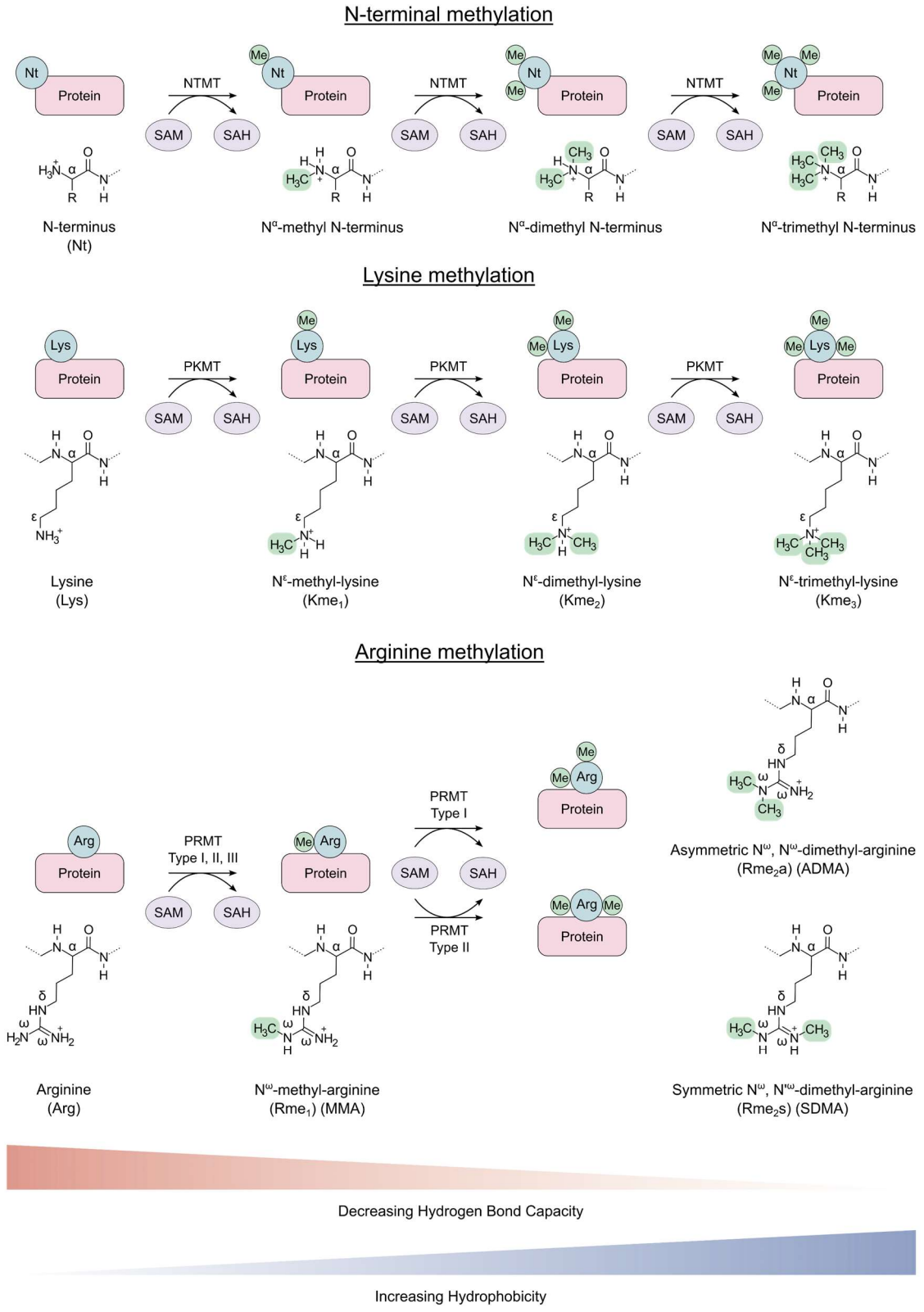


Figure 5. Schematic illustration of the methylation process. This PTM mainly occurs at the N-termini of proteins, the ε-amino group of lysine residues and the terminal ω-guanidino nitrogen atoms of arginine residues. Methylation is catalyzed by PMTs, which are termed NTMTs, PKMTs and PRMTs in the case of N-terminal methylation, lysine methylation and arginine methylation, respectively. At the N-terminus and lysine residues of proteins, the addition of methyl groups to the same nitrogen atom can occur up to three times, while in the case of arginine methylation the addition of methyl groups can occur up to two times at the same nitrogen atom. The consecutive addition of methyl groups results in a decrease of hydrogen bond capacity and a concomitant increase in hydrophobicity. This figure was adapted from [87,96,97].

The biological significance of Nt-methylation was initially identified in the RCC1 (Regulator of Chromatin Condensation 1) protein, a guanidine nucleotide-exchange factor for Ran GTPase. This PTM promotes association of RCC1 with chromatin during mitosis and is critical to prevent genomic instability, with the permanent positive charge of the α-amino group facilitating its electrostatic interaction with the negatively charged phosphate groups of DNA.[98] The first eukaryotic NTMT was identified as the enzyme that methylates the N-terminus of RCC1 and was thus named NRMT1 (N-terminal RCC1 Methyltransferase 1).[93] A second eukaryotic NTMT, named NRMT2, was later identified.[99] Both enzymes recognize and methylate the "canonical" X-Pro-Lys consensus sequence (with X being any amino acid other than leucine, isoleucine, tryptophan, aspartate or glutamate). However, the presence of the proline residue at the second position is not strictly required and can be substituted by alanine, glutamate, methionine, asparagine, glutamine, glycine and serine. The lysine at third position is crucial for efficient methylation but can be substituted by arginine in certain cases.[100] While both enzymes display similar tissue expression and subcellular localization patterns, predominantly in the nucleus, their catalytic functions differ. NRMT1 can catalyze the transfer of one, two and three methyl groups to its substrates, while NRMT2 is mostly an N-terminal mono-methyltransferase. On the basis of their overlapping localization and expression patterns, as well as their differing catalytic specificities, it has been postulated that both enzymes work synergistically, with NRMT2 priming substrates for subsequent NRMT1-mediated di- and trimethylation.[99]

In addition to its role in protein-DNA binding, many Nt-methylated proteins are components of multi-subunit complexes, suggesting that Nt-methylation is also important to mediate protein-protein interactions. This PTM is also protective against digestion by cellular aminopeptidases and may compete with the modifications associated with the N-end rule pathway, which indicates a role in protein stability.[99]

1.1.5 Methionine excision

In eukaryotes, nascent polypeptide chains usually emerge from the ribosomal exit tunnel with an initiator methionine (iMet) at their N-terminus. In over half of the proteins of any proteome, this residue is removed in an evolutionary conserved process termed N-terminal methionine excision (NME).[101] Cleavage of iMet by methionine aminopeptidases is greatly influenced by the adjacent amino-acid residues and generally occurs when small residues such as alanine, cysteine, glycine, proline, serine, threonine or valine appear on the second position of the sequence. The side chains of these amino acids are uncharged with a radius of gyration of 1.29 Å or less. In contrast, when the second position is occupied by any of the other naturally encoded amino acids, then NME usually does not occur and the iMet is fully retained in the sequence.[4] While the functional implications of NME are not completely understood, this process is considered to be the greatest source of N-terminal amino acid diversity.[102]

1.2 Histones as model proteins to study post-translational modifications

Within the nucleus of eukaryotic cells, a family of small and basic proteins termed histones are essential to organize the negatively charged DNA into a complex known as chromatin. In addition to enabling the packaging of DNA into a compact structure, these highly conserved proteins are also crucial to regulate its accessibility to transcription, replication and repair machineries. To accomplish this, histones are extensively and dynamically decorated on specific residues with various PTMs. Such chemical modifications, which include phosphorylation, acetylation and methylation, are enzymatically added by “writers” and removed by “erasers”. Histone PTMs constitute an essential epigenetic mechanism that can modulate chromatin structure and function directly by influencing the strength of the electrostatic interactions with DNA or indirectly by recruiting chromatin-associated proteins, known as “readers”.[103]

Therefore, histone modifications are orchestrated by an intricate interplay of regulatory enzymes, whose combined actions ultimately determine the type and stoichiometry of PTMs installed at each possible site. Due to their pivotal role in the epigenetic regulation of the eukaryotic genome, which is modulated by a wide range of PTMs, histones can be considered as model proteins to study the structural and functional effects of these chemical modifications. Indeed, histones have served as invaluable targets to develop novel *in vitro* chemical strategies to generate homogeneous and site-specifically modified proteins. These modified histones can then be reconstituted into the basic repeating units of chromatin, termed nucleosomes, to perform functional and structural assays that elucidate the underlying molecular mechanisms by which PTMs control chromatin dynamics.

1.2.1 Overview of nucleosome structure and dynamics

Each nucleosome contains a core particle that is connected to the adjacent one by a segment of linker DNA. The latter is typically associated with linker histone protein (H1 or H5) which helps organize nucleosomes into higher-order chromatin structures. In the nucleosome core particle, 145 to 147 base pairs of DNA are tightly wrapped around a histone octamer, which is composed of two copies each of the canonical histones H2A, H2B, H3 and H4.[104] Even though their primary sequence homology is low, these proteins share a common structural motif termed the histone-fold through which the core histones can interact to form four heterodimers of H2A-H2B and H3-H4. The two H3-H4 dimers constitute a $(H3\text{-}H4)_2$ tetramer which is flanked by two H2A-H2B dimers to establish a tripartite assembly.[105] In addition, each of the core histones contains an N-terminal tail of variable length. Together with the C-terminal tails of H2A, these flexible and unstructured regions protrude from the compact nucleosome and constitute the main targets for a wide variety of PTMs.[106] Therefore, the interactions with DNA occur mostly at the structured regions of histones, while the tails interact primarily with neighbouring nucleosomes or nuclear factors.[107]

The packaging of eukaryotic DNA into nucleosomes serves several functions. First, it constitutes the first level of genomic compaction, protecting nuclear DNA from nucleases and limiting the access of transcription factors. Nucleosomes can also self-assemble to form higher-

order chromatin structures, thereby enabling additional compaction of the genome. Second, these structures display a wide variety of PTMs, whose combinatorial presence acts as a signaling scaffold that modulates the recruitment of chromatin-associated proteins.[104] Thus, nucleosomes play the seemingly divergent role of condensing eukaryotic DNA while allowing its regulated access to distinct factors. This is achieved through the dynamic properties of nucleosomes. A schematic illustration of the nucleosome and primary chromatin structures is shown in **Figure 6**.

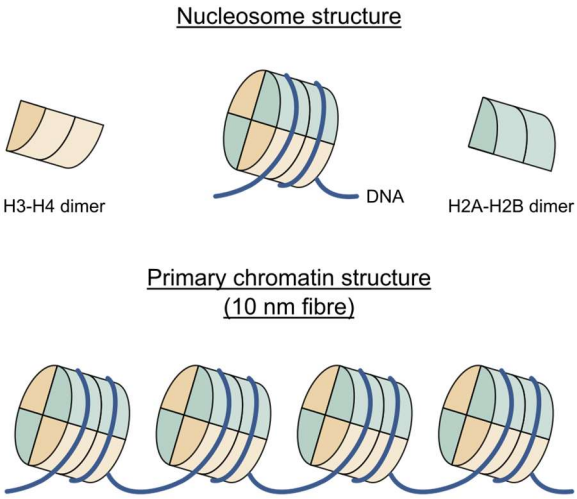


Figure 6. Schematic illustration of the nucleosome and primary chromatin structures. The structure of a nucleosome consists of two copies each of the canonical histones H2A, H2B, H3 and H4. The two H3-H4 dimers (shown in yellow) form a $(H3\text{-}H4)_2$ tetramer which is flanked by two H2A-H2B dimers (shown in green) to establish the histone octamer, which is wrapped by 145 to 147 base pairs of DNA (shown in blue) to form a nucleosome. Nucleosomes with canonical and histone variants can bear distinct PTMs in both type and stoichiometry, thus expanding the diversity of possible nucleosome structures. These can assemble into nucleosomal arrays to form the primary chromatin structure known as the 10 nm fiber. This figure was adapted from [107].

Nucleosomes are highly dynamic structures that can adopt distinct conformational states from those observed in X-ray crystal studies. For instance, nucleosomes can change their positioning by sliding along DNA or disassemble, either partially or completely, by histone eviction and be reconstructed later. This is a particularly relevant occurrence during DNA-templated events, such as transcription, replication and repair. Additionally, even in the absence of these events, DNA can transiently disengage from the nucleosome, a process known as nucleosome breathing. This dynamic behavior, in which a basal level of nucleosomal association-disassociation exists, provides a mechanism by which factors with sufficiently high affinity and local concentration can compete with histones for the binding of DNA.[108,109]

Nucleosome organization can be characterized by a combination of both occupancy and positioning. Although distinct, the concepts of occupancy and positioning are interconnected. Occupancy refers to the average number of nucleosomes that exist in a specific genomic region in a population of cells and thus can be considered as a measure of the probability that a nucleosome is present at a given DNA sequence. Positioning is defined as the probability of finding a reference point of a nucleosome at a given genomic coordinate relative to the surrounding coordinates in a particular population of cells. Nucleosome dynamicity is regulated by a plethora of factors that can influence the occupancy and positioning of these structures on

DNA, including DNA sequence and associated processes, such as transcription, replication and repair, as well as subnucleosomal composition and the action of multiple regulatory factors.[110]

The DNA sequence can affect nucleosome dynamicity by influencing the positioning of this structure. While site-specific interactions do not exist between histones and DNA bases, octamers display considerable preference for some sequences over others. This preference is primarily determined by the propensity of a DNA sequence to bend around the histone octamer. Thus, nucleosome formation is favored when more bendable sequences are in contact with core histones and less bendable ones are solvent-exposed. Empirically, it has been demonstrated that the presence of particular dinucleotides at the face of the helical repeat that interacts directly with histones can stabilize nucleosomes. These dinucleotides, which are typically AA/TT/TA dimers that occur with a periodicity of 10 base pairs, promote the bending of the double-helix around the histone octamer.[111,112] In addition, homopolymeric stretches of poly(dA:dT) and poly(dG:dC) are strongly resistant to nucleosome deposition due to the intrinsic rigidity of these elements.[113] In particular, poly(dA:dT) sequences are abundant in eukaryotic genomes and are associated with nucleosome depletion at promoters, thereby facilitating the accessibility of binding factors to these DNA regions and enhancing transcriptional activity.[114,115]

Through the incorporation of histone PTMs and variants, the composition of nucleosomes can also influence the dynamicity of these structures. A well-known example of the effects of histone PTMs on nucleosome dynamics is given by the acetylation of lysine 56 of histone H3 (H3K56Ac). This epigenetic mark is particularly abundant in yeast and occurs predominantly during the S phase of the cell cycle, as newly synthesized H3 molecules are assembled into nucleosomes following DNA replication. Unlike other acetylated sites, K56 is not located within the N-terminal tail and is instead positioned at the end of the αN-helix, a unique structural element that exists only in H3 and precedes the histone-fold domain. As a result, this residue is located at the entry and exit points of DNA on the nucleosome, where its ε-amino groups form water-mediated contacts with the phosphodiester backbone. Acetylation at H3K56 weakens histone-DNA interactions at the entry and exit points on the nucleosome, facilitating the breathing of these structures and regulating chromatin structure at a higher order level.[116,117] In addition, H3K56Ac plays a crucial role in genomic stability by creating a favorable environment that facilitates chromatin assembly following DNA replication and repair, as well as nucleosome disassembly at promoters during transcriptional activation.[118,119]

Histone proteins are typically divided into two groups. The first is composed of replication-dependent histones whose expression and deposition into nucleosomes is limited to the S phase of the cell cycle when DNA synthesis occurs. This group encompasses the canonical histones H2A, H2B, H3 and H4 that comprise the nucleosome core particle, as well as the linker histones that bind to the linker DNA positioned between nucleosomes. In contrast, the second group includes variants of the canonical histones whose expression and deposition occurs at relatively low levels throughout all stages of the cell cycle. These variants can replace missing histones or be recruited to particular genomic locations to carry out specialized functions.[120,107] The replacement of replication-dependent histones with replication-independent variants influences

the dynamics of nucleosomes, as well as the organization of these structures in chromatin and their interactions with regulatory factors. For instance, the incorporation of the histone H3 variant known as CENP-A (Centromere Protein-A) in vertebrates into nucleosomes rather than canonical H3 is essential for the epigenetic specification of centromeric chromatin.[121]

Nucleosome dynamicity is also in part governed by the action of various regulatory factors, including histone chaperones, chromatin modifiers, and polymerases. Nucleosomes can be reconstituted *in vitro* by combining histone octamers and DNA at high salt concentrations (such as 2 M NaCl or KCl) and gradually dialyzing the mixture into a low ionic strength buffer.[122,123] However, insoluble aggregates are formed when histones and DNA are mixed directly in solution at physiological conditions. Studies by Laskey and colleagues showed that this precipitation is averted and nucleosome assembly is promoted in an ATP-independent manner when cell-free extracts from unfertilized *Xenopus laevis* (*X. laevis*) eggs are added. These contain large amounts of histone chaperones that prevent promiscuous interactions of histones with any acidic macromolecules in the cell, including DNA. This discovery led to the identification of the first histone chaperone known as nucleoplasmin in 1978.[124] Curiously, the term “chaperone” was subsequently adopted to broadly describe a set of proteins that assist in the proper folding of others, such as the well-known heat-shock proteins.[125]

Histone chaperones play essential and specific functional roles during the various stages of the nucleosome assembly process. Following cytosolic synthesis, histones are translocated to the nucleus with the aid of histone chaperones such as NAP-1 (Nucleosome Assembly Protein 1) and ASF1 (Anti-Silencing Function 1), which partially act by modulating the importin-histone complex interaction.[126,127] In addition, the existence of a soluble reservoir of histones that are available to respond to adverse circumstances, such as histone overload and acute replication stress, is crucial. Certain histone chaperones, such as NASP (Nuclear Autoantigenic Sperm Protein), regulate the supply of histones by protecting these proteins against degradation.[128] Histone chaperones, such as ASF1, can also serve as intermediates to facilitate the interaction of histones and histone-modifying enzymes.[129] Finally, histone chaperones, such as CAF-1 (Chromatin Assembly Factor-1), also participate in the process of histone deposition onto DNA, thus actively contributing to the assembly of nucleosomes.[130]

Chromatin modifiers are essential modulators of nucleosome structure and dynamics, constituting the main co-regulatory complexes associated with gene transcription. These can be divided into two functionally distinct categories, namely ATP-dependent chromatin remodeling complexes and histone-modifying complexes.[131] The former complexes are large (>1 MDa), multi-component and highly conserved molecular machineries among eukaryotes.[132] As their name implies, such complexes rely on the energy from ATP hydrolysis to physically change chromatin structure by sliding, ejecting or structurally altering nucleosomes. Remodeling complexes are essential to modulate several DNA-associated processes through their involvement in the packaging of the genome, the establishment of specialized chromatin regions and the control of DNA accessibility to enzymatic machineries.[133] For instance, during DNA replication nucleosomes located ahead of the replication fork need to be disassembled, while new

ones containing both parental and *de novo* synthesized histones must be deposited on daughter strands behind the replication fork.[134] In this context, specific chromatin remodeling complexes are recruited to either evict or guide the histone octamer around the advancing DNA polymerase and properly space nucleosomes following DNA replication.

ATP-dependent chromatin remodeling complexes can be classified into four different families, namely SWI/SNF (SWItch/Sucrose Non-Fermentable), ISWI (Imitation SWItch), CHD (Chromodomain Helicase DNA-binding) and INO80 (INOsitol requiring 80). While sharing a conserved ATPase domain, each of these remodelers plays specialized roles to control particular biological processes such as DNA replication and repair, transcriptional regulation, homologous recombination and chromosomal segregation.[135] The presence of distinct domains in their catalytic ATPases, as well as unique associated subunits enables remodeling complexes to selectively interact with and influence specific target nucleosomes, thereby shaping the chromatin landscapes that govern replication, repair, transcription and organization of the genome.[133,136]

The SWI/SNF family of chromatin remodelers promotes the sliding and eviction of nucleosomes by translocating DNA around the histone octamer.[137] It serves many functions, performing crucial roles in transcriptional regulation and tumor suppression, with nearly 25% of all malignancies containing mutations in one or more genes encoding subunits of the SWI/SNF chromatin-remodeling complexes.[138,139] Several ISWI complexes can assemble nucleosomes and optimize their spacing by adjusting the length of linker DNA, thereby restricting chromatin accessibility and gene transcription.[140] Nevertheless, some complexes can promote gene expression by mobilizing nucleosomes and randomizing their spacing.[133] The CHD family of chromatin remodelers is commonly lost or inactivated in various types of human cancers.[141] Some of these complexes can slide or evict nucleosomes to create nucleosome-depleted enhancer and promoter regions, thereby promoting transcription.[142,143] Others, however, are implicated in transcriptional repression as is the case of the Mi-2/NuRD (Nucleosomal Remodeling and Deacetylase) complex, which combines histone deacetylation and chromatin remodeling activities.[144] The INO80 family of chromatin remodelers has many functions, particularly during DNA transcription where it promotes the recovery of stalled replication forks and the eviction of RNA polymerase at sites of replication-transcription conflict during replicative stress.[145,146] In addition, the INO80 family of remodelers displays unique editing capabilities, mediating the exchange of canonical with variant histones and vice-versa.[147,148]

1.2.2 The “histone code”

Histone-modifying enzymes catalyze the addition and removal of histone PTMs and are often present in large protein complexes, working in tandem with other chromatin-associated proteins. Among others, these histone-modifying complexes may include, for instance, DNA-binding and epigenetic reader modules that target specific genomic regions and histone PTMs, respectively.[149] As already mentioned, histones are subject to a wide variety of PTMs at several of their residue sites and these modifications play an important regulatory role in DNA-templated processes, including gene transcription. In addition, these modifications are also associated with

establishing and preserving a heritable epigenetic code that influences gene expression without directly altering the underlying DNA sequence.[150] The current understanding concerning the functional role of histone PTMs is encapsulated by the “histone code” hypothesis, a concept first introduced by Bryan Turner in 1993 and later named as such by C. David Allis in the early 2000s.[151,152] The latter postulates that specific combinations histone PTMs, which are collectively determined through the action of several histone-modifying enzymes, act as docking sites for the recruitment of downstream regulatory proteins that subsequently modulate chromatin structure and function.[153] In this way, the “histone code” provides a molecular framework for understanding the complex interplay between histone PTMs and chromatin-associated processes.

A well-established example of the effects of histone PTMs on the recruitment of effector proteins that drive chromatin-associated processes is given by the acetylation of lysine 9 of histone H3 (H3K9Ac). This epigenetic mark is commonly found in the promoter regions of actively transcribed genes and is required to recruit TAFII250.[154,155,156,157] The latter is the largest subunit of a multiprotein complex known as transcription factor IID (TFIID), which plays a central role in transcription initiation in eukaryotic cells.[158] Additionally, H3K9Ac at select gene promoters also mediates transcription elongation by recruiting the super elongation complex, which is important to release paused RNA polymerase II just downstream of the transcription start site.[159]

The “histone code” hypothesis was initially investigated in the context of acetylation due to the identification of HATs and HDACs, the enzymes responsible for maintaining the steady-state equilibrium of this PTM, as transcriptional activators and repressors, respectively.[160] Later, the discovery of the bromodomain as the first protein module capable of selectively interacting with a particular epigenetic mark, in this case acetylated lysine residues in histone N-terminal tails, offered compelling evidence supporting the “histone code” hypothesis and further fueled the interest in this PTM.[161] Indeed, most nuclear HATs contain bromodomain modules that recognize acetylated lysine residues in histones.[162] Furthermore, in accordance with the combinatorial nature of the “histone code” hypothesis, TAFII250 contains two bromodomains in tandem in addition to a HAT domain. This arrangement allows the preferential binding of TAFII250 to di-acetylated histone peptides containing properly spaced acetyl-lysine residues.[158]

Bromodomains are one of many protein modules known to bind acetylated lysine residues in histones and their affinity is higher for regions with multiple, closely spaced acetylation sites.[163] These readers of histone acetylation consist of approximately 110 amino acids that are arranged in a left-handed fold of four α-helices (α_Z, α_A, α_B and α_C). In addition, its structure also contains two inter-helical loops of variable charge and length, with one connecting the α_Z and α_A helices and denoted as the ZA loop and the other connecting the α_B and α_C helices and denoted as the BC loop, which form a hydrophobic pocket that binds to the acetylated lysine residue of histones.[164] The human genome encodes 61 bromodomains which have been identified in 46 nuclear and cytoplasmic proteins. These include HATs, histone methyltransferases, helicases, ATP-dependent chromatin remodeling complexes, transcriptional co-activators, and the BET (Bromodomain and Extra-Terminal domain) family of proteins.[165] In addition to bromodomains, acetylated lysine residues in histones are also recognized by other domains, such as tandem

PHD (Plant Homeodomain) fingers,[166,167,168] double PH (Pleckstrin Homology) domain,[169] and the YEATS (Yaf9, ENL, AF9, Taf14, Sas5) domain.[170]

Other histone PTMs besides acetylation are also recognized by specific reader modules. For instance, methylation at histone lysine residues is “read” by members of a structurally related family of protein domains known as the Royal family.[171]

1.3 Strategies for the site-specific acetylation of histones

To answer fundamental questions related to the biological role of PTMs in healthy and diseased states, access to homogeneously modified proteins is essential. However, purification of endogenous proteins with the desired PTM from biological sources can be difficult to achieve due to the diversity of sites, states and types of possible modifications.[172] Enzymatic methods, while successfully employed to introduce PTMs such as acetylation and phosphorylation, also present several limitations. First, the enzyme responsible for a PTM at a particular position in the protein of interest may be unknown or difficult to prepare in their active form. Second, it may be difficult to drive enzymatic reactions, many of which are not site-specific, to completion which results in the formation of heterogeneous samples containing unmodified and modified proteins with a particular PTM at more than one location. Third, some PTMs - notably, the more complex ones such as ubiquitination and glycosylation - can involve more than one step in succession and thus require multiple enzymes until the actual modification is installed.[173] Thus, to overcome the challenges imposed by enzymatic or purification-only methods, the field of chemical biology has developed a wide range of strategies to install PTMs or their mimics with site-specificity.

As evidenced in the previous section, acetylation is one of the most significant PTMs in histones. Although many acetylation marks have been structurally and functionally characterized in these proteins, a biochemical understanding of how these modifications combine with others at specific sites to regulate chromatin structure and function is essential. However, access to homogenous samples of acetylated histones with site-specificity in the required quantities for these studies is difficult to achieve from biological sources or enzymatic methods. Traditional approaches to characterize the functional consequences of lysine acetylation relied on the use of recombinant mutants in which lysine residues are replaced with either glutamine or arginine. The glutamine substitution serves to mimic lysine acetylation due to the loss of positive charge resulting from the formation of an amide bond, whereas the arginine substitution mimics non-acetylated lysine by preserving the positive charge of this residue.[174]

Although widely employed due to its simplicity and applicability in both *in vitro* and *in vivo* scenarios, the substitution approach is not ideal due to structural differences between the intended residue (lysine or acetyl-lysine) and its respective analogue (arginine or glutamine).[175] For instance, the length of the glutamine side chain is substantially shorter than acetyl-lysine’s, while the side chain of arginine is bulkier with a different geometry compared to lysine’s and can establish a higher number of electrostatic interactions, such as hydrogen bonds and salt-bridges, due to its guanidinium group.[176,177] Thus, the results obtained from these substitutions should be thoroughly validated in each case by independently controlling for the acetylation of single lysine

residues.[177,178] To this end, it is imperative to develop a set of strategies to achieve site-specific acetylation of histones. In this context, several strategies have emerged and these can be grouped into four main categories, namely chemical or post-expression mutagenesis, genetic code expansion, protein synthesis by native chemical ligation (NCL) and expressed protein ligation (EPL), and the use of affinity ligands tethered to acetyl transfer catalysts. These strategies, which are schematically illustrated in **Figure 7**, display distinct advantages and disadvantages that complement each other to create a diverse set of tools for installing PTMs in histones.[179]

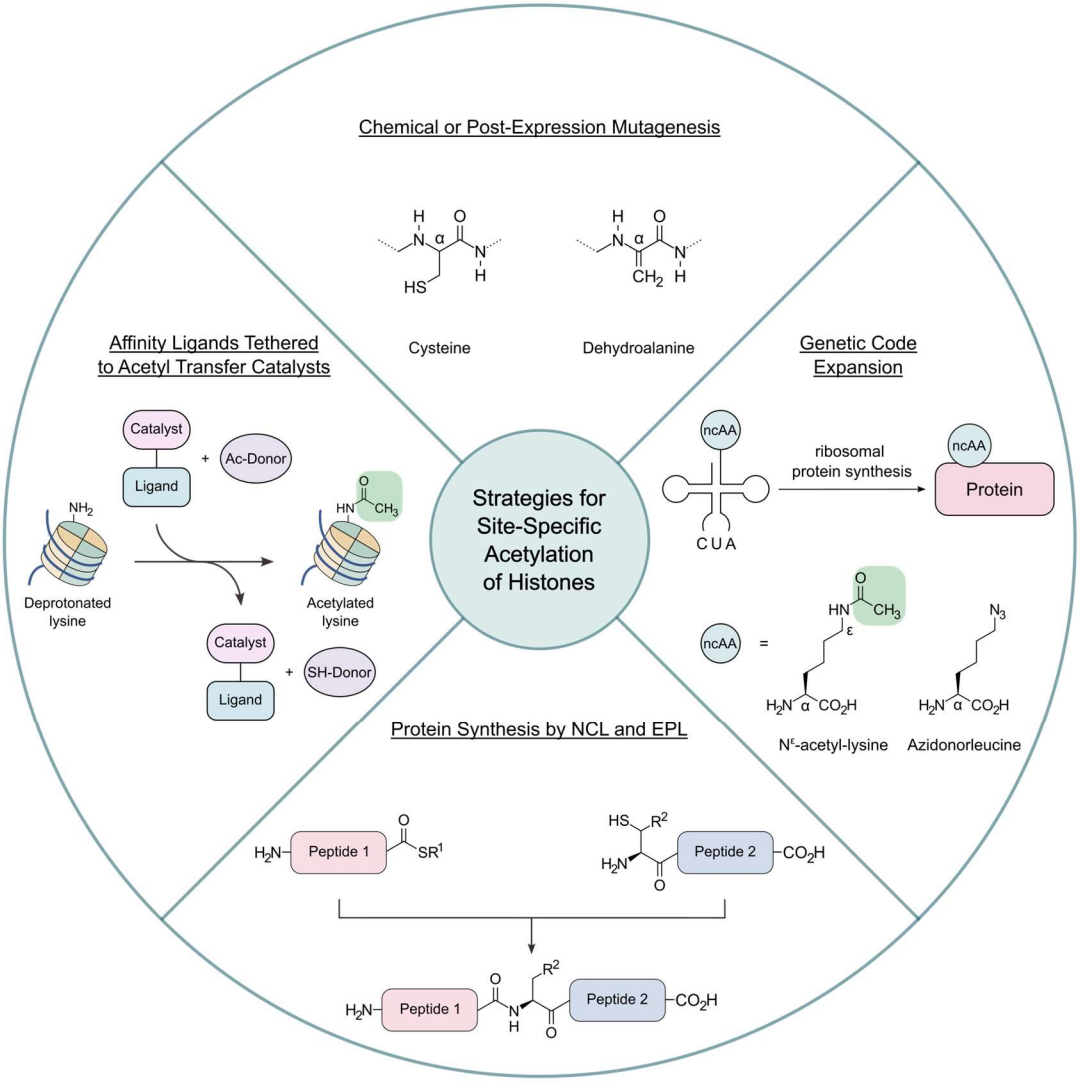


Figure 7. Schematic illustration of the existing strategies for the site-specific acetylation of histones. Lysine residues in these proteins can be acetylated with site-specificity by four main strategies, namely chemical or post-expression mutagenesis, genetic code expansion, protein synthesis by NCL and EPL, and the use of affinity ligands tethered to acetyl transfer catalysts. This figure was adapted from illustrations present in subsections of 1.3 describing the respective strategies.

1.3.1 Chemical or post-expression mutagenesis

Site-directed mutagenesis, a powerful and well-established molecular biology method to insert, substitute or delete specific nucleotides of a DNA sequence within a plasmid vector, is routinely used to change the amino acid composition of proteins. By enabling the conversion of

one residue into another or eliminating it altogether, this method constitutes an invaluable tool to probe the relative importance of a given amino acid in the structure and function of proteins.[180] However, traditional site-directed mutagenesis is limited to the twenty canonical amino acids that are found in proteins. With the technological advances that naturally arose with time in the field of protein science, a novel demand for expanding the chemical diversity afforded by the side chains of these canonical residues also emerged.[181]

The ability to install functional groups that are not directly accessible using site-directed mutagenesis and conventional expression systems enables experimentalists to create precisely modified proteins at will, something which is particularly relevant if one wishes to investigate the biological significance of natural PTMs. In this context, several strategies were developed to either incorporate non-canonical amino acids (ncAAs) or modify the side chains of existing residues. Genetic code expansion, NCL and EPL enable the introduction of ncAAs, such as acetyl-lysine, into proteins at precisely defined sites.[181] These strategies and their applications to the site-specific acetylation of histones are addressed in detail in later subsections of 1.3. An alternative strategy consists of using chemical methods to selectively and/or specifically modify the side chains of proteins following their translation in the ribosome.[181] The current subsection focuses on the latter strategy, known as chemical or post-expression mutagenesis, within the scope of installing acetylation marks in histones with site-specificity.

The concept of chemical mutagenesis in proteins was originally introduced in 1965 with the aim of converting the active serine of esterases into a cysteine residue. To evaluate the feasibility of this idea, serine-containing peptides were used. Following serine tosylation, reaction with thioacetate yielded the corresponding S-acetyl-L-cysteine peptides via an S_N2 displacement mechanism. Subsequent hydrolysis afforded the desired cysteine peptides. Furthermore, several non-canonical cysteine analogues were generated by reacting the tosylated serine peptides with other thiol nucleophiles.[182] Later in 1966, two groups independently described the chemical generation of a cysteine residue from the active serine of subtilisin. In both cases, sulfonylation of the active serine was achieved with phenylmethanesulfonyl fluoride to create a good leaving group which was displaced by thioacetate. Subsequent hydrolysis of the acetyl group from the thiol ester, possibly assisted by the intrinsic activity of the enzyme, afforded the desired thiol-subtilisin containing a cysteine rather than a serine residue at the active center.[183,184] This chemical conversion of a hydroxyl to a sulfhydryl group rendered the resulting thiol-subtilisin significantly less active than the native subtilisin.[184]

Owing to its low abundance (1.5%), the high nucleophilicity of the thiolate anion following deprotonation of its sulfhydryl group and ease of introduction by genetic mutation, cysteine is the natural amino acid of choice for chemical mutagenesis.[185] The selective modification of cysteine is especially relevant as an approach to introduce PTMs in histones, since only H3 contains such a residue which can be mutated to an alanine (C110A) by site-directed mutagenesis without loss of function. The cysteine-selective modification approach provides a relatively simple, economical and flexible avenue for installing distinct PTM mimics in histones, either via direct functionalization of the residue itself, or indirectly by converting it first into dehydroalanine (Dha).[186] In both cases,

if an original cysteine exists within the histone sequence at a position that is not of interest, it is necessary to mutate it to another residue which is typically an alanine due to its small size and non-polar nature. The amino acid at the acetyl-lysine position of interest is subsequently mutated to a cysteine.[187] In the direct route, this cysteine is alkylated to afford the histone containing the site-specific acetyl-lysine or its mimic in a single step. In the indirect route, the engineered cysteine is first converted to Dha. With its reactive α,β-unsaturated carbonyl, Dha can subsequently undergo a Michael addition reaction with an appropriate nucleophile to generate the histone containing the site-specific acetyl-lysine or its mimic.[188] In addition to cysteine, Dha can also be generated from other amino acids, as explained later in this subsection.

Direct functionalization via cysteine-selective modification

The cysteine-selective modification route has been successfully employed to introduce various PTMs and their mimics into proteins. A schematic illustration of the most relevant approaches described for installing acetyl-lysine and its mimics through direct functionalization of the cysteine residue is shown in **Figure 8**. One initial attempt at creating acetyl-lysine mimics used N-(2-bromoethyl)acetamide as the alkylating reagent, but conversion to the desired product was insignificant. The reaction with N-acetyl-aziridine proved to be equally unsuccessful due to the nucleophilic attack at the carbonyl instead, resulting in cysteine rather than lysine acetylation. With its reduced carbonyl reactivity, the use of methylthiocarbonyl-aziridine (MTCA) afforded methylthiocarbonyl-thialysine (MTCTK), the thiocarbamate mimic of acetyl-lysine (**Figure 8a**). While structurally different from acetyl-lysine, MTCTK was recognized by BRDT, a member of the BET family of proteins. However, removal of the methylthiocarbonyl group was not observed with class I and class III deacetylases.[189] A second, more sophisticated approach consisted of a radical-mediated thiol-ene coupling between the side chain of the cysteine residue and the commercially available N-vinyl-acetamide. This reaction afforded the desired acetyl-thialysine mimic, which differs from the natural acetyl-lysine only at the γ position, with excellent yield (> 90%) in histones H3K27C and H4K16C (**Figure 8b**). In addition, the generated H4K16C bearing the acetyl-thialysine mimic was recognized by an anti-H4K16Ac antibody while also being susceptible to NAD^+-dependent deacetylation by the class III Sirt2 deacetylase.[190]

Rather than generating the acetyl-thialysine mimic, a more recent approach which enables the site-specific incorporation of acetyl-lysine through $C(sp^3)$-$C(sp^3)$ bond formation has been published. Here, a cysteine-containing peptide, such as the 10-mer sequence of residues 32-41 of histone H2A, undergoes a visible light-mediated desulfurization reaction with tris(2-carboxyethyl)phosphine (TCEP) and an iridium(III) photocatalyst. The resulting alanyl radical is trapped in the presence of a large excess of N-allylacetamide to afford the peptide containing the acetyl-lysine moiety with the same native stereochemistry at the modified site (**Figure 8c**). However, the yield of this reaction is compromised by a significant by-product consisting of the alanine-containing peptide resulting from trapping of the intermediate alanyl radical through H-atom abstraction. An advantage of this approach is that it enables several lysine PTMs other than

acetylation to be installed, including acyl modifications such as formylation, propionylation, β-hydroxybutylation, succinylation and benzoylation.[191]

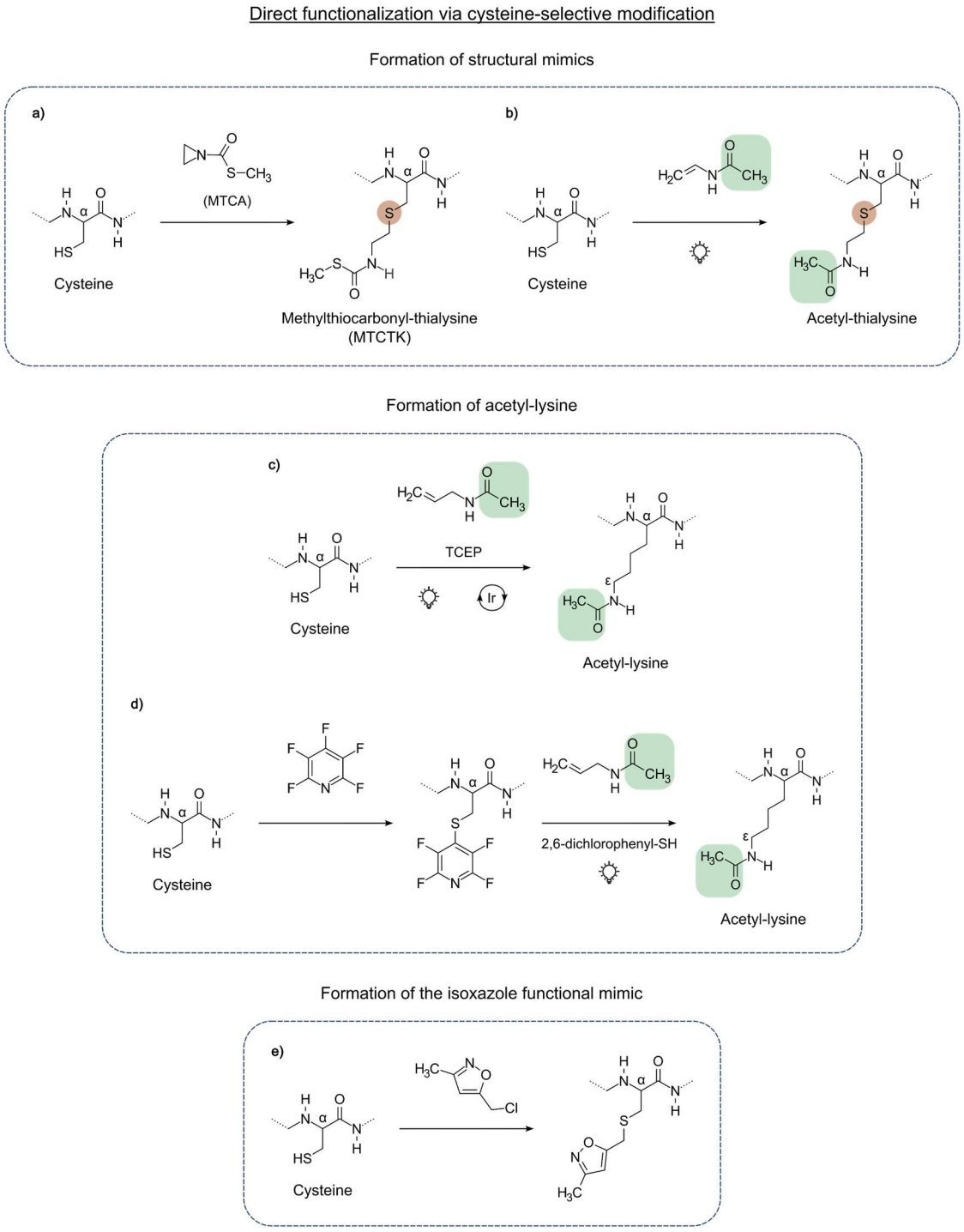


Figure 8. Schematic illustration of the approaches described for the installation of acetyl-lysine and its mimics through direct functionalization via cysteine-selective modification.

A similar radical-mediated desulfurative approach to install acetyl-lysine through carbon-carbon bond formation has recently been published. In this case, site-selective modification of the cysteine residue in histone H3K18C with pentafluoropyridine followed by light-mediated $C_{\beta}–S_{\gamma}$

homolysis and subsequent trapping of the resulting alanyl radical with N-allylacetamide afforded the desired H3 protein acetylated at K18 (**Figure 8d**). Again, the native stereochemistry at the modified site was retained and decreased yields due to the formation of the corresponding desulfurized H3K18A by-product was observed. Importantly, complete deacetylation occurred upon incubation of the acetylated H3K18 protein with Sirt2. In addition to acetylation, this approach also enables other modifications other than naturally-occurring PTMs to be installed.[192]

Another approach, this time focused on installing functional rather than structural mimics of acetyl-lysine in cysteine residues, was also developed. It was based on previous experimental data showing that the 3,5-dimethylisoxazole moiety acts as an acetyl-lysine bioisostere capable of displacing acetylated histone peptides from bromodomains.[193] Expanding this work further, a histone H4 peptide containing the K12C substitution was site-selectively alkylated with structurally distinct isoxazole-containing electrophiles to generate a functional acetyl-lysine mimic with high affinity for BRD4, another member of the BET family of proteins (**Figure 8e**). This isoxazole-derived modification was subsequently introduced into the K18C mutant of histone H3, showing that this approach can be extended to proteins.[194]

Indirect functionalization via intermediate Dha formation

The approaches outlined so far to introduce acetyl-lysine or its mimics have focused only on the direct modification of the cysteine residue. However, an alternative avenue also exists in which the reactive groups of some amino acids are eliminated to generate Dha. The latter is a relatively rare, non-proteinogenic amino acid that is naturally found in some microbial peptides such as lanthipeptides following serine dehydration, as well as in a few human proteins due to spontaneous non-enzymatic elimination of phosphoric acid from phosphoserine as a result of molecular aging.[195,196] Owing to its unsaturated double bond, the incorporation of Dha imposes a conformational rigidity into peptides and proteins.[197] In addition to its biological role, Dha is of particular interest in protein chemistry due to its synthetic usefulness as a precursor to various chemical modifications.

The site-selective incorporation of Dha into proteins can be accomplished by targeting residues other than cysteine. In every case, however, the precursor residues to Dha are either naturally present at the site of interest or placed there by either site-directed mutagenesis or, more infrequently, by genetic code expansion. The side chains of these residues are subsequently modified into good leaving groups, generating Dha following their elimination.[188] Initial attempts at installing Dha in proteins were made in an effort to elucidate the role of the active serine at the catalytic site of chymotrypsin. In this context, the enzyme was treated with tosyl chloride (and later, tosyl fluoride) to promote selective sulfonylation of the nucleophilic serine at the active site. Base-catalyzed elimination of the tosyl group resulted in the conversion of the activated serine to Dha.[198,199] Interestingly, it was already envisioned at the time that the chymotrypsin-installed Dha could be used to introduce “chemical mutations” by nucleophilic addition, thereby generating novel proteins with slightly different chemical properties from which the role of the active serine could be further explored.[199]

While innovative for its time, the latter approach to Dha formation uses conditions that are too stringent for most proteins and is suitable only for activated serine residues, such as those present at catalytic triads, which are remarkably nucleophilic by design.[200] Therefore, methods to install Dha in less common and more reactive residues were sought in order to modify a wider range of proteins and achieve selectivity.[188] A schematic illustration of the most relevant Dha precursors and approaches describing the use of this intermediate to install acetyl-lysine and its mimics is shown in **Figure 9**. Dha formation has been achieved from amino acids such as phosphoserine, selenocysteine and its derivatives, and more importantly, cysteine (**Figure 9a**).

One approach described in the literature generated Dha from phosphoserine, a ncAA whose introduction into recombinant proteins was made possible by genetic code expansion. Following expression and purification of the phosphoserine-containing proteins, conversion to Dha proceeded at room temperature under mild basic conditions using barium hydroxide. This reaction was selective for the non-canonical phosphoserine residue, with no side reactions being observed. The subsequent conjugate addition of alkyl radicals, which were generated *in situ* from alkyl iodides by transition metals to Dha-functionalized proteins resulted in the chemoselective incorporation of several lysine PTMs, including acetylation, via carbon-carbon bond formation (**Figure 9b**). In this way, it was possible to produce the acetylated histone H3 at position K79 with the natural acetyl-lysine PTM.[201]

As will be explained in subsection 1.3.2, the incorporation of ncAAs such as phosphoserine can present some challenges and is experimentally more complex than simply producing recombinant proteins with natural residues. An alternative consists of using selenocysteine, which is commonly known as the twenty-first proteinogenic amino acid. As the sulfur-to-selenium substituted isostere of cysteine, the rare selenocysteine shares many of its chemical properties, while also displaying some important differences.[202] In particular, the higher acidity of selenocysteine, as indicated by the lower pK_a value (~5.5) of its selenol group, enables this residue to be almost completely ionized to a selenolate at physiological pH.[203,204] However, despite the higher acidity of selenocysteine, which translates to a lower basicity of the selenolate anion, the latter is more nucleophilic than thiolate due to the higher polarizability of the selenium atom compared to sulfur.[204] As a result, the ultra-low abundant and highly nucleophilic selenocysteine residue constitutes an excellent target for selective protein modification.

The incorporation of selenocysteine into proteins is more flexible than the introduction of phosphoserine. It can be incorporated co-translationally by exploiting the natural biosynthetic pathway of selenoproteins.[205] In addition, selenocysteine can also be introduced by NCL and EPL.[205,206] Similarly to phosphoserine, selenocysteine derivatives can also be introduced by genetic code expansion. Following the incorporation of selenocysteine or its derivatives, such as phenylselenocysteine or alkylselenocysteine, Dha can be generated from these residues through oxidative elimination using hydrogen peroxide. In this way, by exploiting the reactivity of the two mentioned selenocysteine derivatives, it was possible to install the acetyl-thialysine mimic into histone H3 with site-specificity via Dha formation followed by a Michael addition with N-acetylcysteamine (**Figure 9c**). However, a notable disadvantage of this approach consisted of

methionine oxidation, an undesired side reaction which was circumvented by replacing this residue with the structurally similar amino acid leucine.[207,208]

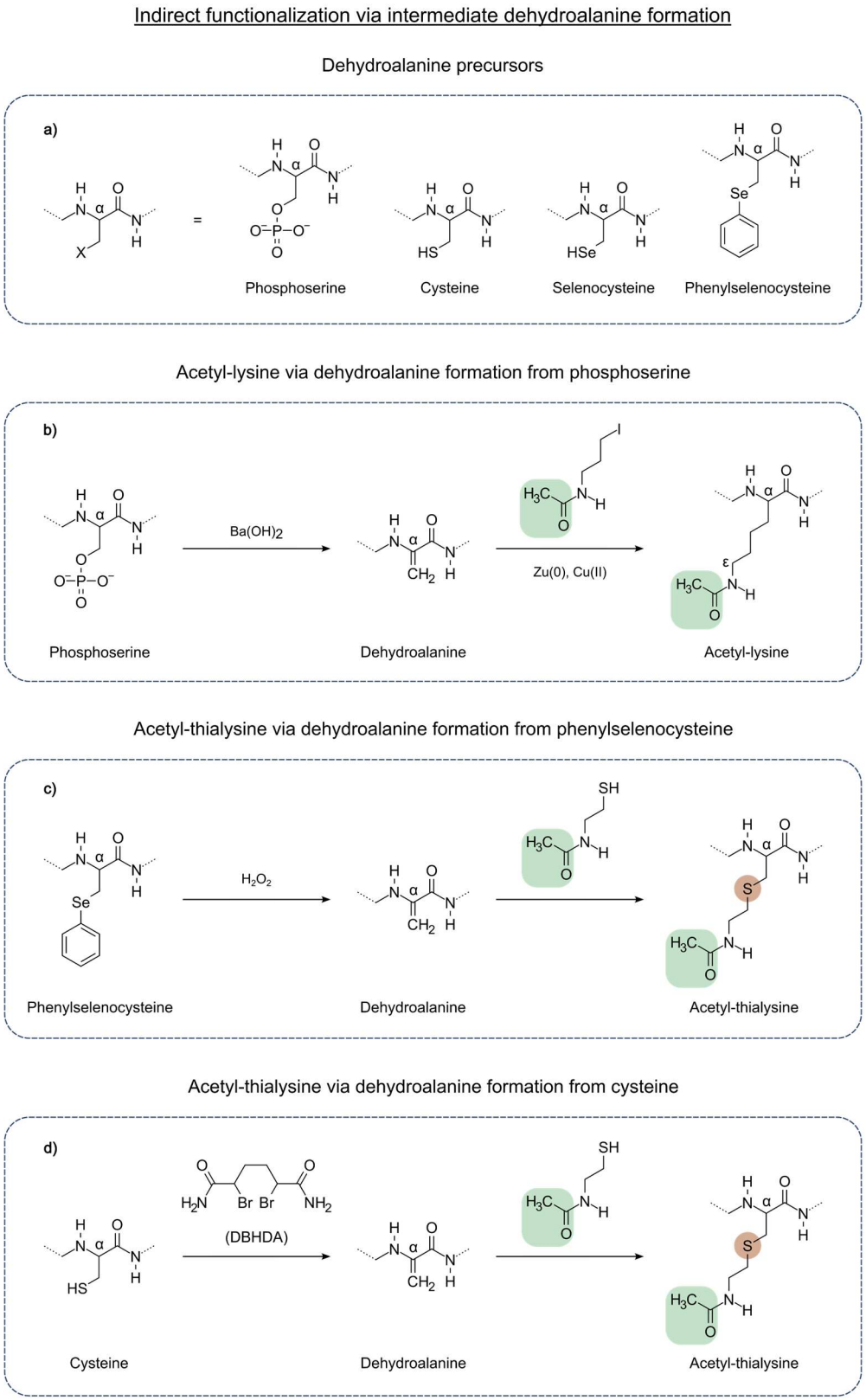


Figure 9. Schematic illustration of the most common dehydroalanine precursors and approaches describing the use of this intermediate for the installation of acetyl-lysine and its mimics.

In contrast to phosphoserine, selenocysteine and its derivatives, cysteine can be easily incorporated into recombinant proteins by site-directed mutagenesis. Several methods have been developed to chemically transform cysteine to Dha using mild conditions that are compatible with protein substrates. The most selective of these rely on the bis-alkylation-elimination of cysteine using reagents containing a central 1,4-dihalobutane structure, with 2,5-dibromohexanediamide (DBHDA) and methyl 2,5-dibromopentanoate being commonly employed for this purpose.[209,210] Importantly, the conversion of cysteine to Dha using the DBHDA reagent was exploited to install the acetyl-thialysine mimic in histone H3. In this context, site-directed mutagenesis was used to insert cysteine at position 9 of H3 from which the Dha residue was site-selectively generated with DBHDA. The addition of N-acetylcysteamine to the H3K9Dha precursor afforded the desired protein product bearing the acetyl-thialysine mimic at position 9, with complete conversion being observed (**Figure 9d**). The generated H3K9 acetyl-lysine mimic was recognized by antibodies raised against the natural modification at this site. Furthermore, a deacetylation of approximately 50% was observed following incubation of the H3K9 acetyl-lysine mimic with HDAC1 and HDAC2. This approach was also used to create a H3 mutant protein containing two acetyl-thialysine mimics at positions 4 and 79.[211]

While the incorporation of acetyl-lysine or its mimics via intermediate Dha formation constitutes a simple and economical chemical strategy for the site-specific introduction of acetylation marks in histones, it also presents a significant drawback. The addition to Dha typically results in a loss of stereochemistry at the α-carbon of the modified residue, thereby producing an epimeric mixture of D and L configuration. The loss of native L-stereochemistry has important implications since deacetylases are often sensitive to D/L-configuration and do not or only poorly process the acetylated D-lysine residue.[211,212] In spite of this isomerism, the resulting acetylated products were nonetheless functionally active in distinct biochemical assays, which demonstrates the usefulness of this approach towards creating histone acetylation marks for future biological studies on chromatin.

1.3.2 Genetic code expansion

Genetic information encoded within the DNA sequence is transcribed into messenger RNA (mRNA), which is subsequently decoded by transfer RNAs (tRNAs). These tRNAs contain a distinct sequence of three nucleotides, known as an anticodon, that specifies which amino acid they should carry. The pairing of a given tRNA with its cognate amino acid is accomplished through the action of aminoacyl-tRNA synthetases. These enzymes catalyze the esterification reaction between an amino acid and the 3'-hydroxyl end of a tRNA, together with the hydrolysis of one ATP molecule.[213] The ribosome, a ribonucleoprotein complex where protein synthesis occurs, contains three tRNA binding sites (A, P and E). At the A-site, the produced aminoacyl-tRNA reads the matching three nucleotide sequence, termed a codon, on the mRNA via complementary base-pairing. Peptide bond formation occurs following nucleophilic attack of the α-amino group of the A-site aminoacyl-tRNA on the carbonyl carbon of the P-site peptidyl-tRNA. This yields an A-site bound peptidyl-tRNA that is one residue longer and a deacylated tRNA in

the P-site.[214] The ensuing translocation of the tRNAs and mRNA through the ribosome displaces the peptidyl-tRNA to the P-site and exposes a new codon in the A-site ready to interact with its respective aminoacyl-tRNA, while the deacylated tRNA binds to the E-site before dissociating from the ribosome.[215]

Nature relies on a limited and conserved set of twenty amino acids to synthesize most proteins. These residues, typically known as canonical amino acids, are genetically encoded by sixty-one codons and are incorporated into proteins by the general components of the endogenous translation machinery. The remaining three codons – UAG (amber), UAA (ochre) and UGA (opal) – are commonly known as stop codons since these do not typically code for any amino acid, instead signaling the end of the translation process.[216] Two additional residues, selenocysteine and pyrrolysine, are also naturally found in proteins. These proteinogenic albeit rare amino acids are incorporated during translation in a context-dependent manner by recoding the UGA (selenocysteine) and UAG (pyrrolysine) stop codons.[217]

Selenocysteine is naturally found in all kingdoms of life as the defining component of selenoproteins.[218] This amino acid lacks a dedicated aminoacyl-tRNA synthetase and is instead synthesized on its cognate tRNA from serine, although the biosynthetic pathway is different in bacteria when compared to archaea and eukaryotes.[219,220] The resulting selenocysteinyl-$tRNA^{Sec}$, delivered to the ribosome by a specialized elongation factor, reassigns the UGA stop codon to selenocysteine in the presence of a mRNA stem-loop structure known as the SECIS (selenocysteine insertion sequence) element.[220,221]

Pyrrolysine, found only in bacterial and archaeal proteins but not eukaryotes, exhibits a distinctive chemical structure composed of a lysine residue attached via its ε-amino group to a methylated pyrroline carboxylate. The imine in the pyrroline ring is a reactive electrophilic moiety that is absent in the repertoire of other proteinogenic amino acids.[222,223] Unlike selenocysteine, pyrrolysine has its own aminoacyl-tRNA synthetase that directly catalyzes the ligation of this amino acid to its cognate tRNA.[224] The incorporation of pyrrolysine in response to the UAG stop codon during translation requires several genetic elements typically encoded in a single operon. Within this operon, pylB, pylC and pylD encode the corresponding enzymes that act sequentially to mediate the biosynthesis of pyrrolysine from two equivalents of lysine. In addition, the gene pylS encodes the pyrrolysyl-tRNA synthetase whereas pylT encodes the respective $tRNA^{Pyl}$.[223]

The standard genetic code was naturally expanded to incorporate selenocysteine and pyrrolysine during protein translation in the ribosome. Inspired by this natural concept of genetic code expansion, researchers have developed approaches that hijack the translational machinery to genetically encode and site-specifically install a wide range of chemical functionalities in proteins during their synthesis in cells. These approaches rely on the introduction of an unassigned codon at the desired modification site within the gene sequence of the protein of interest. In response to this repurposed codon, which is typically the UAG (amber) stop codon, an orthogonal aminoacyl-tRNA synthetase – tRNA pair instructs the ribosomal incorporation of a ncAA into the growing polypeptide chain. The existence of orthogonality is paramount to restrict the possibility of cross-reactivity with the endogenous aminoacyl-tRNA synthetases, tRNAs or

amino acids of the host. This ensures that the orthogonal synthetase only aminoacylates the orthogonal cognate tRNA with the desired ncAA and that the orthogonal cognate tRNA is not recognized by the endogenous synthetases.[225] A schematic illustration of the genetic code expansion strategy using the amber stop codon to incorporate a ncAA into proteins is shown in **Figure 10**.

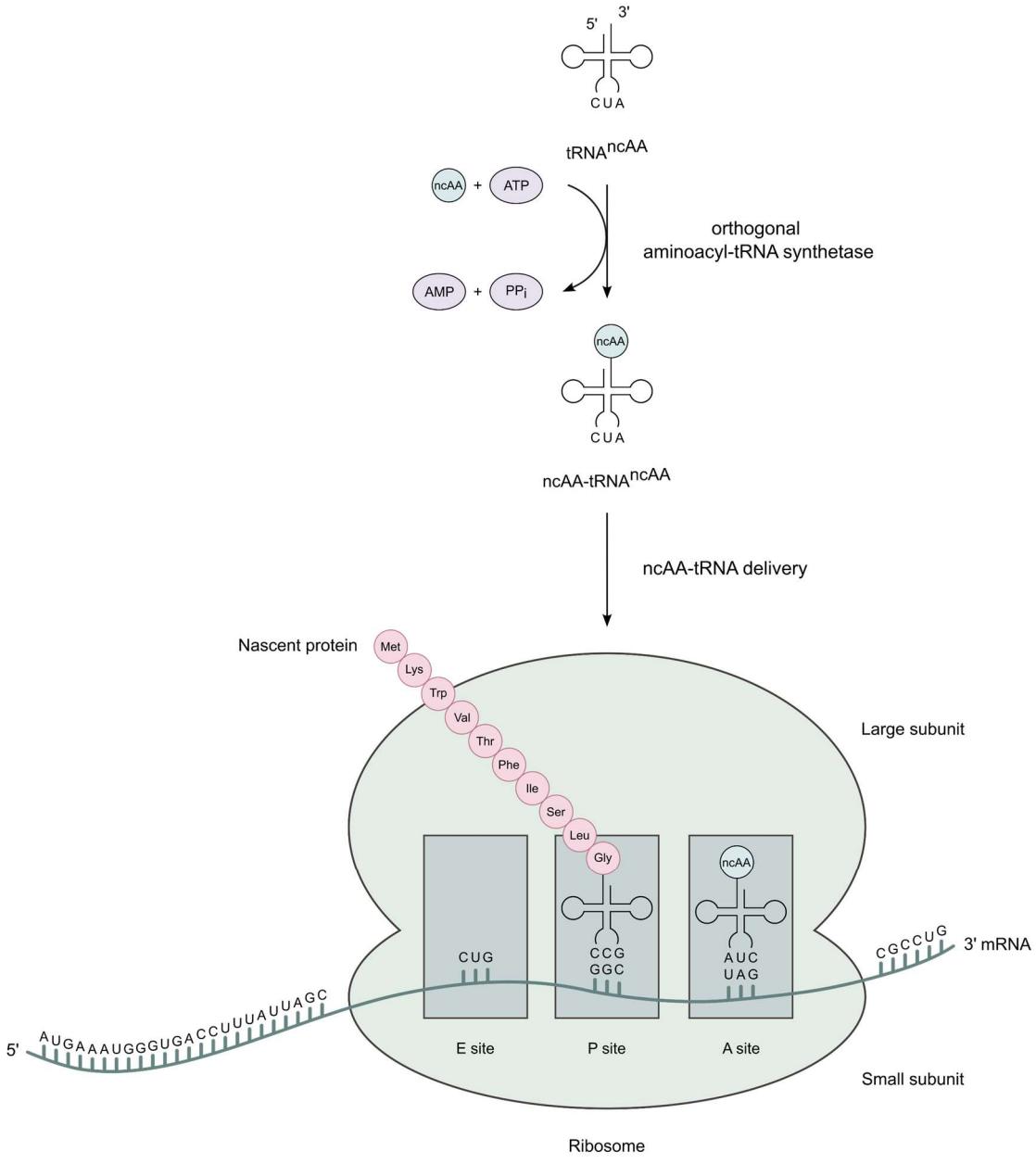


Figure 10. Schematic illustration of the genetic code expansion strategy for the incorporation of a ncAA into proteins. This figure was adapted from [226,227,228].

For an aminoacyl-tRNA synthetase to be orthogonal to the translational machinery of the host, its active site must exclusively recognize the ncAA and transfer it to its cognate orthogonal tRNA.[225] In this context, the archaeal pyrrolysyl-tRNA synthetase is a particularly attractive choice due to being orthogonal in commonly used bacterial and eukaryotic expression systems, and binding exclusively to pyrrolysine but none of the canonical amino acids. As a result, mutation

of its active site is not required to abolish canonical amino acid binding. Although its range of substrates can be expanded by directed evolution, the wild-type pyrrolysyl-tRNA synthetase can incorporate a substantial number of ncAAs due to its high substrate side chain promiscuity, which allows the pyrroline ring to be replaced with a related substituent. This feature enables the synthetase to be used directly in functional studies without the need to reengineer its active site.[226,227] The pyrrolysyl-tRNA synthetase also displays other distinctive characteristics that facilitate the engineering of the pyrrolysine incorporation machinery. Its low selectivity for the α-amine substrate enables α-hydroxy acids to be incorporated in response to amber codons, whereas its low selectivity toward the anticodon of $tRNA^{Pyl}$ allows other codons, such as the remaining stop codons, to be reassigned to ncAAs.[226] Thus, the archaeal pyrrolysyl-tRNA synthetase constitutes an invaluable starting point to develop orthogonal aminoacyl-tRNA synthetase – tRNA pairs.[229]

By enabling the co-translational and site-specific incorporation of ncAAs in the protein of interest within a living organism, genetic code expansion constitutes a strategy of considerable significance to investigate the biological role of PTMs such as acetylation. These diverse chemistries can be installed by direct incorporation of the intended amino acid or indirectly through the introduction of a residue that bears a bioorthogonal handle which is later modified to afford the desired chemical structure. Amino acids with smaller PTMs are more easily accepted by their cognate aminoacyl-tRNA synthetase than those bearing larger modifications. As a result, genetic code expansion approaches using the direct route are best suited to install small PTMs, whereas access to proteins bearing larger and more complex modifications typically rely on the indirect incorporation of ncAAs bearing bioorthogonal handles.[229] Although acetylation is a comparatively small PTM, approaches to install acetyl-lysine via direct and indirect incorporation of this modified amino acid have been described. A schematic illustration of two published approaches to incorporate acetyl-lysine into proteins directly and indirectly through genetic code expansion, which will be described below, is shown in **Figure 11**.

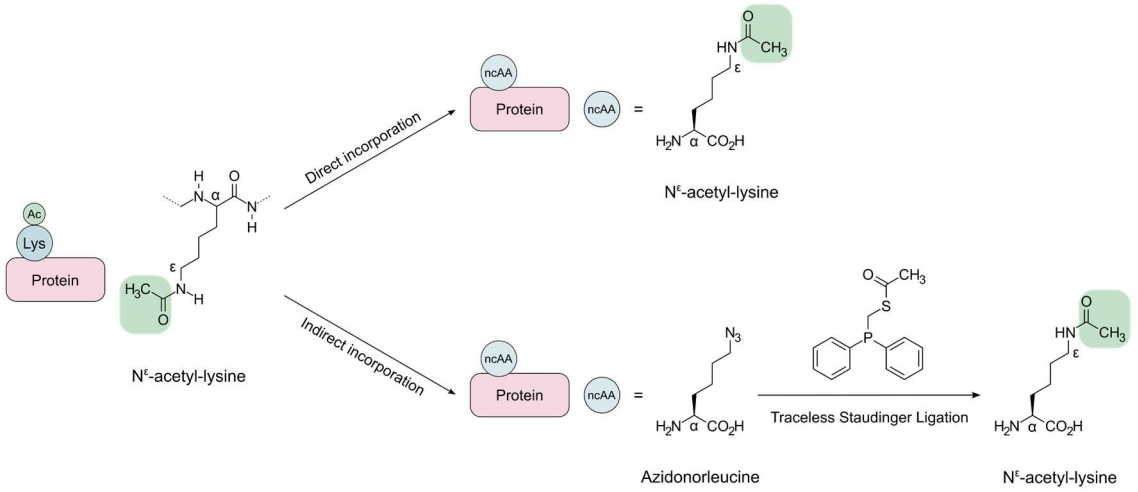


Figure 11. Schematic illustration of two published approaches to produce acetylated proteins through genetic code expansion. In the direct approach, the pyrrolysyl-tRNA synthetase – tRNA pair was engineered to accommodate the ncAA acetyl-lysine. In the indirect approach, a ncAA bearing an azide bioorthogonal handle (azidonorleucine) was genetically incorporated instead, which afforded the desired acetyl-lysine residue upon traceless Staudinger ligation with an acetyl-phosphinothioester.

Direct incorporation of acetyl-lysine

The first genetic code expansion approach described for the site-specific installation of acetyl-lysine in proteins exploited the pyrrolysine translational machinery of *Methanosarcina barkeri* (*M. barkeri).* The pyrrolysyl-tRNA synthetase – tRNA pair of this methanogenic archaeon instructs the incorporation of pyrrolysine in response to UAG codons in various methyltransferase genes. While acetyl-lysine can be considered a substructure of pyrrolysine, it is not, however, a substrate for pyrrolysyl-tRNA synthetase. Thus, the pyrrolysyl-tRNA synthetase – $tRNA^{Pyl}$ pair, which is orthogonal in the *E. coli* expression system, was engineered by directed evolution to accommodate the non-canonical acetyl-lysine amino acid. In addition to the desired protein, in this case myoglobin, containing a single acetylation at the genetically encoded location, the version devoid of this modification was also produced. Since *E. coli* contains a single known endogenous deacetylase, the NAD^+-dependent CobB sirtuin, it was hypothesized that the inhibition of this enzyme with nicotinamide (NAM) would only lead to the expression of the acetylated protein. Subsequent recombinant expression in the presence of NAM abolished the post-translational deacetylation of the ribosomally incorporated acetyl-lysine in myoglobin.[230]

Importantly, the authors envisioned the potential application of their approach to study the biological role of acetylation in chromatin structure and function and were already, at the time of publication, attempting to produce nucleosomes with histones bearing genetically encoded acetyl-lysine residues at natural modification sites. Their efforts subsequently led to the development of an improved approach to produce recombinant histones in *E. coli* with genetically encoded acetylation at defined sites. As a proof-of-concept, position K56 of histone H3 was chosen as the initial target of their approach due to the biological significance of this modification and its location in the middle of the protein sequence. However, to establish the generality of their approach, other histones and acetylation sites were also produced and these include H3K14Ac, H3K23Ac, H3K27Ac, H2BK5Ac, H2BK20Ac, and H2AK9Ac.[231]

Their previously developed acetyl-lysyl-tRNA synthetase – tRNA pair derived from the pyrrolysyl-tRNA synthetase – $tRNA^{Pyl}$ pair of *M. barkeri* was used as a starting point to increase the efficiency of the system and produce milligram quantities of homogeneously acetylated histones. Furthermore, the H3K56Ac protein was subsequently used to investigate the role of the acetylation event at K56 of H3 on chromatin structure and function. To achieve this, H3 and H3K56Ac were separately assembled with histones H2A, H2B and H4 to form octamers, which were mixed with DNA to reconstitute nucleosomes. Similar octamer assembly efficiencies were observed for both types of nucleosomes, revealing that acetylation at K56 of H3 does not influence the *in vitro* formation of these structures.[231]

The authors performed Förster resonance energy transfer (FRET) experiments to study the possible influence of H3K56 acetylation on salt-dependent nucleosome stability. FRET is a powerful experimental technique to study nucleosome stability and dynamics because it provides reliable information regarding molecular proximity at nanometer scale. It involves the transfer of non-radiative energy by dipole-dipole coupling from an excited state donor fluorophore to a ground state acceptor molecule. Two main requirements must be satisfied for energy transfer to

occur, namely overlap between the emission and absorption spectra of the donor and acceptor, respectively, and nanoscale proximity of 1-10 nm between the donor and the acceptor (although some unusual FRET pairs can accommodate up to a 20-nm distance).[232]

In the context of the current approach being described for the genetic introduction of H3K56 acetylation, a Cy3 FRET donor was introduced on the 5'-end of a nucleosome-positioning DNA sequence, whereas a Cy5 FRET acceptor bearing a maleimide moiety was conjugated to the K119C mutant of H2A. While each nucleosome contains two copies of H2A labelled with a Cy5 molecule, only one of these is sufficiently proximal to the Cy3 FRET donor on DNA. The salt-induced dissociation of labelled nucleosomes assembled with either H3 or H3K56Ac was monitored by changes in the Cy3 donor and Cy5 acceptor emissions at 565 nm and 670 nm, respectively. At low salt concentrations in which the DNA is still wrapped around the histone octamer, excitation of the Cy3 donor resulted in its reduced emission at 565 nm and increased Cy5 acceptor emission at 670 nm, which agrees with a distance-dependent FRET signal. As the salt concentration increased until 1.75 M, covering intramolecular conformational changes from partial unwrapping of the nucleosomal DNA to dissociation of H2A/H2B and H3/H4 dimers, acetylated and unacetylated nucleosomes at H3K56 displayed similar stabilities to salt.[231]

The effect of H3K56 acetylation on the spontaneous breathing of DNA was investigated as well by combining native polyacrylamide gel electrophoresis (PAGE), alternating-laser excitation selection and single-pair FRET. To this end, nucleosomes were assembled with two distinct label pairs on their positioning DNA and separated from free DNA by native PAGE. The end label pair was positioned 2 bp from the end of the DNA and within 7 bp of the K56 site. The internal label pair was placed 27 bp from the end of the nucleosomal DNA and 20 bp internally to the K56 site. Following excision and confocal imaging of the corresponding gel bands using alternate bursts of green and red excitation, the FRET efficiency was determined. These experiments suggested that K56 acetylation of H3 resulted in a seven-fold increase in DNA breathing.[231]

The authors also explored whether acetylation at H3K56 influences chromatin compaction, since previous studies had shown that this modification plays an important role in telomere silencing. To accomplish this, nucleosome arrays reconstituted from histone octamers containing either H3 or H3K56Ac on 61 repeats of a nucleosome-positioning DNA sequence in the presence of H5 linker histone were subjected to sedimentation velocity analysis. The results from these experiments revealed that acetylation at H3K56 does not significantly alter chromatin compaction and thus the effects described in other studies are either mediated by extra factors that are present *in vivo* or dependent on the cellular context.[231]

Finally, the genetically incorporated acetylation at K56 of H3 also enabled the authors to evaluate the effect of this PTM on chromatin remodeling. The SWI/SNF remodeling complex contains a bromodomain and previous studies had shown an association between its recruitment to active genes and H3K56 acetylation.[233] In this context, the effect of H3K56 acetylation on the direct binding of the SWI/SNF remodeling complex was studied using the electrophoretic mobility shift assay (EMSA).[231] EMSA is a rapid and sensitive technique to investigate the DNA-binding properties of a protein. It is based on the observation that protein complexes with nucleic acids

typically migrate more slowly through a gel than the corresponding free nucleic acids when subjected to electrophoresis.[234] Binding differences to SWI/SNF were not observed between nucleosomes containing H3 or H3K56Ac using EMSA, which suggested that recruitment of the chromatin remodeling complex following acetylation at this site is mediated by additional factors or is context-dependent.[231] Another alternative explanation is that H3K56 acetylation influences SWI/SNF-mediated chromatin remodeling by facilitating access to DNA at the entry-exit site.[231,233] To explore this possibility, the authors performed mononucleosome repositioning assays in the presence of SWI/SNF and RSC (Remodeling the Structure of Chromatin), two remodelers of the SWI/SNF family. Both chromatin remodeling complexes repositioned the H3K56 acetylated nucleosomes slightly faster than the unmodified nucleosomes.

In addition to repositioning nucleosomes along DNA *in cis*, chromatin remodelers can also drive the exchange of H2A/H2B dimers from chromatin fragments.[235] Thus, the authors also evaluated whether H3K56 acetylation has an effect on this type of chromatin remodeling outcome using H2A/H2B dimer transfer assays. In this context, donor nucleosomes containing H3 or H3K56Ac were assembled with Cy5-labelled H2A onto Cy3-labelled DNA sequences. A wild-type H3/H4 tetramer assembled onto DNA fragments served as the H2A/H2B dimer acceptor. Following incubation with SWI/SNF and RSC chromatin remodelers, samples were resolved by native PAGE and the Cy5-labelled histone dimer fluorescence was evaluated. Nucleosomes assembled with H3 or H3K56Ac exhibited similar H2A/H2B dimer transfer behavior promoted by RSC and SWI/SNF, suggesting that K56 acetylation of H3 does not influence this type of remodeling process.[231] As perfectly illustrated by this pioneering work in which acetylation at H3K56 was genetically introduced, the ability to produce histones bearing acetyl groups at specific sites enables a variety of biochemical studies from which the putative functional role of this PTM can subsequently be derived.

Indirect incorporation of acetyl-lysine

Although acetyl-lysine is a relatively small PTM and can thus be directly incorporated into proteins, it can also be introduced by coupling genetic code expansion with bioorthogonal chemistry strategies. In this context, a ncAA bearing a reactive handle is genetically encoded at the desired site and is subsequently reacted with the appropriate counterpart to generate an acetyl-lysine at the site of interest. An example of this indirect route consisted of the genetic incorporation of azidonorleucine followed by a traceless Staudinger ligation reaction. With its azide group, azidonorleucine contains a bioorthogonal handle that can react with a phosphinothioester to produce an amide bond. One advantage of this approach is its versatility, since it enables the installation of a variety of acylations, including acetylation, purely by modifying the acyl group within the phosphinothioester reagent, without the need to engineer a pyrrolysyl-tRNA synthetase – tRNA pair that recognizes a particular acyl-lysine. Indeed, encoding azidonorleucine into ubiquitin and histone H3 via genetic code expansion and reacting these with acetyl-phosphinothioester or succinyl-phosphinothioester, enabled the site-specific acetylation and succinylation of ubiquitin at K48, as well as the site-specific succinylation of histone H3 at K4.

However, the existence of a few noteworthy drawbacks limits the usefulness of this approach. First, the traceless Staudinger ligation displays significantly slow kinetics, which led to acylation reaction times of approximately 48 hours to reach completion. Secondly, the intrinsic reductive nature of the phosphinothioester reagent inevitably leads to the partial conversion of azidonorleucine to lysine, which results in the formation of the non-acylated protein as a side product.[236]

Advantages, disadvantages and challenges of genetic code expansion

Genetic code expansion constitutes a powerful synthetic biology strategy for obtaining proteins with precisely defined PTMs, including acetylated histones. It presents the advantage of enabling the synthesis of modified histones within a biological context, thereby exploiting the endogenous translation and folding machineries. The possibility of incorporating bioorthogonal handles also endows this strategy with the ability to probe the dynamics and function of acetylated histones in a cellular context, from its ribosomal synthesis until its ultimate delivery in the nucleus. This is particularly relevant to decode the spatiotemporal role of histone acetylation within living cells, thereby contributing to the growing mechanistic understanding of this PTM. In addition, any site can in principle be targeted to incorporate acetyl-lysine or its mimic in histone proteins, requiring conventional molecular biology techniques that are typically in use in many laboratories.[229]

While genetic code expansion constitutes an indispensable strategy to study the biological role of any PTM, it also presents some important disadvantages that limit the obtained yields of modified protein. First, it requires the development of an orthogonal aminoacyl-tRNA synthetase – tRNA pair for each ncAA. This can be especially problematic for bulky PTMs, as the orthogonal aminoacyl-tRNA synthetase will likely require extensive engineering to accommodate the stereochemical diversity of such non-canonical substrates in its binding pocket.[229] Further optimization may also be necessary for expression of proteins containing ncAAs in different organisms since the translational machinery between prokaryotes and eukaryotes is not well conserved.[237] Importantly, orthogonality of the aminoacyl-tRNA synthetase – tRNA pair is seldom absolute, with the suppressor tRNA becoming acylated with other amino acids.[238] Another aspect to consider when using genetic code expansion is the ability of the ncAA to cross the plasma membrane and remain metabolically stable within cells in sufficient quantities to sustain recombinant protein expression. In addition, the basal activity of endogenous enzymes that catalyze the removal of PTMs, such as deacetylases, poses an extra constraint to the genetic incorporation of certain modified amino acids in proteins.[229] In these cases, enzymatic inhibition may be required as described above for acetylation by using NAM to abolish the activity of the NAD^+-dependent CobB sirtuin in *E. coli*.

A major challenge in the genetic code expansion strategy concerns the introduction of multiple PTMs. When these are of a different type, the difficulty arises due to the restricted number of yet unassigned codons and the availability of orthogonal aminoacyl-tRNA synthetase – tRNA pairs. Early attempts to overcome this limitation centered on the use of quadruplet rather than

triplet codons, which results in a +1 frameshift.[239,240] In one of these instances, an orthogonal ribosome capable of decoding quadruplet codons and the amber codon was also synthetically evolved.[240] More recently, synthetic nucleotides that can form base pairs via interactions other than hydrogen bonding to store genetic information and translate it into mRNAs containing unnatural codons have also been employed.[241,242] Another avenue to expand the repertoire of ncAAs that can be incorporated in a single protein consists of reducing the existing codon redundancy. In this context, a synthetic variant of *E. coli* was developed in which the two UCG and UCA serine codons, and the UAG stop codon were substituted genome-wide by their synonyms AGC, AGU and UAA, respectively.[243]

Competition between the orthogonal tRNA and the release factors RF1 and RF2 in *E. coli* constitutes an additional barrier to the introduction of more than one PTM in a single protein, even of the same type.[229] These release factors recognize stop codons to terminate protein synthesis, with RF1 binding to UAA and UAG and RF2 binding to UAA and UGA.[244] Efforts at bypassing this competition involved replacing all known occurrences of the UAG stop codon in *E. coli* with synonymous UAA codons. This enabled the deletion of RF1 and reassignment of the UAG stop codon.[245] Another approach consisted of inhibiting RF1 with antimicrobial peptides to improve the incorporation of ncAAs at various sites and enhance genetic code expansion strategies.[246]

1.3.3 Protein synthesis by native chemical ligation and expressed protein ligation

Another robust and widely employed strategy to generate proteins with precisely defined PTMs consists of their chemical synthesis. In this context, protein fragments containing the desired modification(s) are generated and ligated through amide bond-forming reactions. First described in 1994,[247] NCL is the most widely used method to assemble proteins from their cognate segments, as it enables chemoselective peptide bond formation from accessible building blocks under aqueous conditions and physiological pH. The first step of NCL consists of a transthioesterification between a fragment bearing a C-terminal α-thioester with another containing an N-terminal 1,2-amino thiol functionality, which is typically a cysteine residue.[172] This reaction results in the formation of a thioester intermediate that undergoes a spontaneous and irreversible intramolecular S-to-N acyl transfer to yield a native peptide bond adjacent to cysteine at the ligation site.[247]

Chemoselective protein desulfurization was largely developed in the context of NCL to convert the cysteine at the ligation site into an alanine residue, thereby enabling the synthesis of wild-type peptides and proteins without a native cysteine or with one such residue located at an unsuitable location. Additionally, in cases where a synthetically derived thiol-containing amino acid is used rather than cysteine for NCL, desulfurization is employed to remove the sulfhydryl group and thus provide the target protein with the canonical amino acid at the ligation site.[248] A schematic illustration of NCL followed by desulfurization is shown in **Figure 12**.

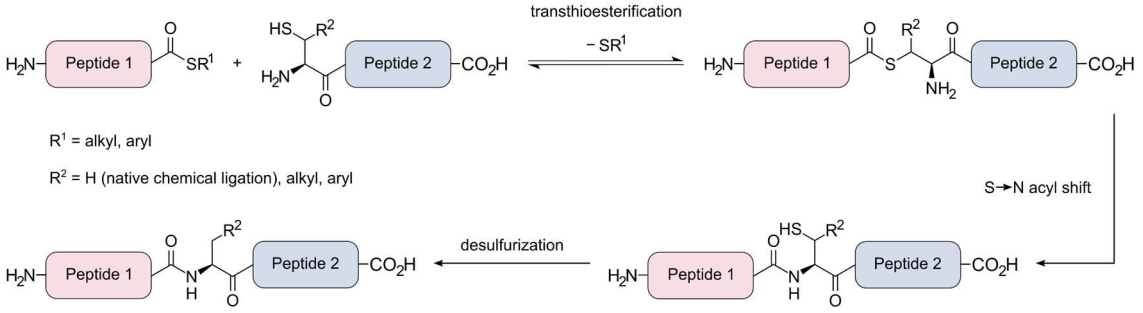


Figure 12. Schematic illustration of NCL followed by desulfurization to obtain the native amino acid at the target peptide or protein. This figure was adapted from [249, 250].

The required segment with an N-terminal 1,2-amino thiol can be produced by recombinant expression of protein fragments bearing an N-terminal cysteine. In addition, peptide segments containing this functionality may also be generated by solid-phase peptide synthesis (SPPS) through a coupling reaction with cysteine or a synthetically derived thiol-containing amino acid, such as glutamine,[249] aspartate,[250] phenylalanine,[251] valine,[252] lysine,[253] threonine,[254] leucine,[255] and arginine.[256] SPPS is the technique of choice for the chemical synthesis of peptides, including those bearing the desired PTM(s). First introduced by Robert Bruce Merrifield in 1963,[257] it consists of sequential cycles of alternate N-terminal deprotection and N-protected amino acid coupling reactions to a growing polypeptide chain immobilized on a solid support. Briefly, the N-protected C-terminal amino acid is first anchored, typically via its C-terminal carboxyl group, to a resin derivatized with a cleavable linker.[258] The functional groups of the residue side chains must also be protected with chemical groups that are resistant to the reaction conditions employed during peptide elongation. Following attachment of the first residue, its N^{α}-amine is deprotected to enable the subsequent coupling to the activated carboxyl group of the second amino acid in the desired sequence. This succession of deprotection and coupling reactions proceeds until the peptide is completely assembled in a linear C-to-N direction. At this end stage, the peptide is released from the solid support and the side chain protecting groups are cleaved.[259]

Besides providing a reversible point of attachment between the peptide chain and the solid support, the linker also protects the C-terminal carboxyl group of the peptide from unwanted side reactions during synthesis. The selection of a suitable linker is important, since its structure determines the chemical conditions that can be used during peptide synthesis, including those for repeated N^{α}-deprotection, cleavage from the solid support and C-terminal functionality of the final product.[260] In this context, various linkers have been described to generate the required C-terminal α-thioester for NCL. Importantly, the choice of N^{α}-protecting group also defines the type of linker that can be employed in SPPS. The two most common N^{α}-protecting schemes use either *tert*-butyloxycarbonyl (Boc) or 9-fluorenylmethyloxycarbonyl (Fmoc) chemical groups, with their removal requiring very different conditions. The Boc strategy relies on the use of trifluoroacetic acid (TFA) to successively remove this protecting group and often employs anhydrous hydrofluoric acid (HF), a highly corrosive and toxic chemical, to deprotect the side chains of the amino acid residues and release the peptide from the solid support. In contrast, the Fmoc strategy

uses mild conditions with secondary amines to remove this protecting group, with TFA being often employed to deprotect the side chains of the amino acid residues and detach the final peptide from the resin. Since it avoids the use of the hazardous HF and displays opposite acid-base requirements for peptide release and protecting group removal, respectively, the Fmoc strategy is usually preferred over the Boc one.[258,259]

Peptide fragments containing the required C-terminal α-thioester for ligation reactions may be generated by SPPS or EPL.[261] Originally reported in 1998,[262] EPL is the semi-synthetic version of NCL in which the segment bearing the α-thioester moiety is recombinantly generated following cleavage of an intein domain that is fused to the C-terminus of the peptide of interest.[261] An intein (intervening protein) is a unique domain that can excise itself from its flanking extein (external protein) sequences. This process, known as protein splicing, involves a sequence of acyl-transfer reactions that cleave the two peptide bonds at the intein-extein junction and ligate the N- and C-exteins. Intein-mediated protein splicing occurs spontaneously without the need for any additional factor or energy source.[263,264]

The canonical mechanism for protein splicing begins by activation of the peptide bond between the N-extein and the N-terminus of the intein. This is achieved by the nucleophilic attack of the N-terminal serine or cysteine of the intein (Block A) on the carbonyl carbon of the immediately upstream (-1) C-terminal N-extein residue. While spontaneous N-to-O/S acyl shifts are thermodynamically unfavored, the presence of threonine and histidine in the conserved TXXH motif (Block B), together with an aspartate residue in the downstream Block F facilitates the formation of a linear (thio)ester intermediate. The next step consists of a nucleophilic attack of the first C-extein residue (+1), which is either cysteine, serine or threonine, on the (thio)ester linkage between the N-extein and the intein. This trans-(thio)esterification reaction results in the formation of a branched intermediate, which is subsequently resolved through cyclization of the C-terminal asparagine of the intein following nucleophilic attack of its β-amide nitrogen on the carbonyl carbon of the peptide bond. Once formation of the succinimide occurs, the intein is excised from the branched intermediate containing the two exteins attached via a (thio)ester linkage. The latter structure subsequently undergoes a thermodynamically favored O/S-to-N acyl transfer to form a stable peptide bond, whereas the C-terminal succinimide of the excised intein slowly hydrolyzes into an α- or β-isomer of the carboxylic acid.[263,264] By mutating the intein to inhibit the formation or breakdown of the branched intermediate and performing the splicing in the presence of a capturing thiol, the target peptide bearing the required C-terminal α-thioester for ligation reactions can be generated.[264,265]

Whether the modified protein is produced by semi- or total chemical synthesis and ligated by NCL or EPL, the peptide fragment(s) containing the desired PTM(s) must be synthesized, typically by SPPS. Semi-synthesis involves the chemoselective ligation between recombinantly generated unmodified fragments and synthesized peptides bearing the PTM(s) of interest. In total synthesis, the chemoselective ligation reaction occurs exclusively between synthetized peptide fragments.[172] A schematic illustration of protein semi-synthesis and total chemical synthesis is shown in **Figure 13**.

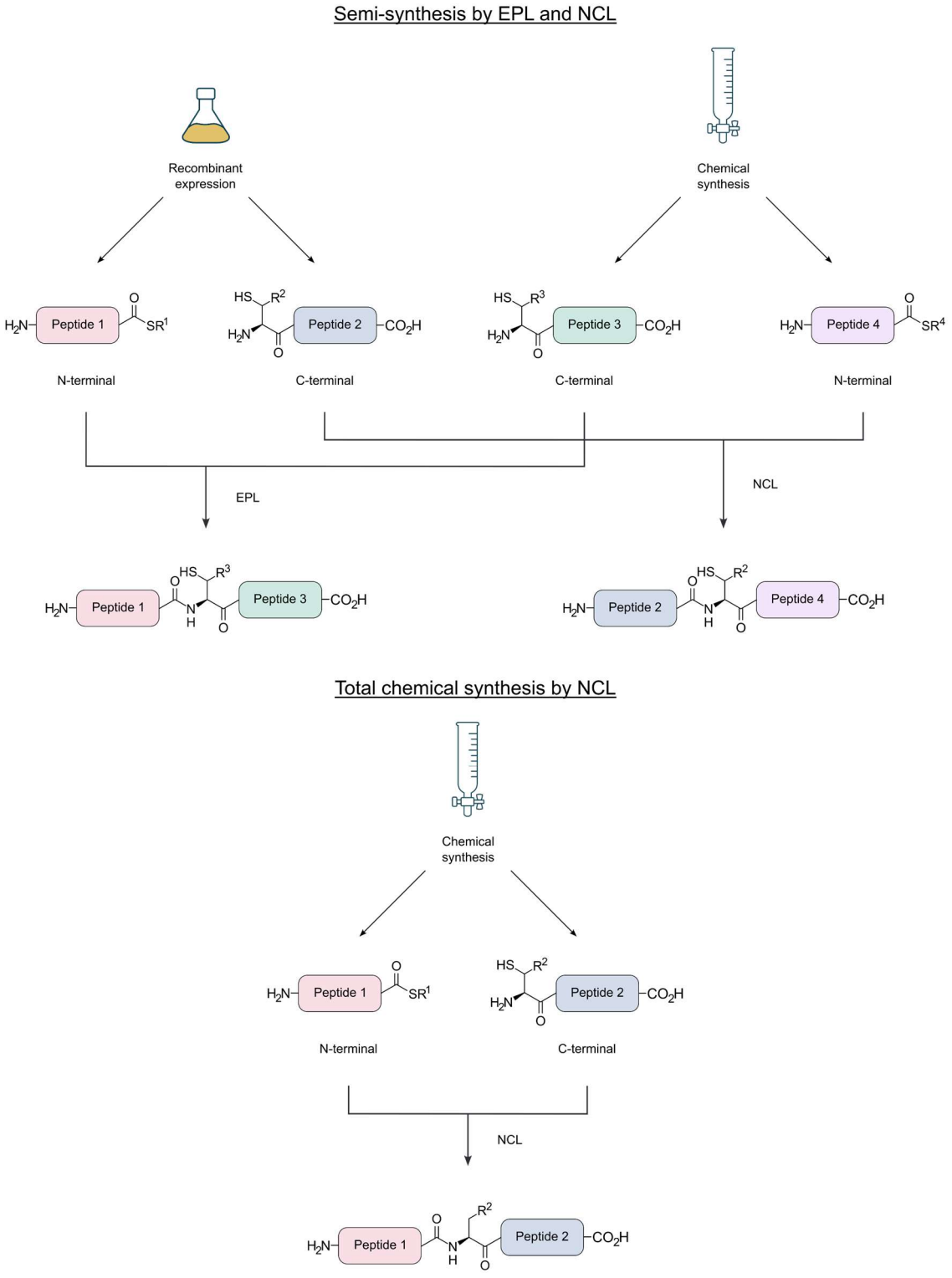


Figure 13. Schematic illustration of semi-synthesis and total chemical synthesis of proteins. In the case of semi-synthesis by EPL, the peptide containing the C-terminal thioester is recombinantly expressed whereas the peptide bearing the N-terminal thiolated residue is synthesized. In contrast, in the case of semi-synthesis by NCL, the peptide containing the C-terminal thioester is synthesized whereas the peptide bearing the N-terminal thiolated residue is recombinantly expressed. In the case of total chemical synthesis, both the C-terminal thioester and N-terminal thiolated peptide fragments are synthesized. This figure was adapted from [172].

Semi-synthesis of acetylated histones

Several semisynthetic routes to generate acetylated histones have been reported in the literature. The first traceless semi-synthesis of acetylated histones consisted of installing several acetylation marks in H3 and H4 through a combination of NCL and desulfurization. In this context, N-terminal H3-(1-24) and H4-(1-14)-derived peptide thioesters with acetyl-lysine residues were synthesized by Boc-SPPS using a 3-mercaptopropionic acid linker on a methylbenzhydrylamine resin. These peptides were then reacted with truncated H3A25C-(25-135) and H4A15C-(15-102) C-terminal domains containing N-terminal cysteine residues to produce the corresponding multi-acetylated histones by NCL. H3 contained five (K4, K9, K14, K18, K23) acetyl-lysines, whereas H4 was prepared with two (K5, K8) or three (K5, K8, K12) of such residues. These products were converted to their corresponding native sequences following cysteine desulfurization.[266]

The penta-acetylated semi-synthetic H3 was then assembled into a heterotetramer with recombinant H4. To this end, equimolar concentrations of H3 (K4Ac, K9Ac, K14Ac, K18Ac, K23Ac) and H4 were mixed under denaturing conditions and dialyzed to remove the denaturant, thereby promoting refolding and self-assembly. The semi-synthetic H3/H4 heterotetramer, which was purified by size-exclusion chromatography and exhibited a similar elution profile to the recombinant one, was subsequently incorporated into chromatin through the action of the ISWI-containing factor RSF (Remodelling and Spacing Factor).[266] The latter mediates ATP-dependent nucleosome deposition in the absence of chaperones.[267] In the presence of ATP and a DNA template, RSF was able to assemble histones H2A and H2B with the semi-synthetic H3/H4 heterotetramer to form nucleosome arrays.[266]

Finally, the ability of heterotetramers containing semi-synthetic H3 (K4Ac, K9Ac, K14Ac, K18Ac, K23Ac) and recombinant H4 to serve as substrates for histone-modifying enzymes was also evaluated. In this context, the human histone methyltransferase G9a, which catalyzes most mono- and di-methylation events at position K9 of H3, was employed.[268] Since the K9 site of penta-acetylated H3 is already blocked by an acetyl group, incubation of the semi-synthetic H3/H4 heterotetramers with G9a did not result in a transfer of [^{3}H]-methyl groups from its [^{3}H]-SAM cofactor. However, treatment with the NAD^+-dependent Sirt1 prior to incubation with G9a and [^{3}H]-SAM enabled deacetylation at K9 to occur first, which was then followed by methyl transfer to form some methylated product.[266] Thus, acetylated H3 proteins generated by traceless semi-synthesis via NCL and desulfurization can be used to study histone modification enzymes. A second semi-synthetic approach was described, in this case to generate tetra-acetylated H3 and study the mechanism of chromatin remodeling enzymes. Here, four acetyl groups at positions K9, K14, K18 and K23 were present in the N-terminal H3-(1-27)-derived thioester peptide, while the H3S28C C-terminal domain truncated from residues 1-27 and containing an N-terminal cysteine residue was recombinantly expressed.[269]

Another semi-synthetic approach, this time to install C-terminal acetylation marks in H3 at positions K115 and/or K122, was developed. To this end, the truncated H3-(1-109) sequence was fused to the GyrA intein and recombinantly expressed, with the corresponding N-terminal thioester generated following thiolysis via addition of sodium 2-mercaptoethanesulfonate. The C-

terminal H3-(110-135)-derived peptides bearing an N-terminal cysteine were synthesized by Boc-SPPS. Site-specific acetylation of these peptides was achieved prior to their release from the resin by incorporating the lysine(s) of interest with orthogonal Fmoc protecting group(s), which were deprotected under basic conditions with piperidine, and treating the resulting side chain amine(s) with acetic anhydride. In this manner, the semi-synthesis of acetylated H3 at K115 and/or K122 was performed by EPL and these proteins were individually assembled into nucleosomes. Acetylation of these lysine residues, which are located at the nucleosome dyad, weakened the interactions between DNA and the histone octamer.[270]

Total chemical synthesis of acetylated histones

The semi-synthetic approaches described above are particularly useful to prepare proteins with terminal PTMs, as is the case of terminally acetylated histones. In contrast, total chemical synthesis enables the modification of proteins at any desired site in the sequence. This approach begins by choosing a suitable ligation site in the sequence, which is followed by the synthesis of the required peptide fragments through SPPS and finally their assembly via NCL.[172] Curiously, the first described approach relying on total chemical synthesis to acetylate H3 targeted the K56 site. The Boc-SPPS strategy was employed to prepare the required peptide fragments, with the C-terminal thioester moiety being synthesized via use of a mercaptopropionamide linker on a 4-methylbenzhydrylamine resin. The acetylated lysine was introduced as the commercially available N^{α}-t-Boc-acetyl-lysine and the protected N-terminal cysteine was incorporated as the sulfur-containing derivative of proline (Boc-L-thiazolidine-4-carboxylic acid) to prevent peptide cyclization or polymerization. The preparation of acetylated H3 at K56 from the synthesized peptide fragments was performed by NCL alone or in combination with desulfurization.[271]

In the first case, two cysteines were introduced and retained in the *X. laevis* H3 (C110A) sequence at positions R40 and S96, thereby designing an approach in which three segments are generated and sequentially ligated to create an acetylated, double-cysteine histone mutant. In this context, the N-terminal and middle peptides, with the latter bearing the acetylated K56, were synthesized with C-terminal thioesters. The first ligation step consisted of connecting the middle and C-terminal peptides, at which point the thiazolidine (Thz) group protecting the N-terminal cysteine residue of the middle segment was removed. The resulting product was ligated to the N-terminal peptide to generate the H3 (R40C, S96C, C110A) protein acetylated at K56. The overall yield of this approach was extremely poor (2%) due to product loss following purification by high-performance liquid chromatography (HPLC). However, acetylated H3 at K56 was still obtained in sufficient amounts to enable its use in preliminary studies.[271]

Since the incorporation of non-native cysteine residues in the H3 sequence led to changes in the structure and dynamics of the assembled nucleosomes, this motivated the need for a desulfurization step following NCL. As a result, a second NCL strategy was devised in which the two native alanine residues A47 and A91 in the *X. laevis* H3 (C110A) sequence were selected as the ligation sites. Three peptides were synthesized, with the A47 residue incorporated as Thz in the middle segment, which again contains the acetylated K56, and the A91 residue introduced as

a cysteine in the C-terminal fragment. Following ligation of the middle and C-terminal peptides, ring opening of Thz with methoxylamine afforded the N-terminal cysteine of this product. The latter was subsequently reacted with the N-terminal segment of H3 (C110A) to give the desired acetylated product at K56 upon desulfurization. Although higher than in the previous case, the overall yield of this approach was again quite low (7%). FRET experiments with nucleosomes reconstituted with either recombinant H3 (C110A) or synthetic H3 (C110A) were performed to establish their equivalence, thereby demonstrating that the NCL-desulfurization approach does not introduce any unwanted changes in the structure of the chemically prepared histone.[271]

As previously described in the literature, nucleosomes can change their conformational equilibrium toward an unwrapped and accessible state that enables proteins to access buried stretches of nucleosomal DNA.[272] Thus, with the prepared H3 (C110A) protein acetylated at K56, the biological role of this PTM on nucleosome dynamics and protein binding to a buried DNA target site was evaluated. To this end, nucleosomes were prepared from Cy3-labelled DNA containing a LexA binding site and histone octamers bearing Cy5-labelled H2A with either unmodified H3, H3K56Q, recombinant H3 (C110A), synthetic H3 (C110A), H3K56Q (C110A) or H3K56Ac (C110A). The juxtaposition of the Cy3 and Cy5 labeling reagents close to the nucleosome entry-exit site enabled the analysis of DNA movement relative to the histone octamer. Experiments revealed a reduced FRET efficiency for nucleosomes reconstituted from H3K56Ac, (C110A), H3K56Q and H3K56Q (C110A), but slightly increased for recombinant H3 (C110A) and unaltered for synthetic H3 (C110A) relative to unmodified H3. These results suggested that H3K56 acetylation increases the distance between DNA and the histone octamer at the nucleosome entry-exit site and that the K56Q substitution mimics the effect of acetylation at this lysine residue on the steady-state nucleosome structure.

The binding of LexA to its nucleosomal DNA sequence resulted in a decrease in FRET efficiency, as expected. The site exposure equilibrium for nucleosomes containing H3K56Ac (C110A) or H3K56Q (C110A) was approximately twice that of nucleosomes reconstituted with either recombinant or synthetic H3 (C110A), which is equivalent to saying that the probability of LexA binding is increased by the same amount. EMSA in combination with hydroxyl radical cleavage experiments using iron-(S)-1-[p-(bromoacetamido)benzyl]ethylenediaminetetraacetate (FeBABE) revealed that the reduction in FRET efficiency and altered mobility shift observed for nucleosomes containing either H3K56Ac (C110A) or H3K56Q (C110A) were due to increased DNA unwrapping rather than its repositioning around the nucleosome.[271] The FeBABE reagent binds to the free sulfhydryl group of proteins, in this case the engineered thiol of H4S47C, to form a tethered iron chelate. The activation of the latter with peroxide and ascorbate generates hydroxyl radicals which cleave polypeptide chains and nucleic acid backbones within 12 Å of the conjugation site.[273] Thus, FeBABE is useful to map the contact sites between DNA and histones within nucleosomes.

Since initial experiments were performed at a low ionic strength of 1 mM Na^+, LexA binding studies were subsequently carried out at the more biologically relevant salt concentrations of 75 mM and 130 mM NaCl. These also revealed that H3K56 acetylation and its corresponding

glutamine mimic similarly increase LexA binding to its target sequence at physiologically-relevant ionic strengths.[271] In summary, this total chemical synthesis approach relying on SPPS and sequential NCL followed by desulfurization enabled the generation of site-specifically acetylated H3 at K56. The combined use of quantitative biophysical methodologies, in this case FRET, with biological studies such as LexA binding assays, FeBABE mapping and EMSA enabled the study of H3K56 acetylation and its respective glutamine mimic on nucleosomal DNA accessibility at differing ionic strengths.

1.3.4 Use of affinity ligands tethered to acetyl transfer catalysts

The last strategy described here to acetylate histones consists of using affinity ligands tethered to acetyl transfer catalysts. Its most representative example employs a catalyst with a 4-dimethylaminopyridine (DMAP) moiety conjugated to a suitable ligand. DMAP can activate a (thio)ester from a donor molecule and thus catalyze the transfer of the latter's acetyl group to a nucleophile, such as the side chain amine of a lysine residue. To promote site-specific acetylation, the DMAP-based catalyst is conjugated to a ligand that can specifically recognize the target protein containing the lysine(s) of interest. Since this transfer is spatially constrained, only the lysine residues in proximity to the ligand-binding site become acetylated. Rather than acetylating histones in nucleosomes, this strategy was originally developed to label carbohydrate-binding proteins known as lectins with a variety of acyl groups.[274]

This ligand-directed concept was expanded for the site-specific acetylation at K120 of H2B in reconstituted nucleosomes. To this end, the N-terminus (residues 5-15) of the Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen (LANA) was employed.[275] This N-terminal LANA peptide specifically interacts with the H2A-H2B acidic patch within the nucleosome.[276] By conjugating it to DSH (that is, a DMAP catalyst containing a mercaptomethyl group in the second position) and using acetyl-CoA as the acetyl donor, site-specific acetylation at K120 of H2B was achieved within nucleosomes. However, the yield of this reaction was 69 ± 6%, which means that a significant amount of H2BK120 remained non-acetylated. Additionally, minor acetylation at K116 of H2B (15.1 ± 0.3%) was detected due to the proximity of this residue to H2BK120. Other acylation types were also carried out by incorporating distinct acyl groups in the donor scaffold. Still, acylation yields did not exceed 65 ± 5-6%. Biological assays performed with reconstituted nucleosomes containing acetylated H2B at K120 revealed that this modification reduces internucleosomal interactions, thereby modulating higher-order chromatin structures.[275] A schematic illustration of the ligand-directed catalysis approach employing DMAP-LANA to acetylate H2B at K120 in reconstituted nucleosomes is shown in **Figure 14**.

Recently, the DMAP-LANA approach has been extended to acetylate H2B at K120 in living cells.[277] However, the catalyst system was altered to enable in-cell acetylation in accordance with prior results. First, rather than using acetyl-CoA due to its low membrane permeability, the cell-permeable N,S-diacetylcysteamine was employed instead as the acetyl donor. Second, the prodrug version of DSH in which the thiol group is protected as a disulfide with methanethiol was used.[278] This was necessary to prevent DSH oxidation to its dimer outside cells, which would

render it less membrane-permeable than the monomer, and a reduction in its catalytic activity within cells due to the presence of numerous nucleophiles, such as glutathione. Third, a PEG moiety was conjugated to the LANA peptide to stabilize it against proteolysis. With this optimized catalyst system, a yield of up to 51% was obtained for the acetylation of H2B at K120 in cellular nucleosomes.[277]

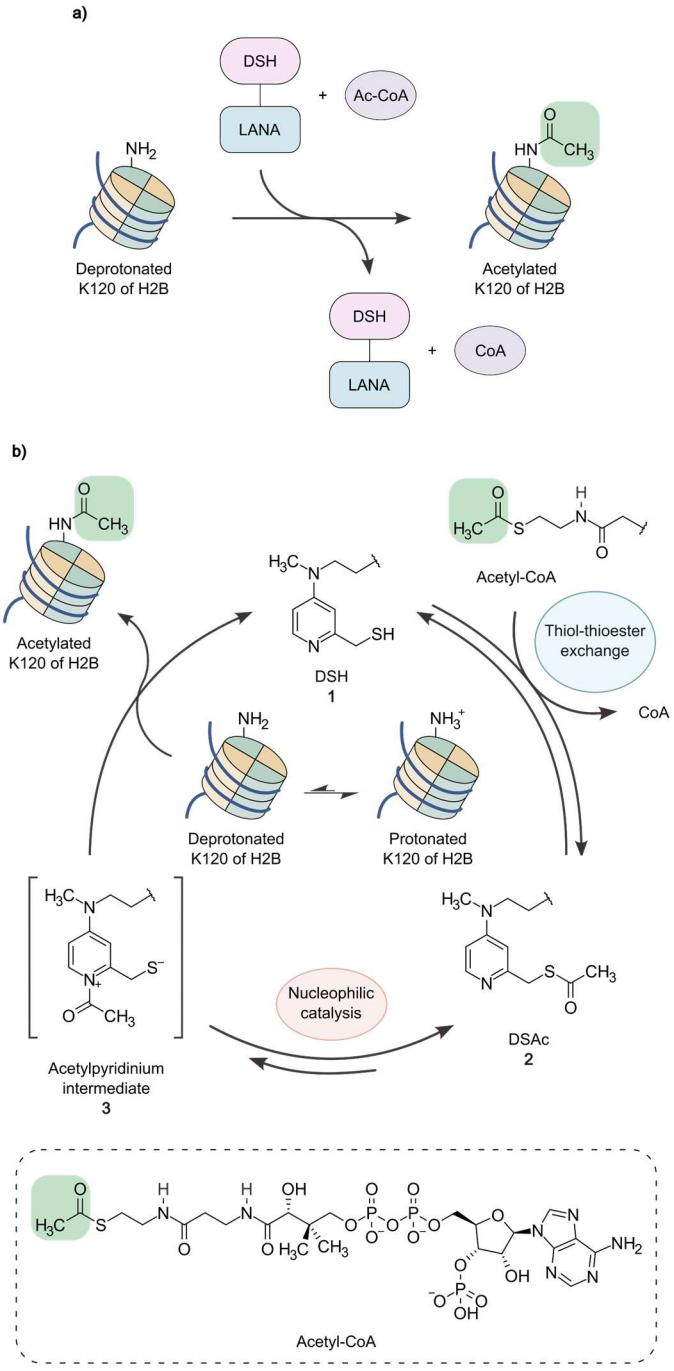


Figure 14. Schematic illustration of the ligand-directed catalysis approach to acetylate K120 of H2B with site-specificity. **a)** The catalyst system composed of the LANA peptide coupled to DSH can promote site-specific acetylation at K120 of H2B in the presence of acetyl-CoA; **b)** Mechanism for lysine acetylation by DSH. First, a thiol-thioester exchange reaction between acetyl-CoA and the mercaptomethyl group of DSH generates DSAc. The intramolecular migration of the acetyl group produces an acetylpyridinium intermediate. The activated acetyl group is subsequently transferred to the ε-amino group of a lysine, thereby acetylating the proximal K120 residue of H2B. This figure was adapted from [275].

1.4 Aims of the thesis

It had been previously hypothesized and demonstrated within the research group that the sequential combination of cysteine conjugation with click chemistry could be applied for the site-specific acetylation of histone peptides. Considering the biological significance of the H3K9Ac epigenetic mark, this strategy was at the time tested to acetylate H3 peptides at position K9. In this context, two peptides – H3K4C and H3K14C – each containing the first 15 amino acids of the H3 protein in which a single cysteine replaces one of the two surrounding lysine residues of K9, were first conjugated to a maleimide reagent containing a dibenzocyclooctyne (DBCO) moiety. This heterobifunctional crosslinker, with its sulfhydryl reactive group and strain-promoted azide-alkyne cycloaddition (SPAAC) clickable handle, connects the peptide to a lysine modifying reagent. In the presence of the latter, which contains a click counterpart (azide) for the SPAAC reaction and a stable acetyl donor (*gem*-dithioacetate) with a fixed spacer, mono-acetylation occurred at the nearest lysine at position 9.

The aim of the present thesis consisted of extending this cysteine-assisted click chemistry concept to the full-length protein, thereby contributing to the development of a novel strategy for the site-specific acetylation of lysine residues in histones that seeks to overcome many of the disadvantages presented by current methods. These include the mutation of the amino acids of interest to generate acetyl-lysine mimics, which may lead to sub-optimal interactions with their natural binding partners in the case of chemical or post-expression mutagenesis, the lower yield of modified histones with genetic code expansion, the more challenging nature of protein ligation, and the dependency on the existence of corresponding interacting ligands that can promote acetylation of proximal lysine residues in the protein sequence, when using affinity ligands tethered to acetyl transfer catalysts. A schematic illustration of the strategy developed in the context of this thesis is shown in **Figure 15**.

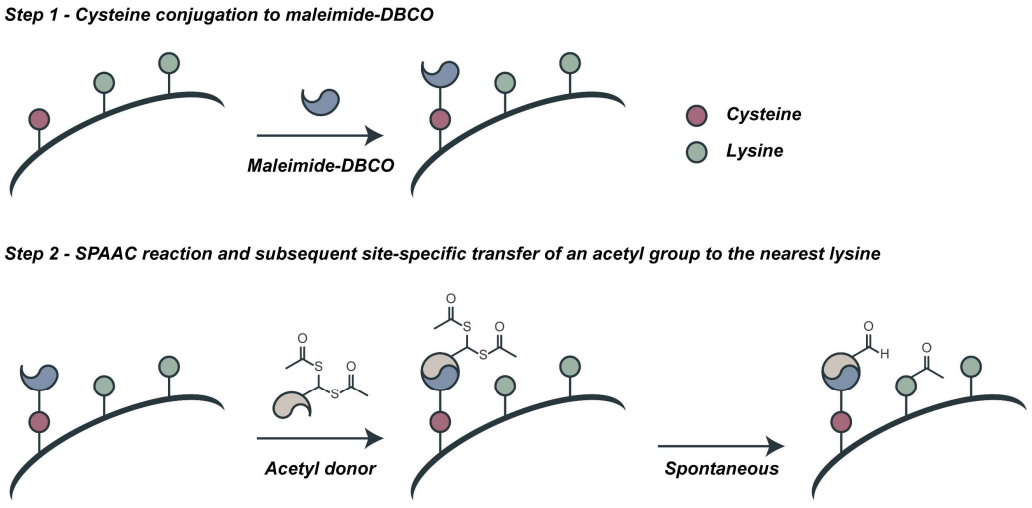


Figure 15. Schematic illustration of the site-specific acetylation strategy developed in this thesis. A cysteine residue is first introduced at a position that is five residues away from the intended lysine to be acetylated. Then, a maleimide-DBCO reagent is installed at this cysteine and in the presence of an azide that acts as an acetyl donor, a SPAAC reaction occurs. This click reaction locks the acetyl donor in proximity to the lysine of interest and acetylation is then triggered by the nucleophilic attack of the side chain amine of this residue.

2. Methods

2.1 General remarks

All chemical reagents and solvents were purchased from commercial sources and used without further purification. The protected histone peptide H3K4C-(1-15) (ARTCQTARKSTGGKA) with N-terminal acetylation and C-terminal amidation was purchased from Biomatik.

Thin layer chromatography was performed using silica gel (60 F_{254} from Merck) and visualized by ultraviolet (UV) light. Synthesized compounds were purified by column chromatography using silica gel (60A, 70-200 μm from Carlo Erba).

Proton (^{1}H) and carbon (^{13}C) nuclear magnetic resonance (NMR) spectra were taken on a Bruker Fourier 300 MHz spectrometer in deuterated chloroform ($CDCl_3$). NMR spectra were assigned according to previous characterization within the group. Chemical shifts are reported on the δ scale in parts per million (ppm) and the spectra were calibrated using the residual solvent peaks as internal standards (^{1}H-NMR: $CDCl_3$ δ = 7.26 ppm; ^{13}C-NMR: $CDCl_3$ δ = 77.16 ppm). Multiplicities are described as s (*singlet*), d (*doublet*), t (*triplet*), q (*quartet*), quint (*quintuplet*), m (*multiplet*), dd (*doublet of doublets*), dt (*doublet of triplets*), tdd (*triplet of doublet of doublets*) and ddd (*doublet of doublet of doublets*). Coupling constants (*J*) are reported in hertz (Hz) to 1 decimal place using MestreNova for signal processing. The center of each peak is reported except for multiplet signals where a range of ppm values is given.

For peptide experiments, liquid chromatography coupled to mass spectrometry (LC-MS) was performed on a Waters QDa mass spectrometer coupled to a Waters Acquity ultra-performance liquid chromatography (UPLC) system using a CORTECS C18 column (2.7 μm, 4.6 mm x 50mm) operated at 35 ^{o}C. Water (solvent A) and acetonitrile (solvent B) containing 0.05% and 0.01% formic acid by volume, respectively, were used as the mobile phases at a flow rate of 0.5 mL/min. The gradient elution was programmed as follows: 99% A for 1.3 min and then a linear gradient to 100% B for 13.7 min. The mobile phase remained 100% B for the next 2 min and a gradient of 1 min back to 99% A was used to re-equilibrate the column. Electrospray ionization (ESI) was conducted with the ion source in positive mode, a capillary voltage of 0.8 kV and a cone voltage of 15 V. The mass-to-charge ratio (m/z) was recorded from 300 to 1250. Nitrogen was used as the nebulizer and desolvation gas. Probe temperature was set to 600 ºC. Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (version 4.1 from Waters). In addition, low resolution mass spectra for the desulfurization experiments in the modified H3K4C-(1-15) peptide were recorded on an ion trap mass spectrometer (Thermo Scientific LCQ Fleet Ion Trap LC-MS) equipped with an electrospray interface operating in positive mode.

For protein experiments, LC-MS was also performed on a Waters QDa mass spectrometer coupled to a Waters Acquity UPLC but using a Waters XBridge Protein BEH C4 column (300 Å, 3.5 μm, 2.1 mm x 100 mm) operated at 40 ^{o}C. Water (solvent A) and acetonitrile (solvent B) containing 0.1% and 0.01% formic acid by volume, respectively, were used as the mobile phases at a flow rate of 0.2 mL/min. The gradient elution was programmed as follows: 90% A for 5 min

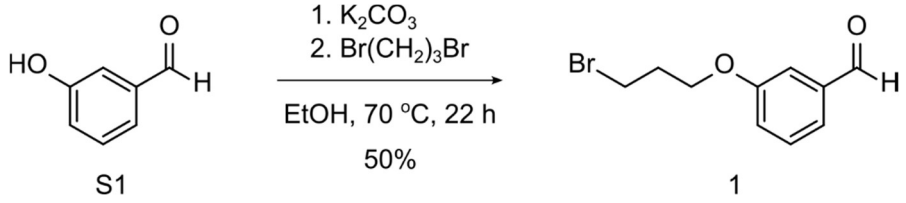
and then a linear gradient to 100% B for 15 min. A gradient of 5 min back to 90% A was used to re-equilibrate the column. ESI was conducted with the ion source in positive mode, a capillary voltage of 1.5 kV and a cone voltage of 20 V. The m/z was recorded from 600 to 1250. Nitrogen was used as the nebulizer and desolvation gas. Probe temperature was set to 400 ºC. Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (version 4.1 from Waters).

Protein concentrations were determined by Nanodrop and the Bradford assay (Bio-Rad).

2.2 Chemical synthesis

The synthesis of the azide that acts as the acetyl donor (compound **4**) was performed within the research group prior to the start of this thesis. However, to generate unacetylated proteins modified with the same SPAAC by-product that remains attached at the cysteine, the synthesis of the corresponding azide counterpart lacking the acetyl donor (compound **2**) was repeated during this doctoral thesis. To elucidate the structure of both compounds, the complete protocol and scheme for their synthesis is illustrated in this section and is referenced from [279].

2.2.1 Synthesis of 3-(3-bromopropoxy)benzaldehyde (compound 1)

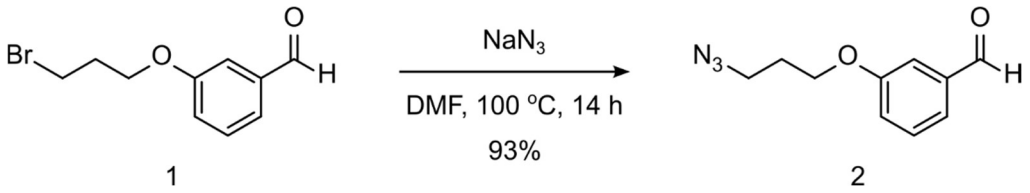


Scheme 1. Synthesis of compound **1**.

Literature Basis: *Chem. Sci.,* **2014**, *5* (2), 489-500 [280]

In a round-bottom flask, under inert atmosphere, 3-hydroxybenzaldehyde (1.0 g, 8.2 mmol) and potassium carbonate (1.7 g, 12.0 mmol) were dissolved in ethanol (47 mL). The reaction mixture was left to react at 70 ºC for 30 min, after which it was cooled to room temperature and 1,3-dibromopropane was added. The mixture was then stirred at 70 ºC for over 22 h. After cooling to room temperature, the reaction mixture was filtrated under vacuum using ethanol as the wash solvent. Concentration under vacuum afforded the crude product that was purified with a flash chromatography to isolate compound **1** in 50% yield (1.0 g). ^{1}H-NMR (300 MHz, $CDCl_3$): δ = 9.93 (s, 1H), 7.45-7.35 (m, 3H), 7.14 (dt, *J* = 6.4, 2.7 Hz, 1H), 4.12 (t, *J* = 5.8 Hz, 2H), 3.58 (t, *J* = 6.4 Hz, 2H), 2.30 (quint, *J* = 6.1 Hz, 2H). ^{13}C-NMR (75 MHz, $CDCl_3$): δ = 192.03, 159.19, 137.75, 130.11, 123.63, 121.76, 112.80, 65.50, 32.10, 29.87.[279] When repeating this synthesis (0.25 g. 2.0 mmol of 3-hydroxybenzaldehyde), a yield of 45% (0.22 g) was obtained for compound **1**.

2.2.2 Synthesis of 3-(3-azidopropoxy)benzaldehyde (compound 2)

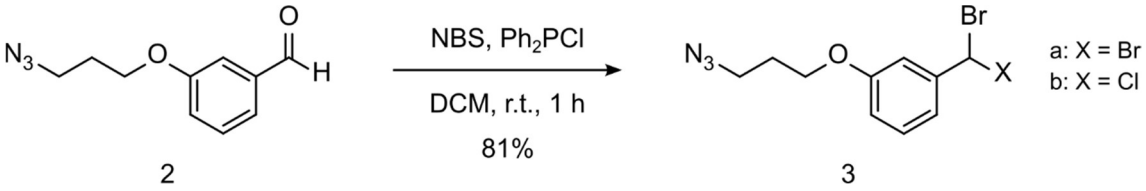


Scheme 2. Synthesis of compound **2**.

Literature Basis: *Chem. Sci.,* **2014**, *5* (2), 489-500 [280]

In a round-bottom flask, compound **1** (0.89 g, 3.7 mmol) was dissolved in DMF (10.2 mL) at room temperature. Sodium azide (0.315 g, 4.8 mmol) was then added to this solution and the reaction mixture heated to 100 °C. After 14 h, the mixture was poured to water and extracted with dichloromethane. Concentration under vacuum afforded the crude product that was purified with a flash chromatography to isolate compound **2** in 93% yield (0.7 g). ^{1}H-NMR (300 MHz, $CDCl_3$): δ = 9.92 (s, 1H), 7.43-7.34 (m, 3H), 7.13 (dt, *J* = 6.3, 2.7 Hz, 1H), 4.06 (t, *J* = 5.9 Hz, 2H), 3.48 (t, *J* = 6.6 Hz, 2H), 2.03 (quint, *J* = 6.3 Hz, 2H). ^{13}C-NMR (75 MHz, $CDCl_3$): δ = 192.02, 159.19, 137.76, 130.09, 123.62, 121.74, 112.73, 64.74, 48.07, 28.59.[279] When repeating this synthesis (0.20 g, 0.82 mmol of compound **1**), a yield of 90% (0.15 g) was obtained for compound **2**.

2.2.3 Synthesis of 1-(3-azidopropoxy)-3-(dibromomethyl)benzene (compound 3a) and 1-(3-azidopropoxy)-3-(bromochloromethyl)benzene (compound 3b)



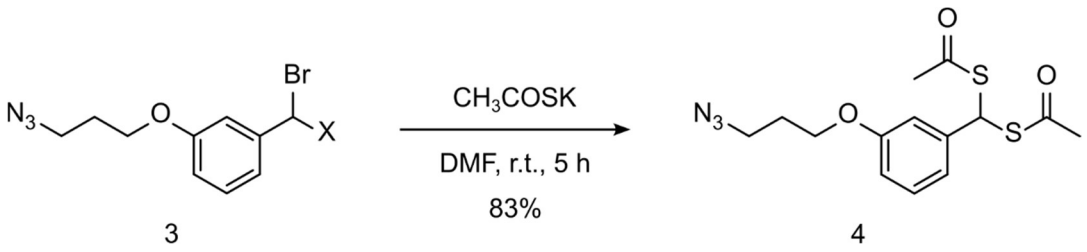
Scheme 3. Synthesis of compound **3a** and **3b**.

Literature Basis: *Synth. Commun.*, **2008**, *38* (22), 4023-4035 [281]

In a flame-dried round-bottom flask, under inert atmosphere, N-bromosuccinimide (NBS, 0.35 g, 1.95 mmol) was dissolved in previously distilled dichloromethane (4 mL), followed by the addition of chlorodiphenylphosphine (0.36 mL, 1.95 mmol) at room temperature. A solution of compound **2** (0.2 g, 0.975 mmol) in freshly distilled dichloromethane (0.9 mL) was added to the reaction mixture and left to stir for 1 h. Then, 5 mL of water were added to the mixture and a multiple extraction was performed with dichloromethane. The combined organic layers were washed with brine, dried over anhydrous magnesium sulfate, filtered and concentrated under vacuum. A flash chromatography was performed to isolate a mixture of compounds **3a** and **3b** in 81% yield (0.28 g, 25% of 3a and 75% of 3b). ^{1}H-NMR (300 MHz, $CDCl_3$): δ = 7.26-7.19 (m, 1H), 7.10-7.05 (m, 2H), 6.82 (tdd, J = 8.3, 2.5, 0.9 Hz, 1H), 6.56 (s, 1H), 4.02 (t, *J* = 5.9 Hz, 2H), 3.47 (t, *J* = 6.6 Hz, 2H), 2.01 (quint, *J* = 6.4 Hz, 2H). ^{13}C-NMR (75 MHz, $CDCl_3$): δ = 158.83, 158.71,

143.32, 142.63, 129.83, 129.72, 118.73, 118.55, 116.15, 112.90, 112.44, 112.28, 64.74, 57.30, 48.21, 40.93, 28.76.[279]

2.2.4 Synthesis of (3-(3-azidopropoxy)phenyl)methylenediethanethioate (compound 4)



Scheme 4. Synthesis of compound **4**.

Literature Basis: *Org. Lett.*, **2014**, *16* (17), 4536-4539 [282]

In a round-bottom flask, under inert atmosphere, compound **3** (0.23 g, 0.66 mmol) was mixed with potassium thioacetate (0.23 g, 2.0 mmol) dissolved in dried DMF (3.4 mL) at room temperature. The reaction was stirred for 5 h, after which 12 mL of water were added. A multiple extraction was performed using dichloromethane and the combined organic layers were dried over anhydrous sodium sulfate, filtered and concentrated under vacuum. A flash chromatography was performed to isolate compound **4** in 83% yield (0.18 g). ^{1}H-NMR (300 MHz, $CDCl_3$): δ = 7.21 (t, *J* = 7.9 Hz, 1H), 7.03-6.96 (m, 2H), 6.79 (ddd, *J* = 8.3, 2.5, 0.9 Hz, 1H), 6.21 (s, 1H), 4.03 (t, *J* = 5.9 Hz, 2H), 3.50 (t, *J* = 6.6 Hz, 2H), 2.31 (s, 6H), 2.03 (quint, *J* = 6.5 Hz, 2H). ^{13}C-NMR (75 MHz, $CDCl_3$): δ = 192.59, 158.75,140.88, 129.85, 120.28, 114.29, 113.91, 64.52, 48.21, 48.01, 30.04, 28.75.[279]

2.3 Experiments in the H3K4C-(1-15) peptide

The H3K4C-(1-15) peptide was reconstituted in water (1 mg/mL) and aliquoted at -80 °C without any further purification steps. Stock solutions of maleimide-DBCO (4 mM in DMF), compound **4** (8 mM in DMF) and ammonium acetate buffer (20 mM, pH 8.0) were freshly prepared prior to use.

2.3.1 General procedure to generate the H3K4C**-(1-15)-K9Ac peptide

Conjugation at the free cysteine of the H3K4C-(1-15) peptide with the maleimide-DBCO clickable handle and site-specific acetylation following SPAAC reaction with compound **4**

An aliquot of histone peptide H3K4C-(1-15) was diluted to 50 µM in ammonium acetate buffer (20 mM, pH 8.0) and 4 equiv. of maleimide-DBCO (400 µM) were then added. The reaction mixture was left to react at 37 °C for 1 h, after which an aliquot was analyzed by LC-MS. Subsequently, 16 equiv. of compound **4** (800 µM) were added and the reaction was shaken for an additional hour at the same temperature. Following centrifugation, the mixture was analyzed by LC-MS.

2.3.2 Desulfurization attempt in the H3K4C**-(1-15)-K9Ac peptide

The desulfurization reaction of the H3K4C**-(1-15)-K9Ac peptide was carried out by adding 100 equiv. of a premixed TCEP and sodium borohydride ($NaBH_4$) solution (1:1 molar ratio) in ammonium acetate buffer (20 mM, pH 7.0). The reaction mixture was left to react at 37 °C for 3 h, after which it was analyzed by MS.

2.4 Expression and purification of mutant histone proteins

The pET-3a plasmid encoding histone H3 (C110A) from *X. laevis* was a kind gift of Professor Ernest Laue (Department of Biochemistry, University of Cambridge). This plasmid contained the T7 promoter for the expression of H3 (C110A). The K4C, R52C and K4CR52C mutants of the H3 (C110A) coding sequence were constructed by site-directed mutagenesis using the NZYMutagenesis kit (NZYTech) according to the manufacturer's instructions. Primers were obtained from Sigma and the mutations were confirmed by Sanger DNA sequencing at Eurofins Scientific. The following primer sequences were used to introduce the K4C and R52C mutations in the pET-3a-H3 plasmid by polymerase chain reaction:

K4C

5'- ATGGCCCGTACCTGCCAGACCGCCCGT-3'

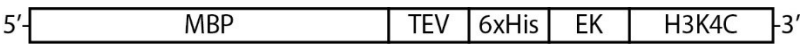
5'- ACGGGCGGTCTGGCAGGTACGGGCCAT-3'

R52C

5'-CTCTCCGCGAGATCTGCCGCTACCAGAAA-3'

5'-TTTCTGGTAGCGGCAGATCTCGCGGAG-3'

Prior to the start of this thesis, the H3K4C (C110A) coding sequence had been cloned into a plasmid downstream from a maltose-binding protein (MBP) sequence. In addition, the plasmid backbone contained a six-residue polyhistidine (6xHis) tag between the MBP and H3K4C sequences, which was flanked upstream and downstream by a Tobacco etch virus (TEV) protease and an enterokinase (EK) recognition sites, respectively. For the expression of the MBP-H3K4C fusion protein, the tac promoter was used. A schematic representation of the construct corresponding to the fusion protein just described is shown below (**Scheme 5**).



Scheme 5. Construct corresponding to the MBP-H3K4C fusion protein.

2.4.1 Expression, purification and enzymatic cleavage of MBP-H3K4C

Expression of MBP-H3K4C

The mutant MBP-H3K4C protein was recombinantly expressed in *E. coli* BL21(DE3) cells (NZYTech) following bacterial transformation with its respective expression plasmid. Cells were grown at 37 °C and 220 rpm in Luria broth (LB) medium containing carbenicillin (100 µg/mL) until the optical density reached 0.5-0.7. Protein expression was then induced with 1 mM isopropyl β-

D-thiogalactopyranoside (IPTG) for 2 h at 37 ºC and 220 rpm. Cells were harvested by centrifugation at 7,000 rpm and 4 ºC for 10 min and stored at -80 ºC. Protein expression was confirmed by sodium dodecyl-sulfate (SDS)-PAGE.

Purification of MBP-H3K4C

The MBP-H3K4C protein was purified by affinity chromatography. Briefly, the bacterial pellet was resuspended in wash buffer containing 50 mM Tris pH 7.4, 2 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM dithiothreitol (DTT), supplemented with DNase I (Roche) and protease inhibitors (Roche). The resuspended pellet was sonicated on ice with an amplitude of 20% using 5 pulses of 30 seconds, each followed by a pause of 30 seconds. The extract was cleared by centrifugation at 22,000 g and 4 ºC for 20 min and the pellet resuspended in wash buffer (50 mM Tris pH 7.4, 200 mM NaCl, 1 mM EDTA) with 1% Triton X-100 to partially solubilize the inclusion bodies and thus recover a fraction of protein present in these structures. The resulting suspension was sonicated with an amplitude of 20% using 2 pulses of 30 seconds, each followed by a pause of 30 seconds. The extract was cleared by centrifugation at 22,000 g for 10 min and the resulting supernatant separated from the pellet. The latter was again resuspended in wash buffer with 1% Triton X-100, sonicated and the suspension centrifuged. The resulting pellet was resuspended in wash buffer without Triton X-100, sonicated and the extract cleared by centrifugation. This step was repeated once more.

The washing samples containing Triton X-100 were combined and passed through a PD-10 desalting column (Cytiva) to remove this detergent. The eluted fraction was combined with the washing samples that did not contain Triton X-100 and concentrated with Amicon Ultra-15 mL centrifugal filter units (Merck Millipore). The resulting sample was filtered and loaded into a MBPTrap column (GE Healthcare, 1 mL), which was pre-equilibrated with binding buffer (20 mM Tris pH 7.4, 200 mM NaCl, 1 mM DTT, 0.1 mM EDTA) and connected to a fast protein liquid chromatography (FPLC) system (ÄKTA, GE Healthcare). The MBP-H3K4C protein was then eluted with a maltose gradient. Histone-containing fractions were pooled, buffer-exchanged to the digestion buffer of TEV or EK proteases and kept on ice until further use.

Enzymatic cleavage with TEV protease

Cleavage of MBP-H3K4C (45 µg) with HaloTEV protease (Promega, 10 µL, 50 units) was attempted in reaction buffer (50 mM Tris pH 8.0, 0.5 mM EDTA, 1 mM DTT) using two different experimental conditions, namely (1) 30 min at 30 ºC and then overnight at 4 ºC with agitation, and (2) 1 h at room temperature, followed by overnight incubation at 4 ºC without agitation. Aliquots of the two reactions were taken for SDS-PAGE analysis.

Enzymatic cleavage with EK protease

The EK enzyme (100 µg, Peprotech) was reconstituted in 100 µL of water, aliquoted, flash-frozen in liquid nitrogen and stored at -80 ºC. Prior to use, a working solution of EK was prepared by diluting 5 µL of the enzyme in 5 µL of reaction buffer (40 mM Tris pH 7.4, 100 mM NaCl.

Cleavage of MBP-H3K4C (at a concentration of 1 mg/ml) with EK (at a concentration of 0.5 µg/µL) was tested at a 40:1 ratio (w/w) of protein to enzyme using three different experimental conditions, namely (1) 23 °C, (2) 37 °C, and (3) 23 °C in the presence of 4 M urea and 20 mM lysine. Aliquots of the three reactions were taken at the 2 h and 4 h time-points for SDS-PAGE analysis.

2.4.2 Expression and purification of H3 mutant proteins without tags

Expression of H3K4C, H3R52C and H3K4CR52C

The mutant H3K4C, H3R52C and H3K4CR52C proteins were recombinantly expressed in *E. coli* BL21(DE3) cells (NZYTech) following bacterial transformation with their respective expression plasmids. Cells were grown at 37 °C and 220 rpm in LB containing carbenicillin (100 µg/mL) until the optical density reached 0.5-0.7. Protein expression was then induced with 1 mM IPTG for 2 h at 37 °C and 220 rpm. Cells were harvested by centrifugation at 7,000 rpm and 4 °C for 10 min and stored at -80 °C. Protein expression was confirmed by SDS-PAGE.

Purification of H3K4C, H3R52C and H3K4CR52C

Histone H3K4C, H3R52C and H3K4CR52C mutants were purified by anion and cation exchange chromatography in tandem, essentially as described in the literature.[283] Briefly, the bacterial pellet was resuspended in sodium-acetate-urea (SAU) buffer (40 mM sodium acetate pH 5.2, 7 M urea, 10 mM lysine, 1 mM EDTA pH 8.0) containing 5 mM β-mercaptoethanol (BME), 200 mM NaCl, protease inhibitors and DNase. Defined buffer conditions were achieved by directly adding to the pellet DNase I (Roche) and protease inhibitors (Roche) in powder, followed by the addition of 10x SA buffer (400 mM sodium acetate pH 5.2, 10 mM EDTA pH 8.0, 100 mM lysine), BME and NaCl to a final concentration of 5 mM and 200 mM, respectively. Once the cells were resuspended, urea was added to a concentration of 7 M and the suspension filled up to its final volume with water. All steps during lysis were performed on ice.

The suspension was sonicated on ice with an amplitude of 20% using 20 pulses of 15 seconds, each followed by a pause of 30 seconds. The extract was cleared by centrifugation at 40,000g and 4 °C for 30 min. The resulting supernatant was filtered and loaded into a HiTrap Q HP column (GE Healthcare, 5 mL) stacked on top of a HiTrap SP HP column (GE Healthcare, 1 mL), which were pre-equilibrated with SAU buffer containing 5 mM BME and 200 mM NaCl, and connected to an FPLC system (ÄKTA, GE Healthcare). When the extract had passed completely through the Q column, the latter was removed from the FPLC system and the SP column was washed with 200 mM NaCl for several column volumes. Histones were then eluted with a NaCl gradient. Histone-containing fractions were pooled, buffer-exchanged to Tris (15 mM, pH 7.5) with Amicon Ultra-15 mL centrifugal filter units (Merck Millipore) and lyophilized.

The amino acid sequences of the three histone mutants used in this work are shown below and these were used to compute the expected masses of these proteins using ExPASy.[284] This was essential to determine whether purified fractions contained the expressed proteins suitable for conjugation reactions and later evaluate if these proteins were successfully modified by comparing the expected masses with the observed ones in MS experiments.

Histone H3K4C (C110A)

ART**C**QTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKST ELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTIMPKDI QLARRIRGERA

Histone H3R52C (C110A)

ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREI**C**RYQKST ELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTIMPKDI QLARRIRGERA

Histone H3K4CR52C (C110A)

ART**C**QTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREI**C**RYQKST ELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTIMPKDI QLARRIRGERA

2.5 Optimized procedure to generate acetylated H3 proteins

Stock solutions of maleimide-DBCO (2.5 mM), N-(5-nitro-ortho-tolyl)-maleimide (2.5 mM), compound **2** (1 mM and 2 mM), compound **4** (1 mM and 2 mM) and 5,5′-dithiobis(2-nitrobenzoic acid (Ellman's reagent, 2.5 mM) were freshly prepared in anhydrous DMF prior to use.

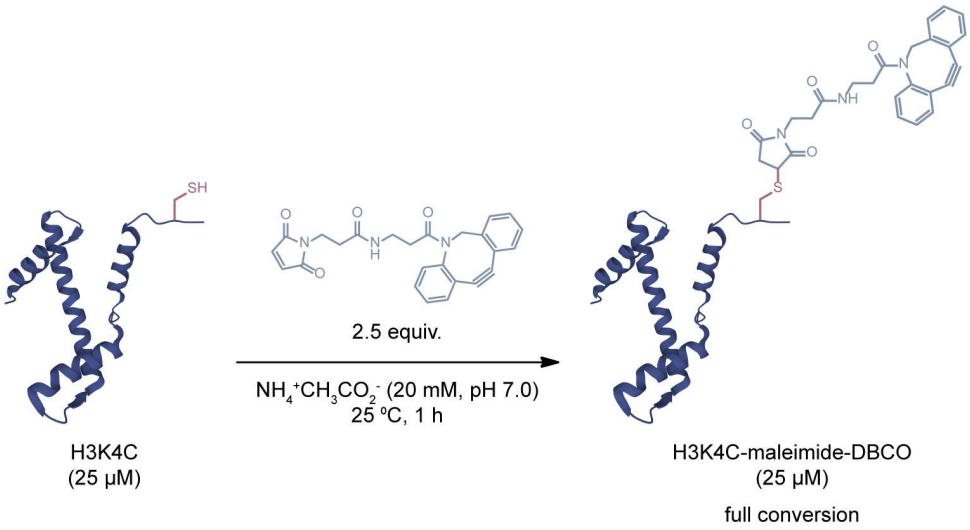
2.5.1 Preparation of H3K4C, H3R52C and H3K4CR52C samples for conjugation reactions

The preparation of histone samples for subsequent chemical modification was based on work described in the literature.[211] Following lyophilization, approximately 2 to 4 mg of histones were dissolved in 400 µL of Milli-Q water / ammonium acetate buffer (20 mM, pH 7.0) and treated with a slight excess of TCEP for 10 to 30 min at 21 °C to reduce any contaminant disulfide. The reduced sample was then passed through a PD MiniTrap G-25 column (Cytiva) and eluted with 1.0 mL of Milli-Q water. The resulting protein sample was kept on ice and quantified by the Bradford assay.

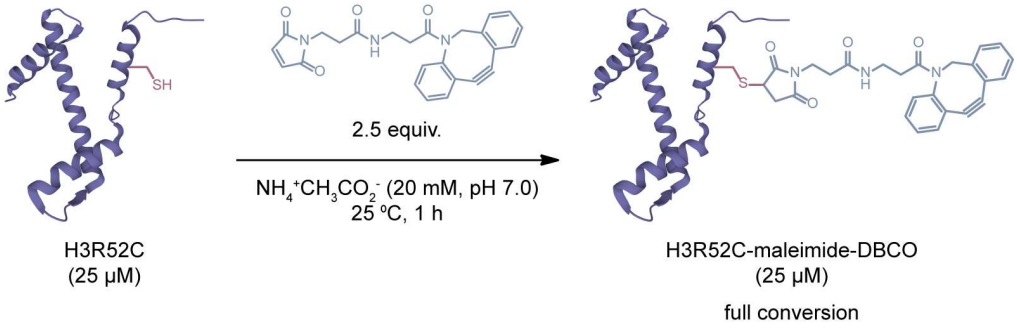
2.5.2 Conjugation of H3K4C, H3R52C and H3K4CR52C with the maleimide-DBCO clickable handle

Aliquots of reduced H3K4C, H3R52C and H3K4CR52C were diluted to 25 µM with ammonium acetate buffer (20 mM, pH 7.0). Then, an aliquot of 2.5 equiv. of maleimide-DBCO was added to each mixture and the reactions were shaken at 400 rpm and 25 °C. After 1 h, the reaction mixtures were analyzed by LC-MS and complete conversion to the H3K4C (calculated mass = 15641 Da; observed mass = 15639 Da), H3R52C (calculated mass = 15613 Da; observed mass = 15611 Da) and H3K4CR52C (calculated mass = 16016 Da; observed mass = 16022 Da) proteins containing the maleimide-DBCO clickable handle was observed. A schematic representation of these reactions is shown in **Scheme 6, Scheme 7** and **Scheme 8**. The resulting deconvoluted ESI-MS spectra are shown in **Figure 55**, **Figure 56** and **Figure 64**. The independent aliquots of H3K4C-maleimide-DBCO, H3R52C-maleimide-DBCO and H3K4CR52C-

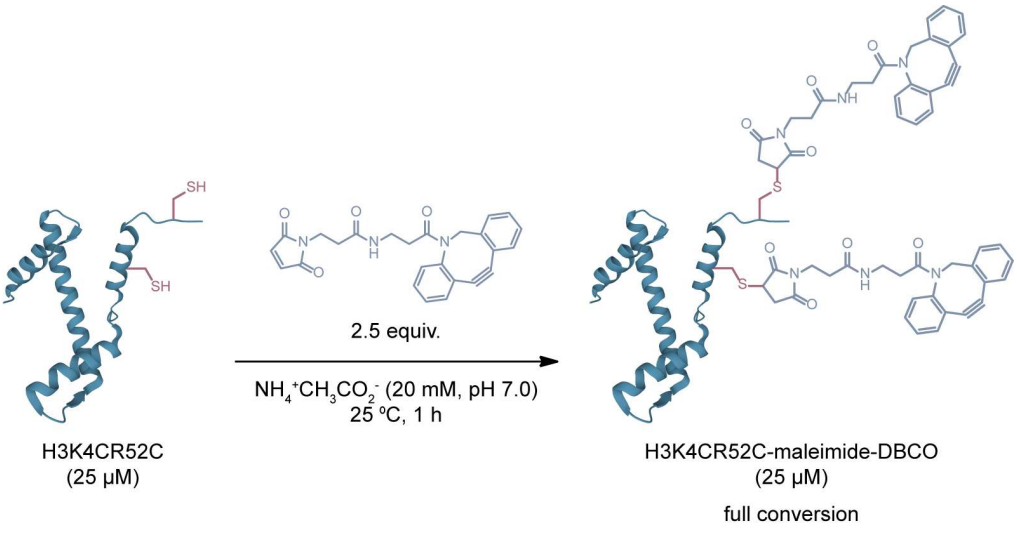
maleimide-DBCO were desalted into water and concentrated using Amicon centrifugal filter units (Merck Millipore). Protein concentrations were then determined by the Bradford assay.



Scheme 6. Site-selective modification of the H3K4C protein with maleimide-DBCO.



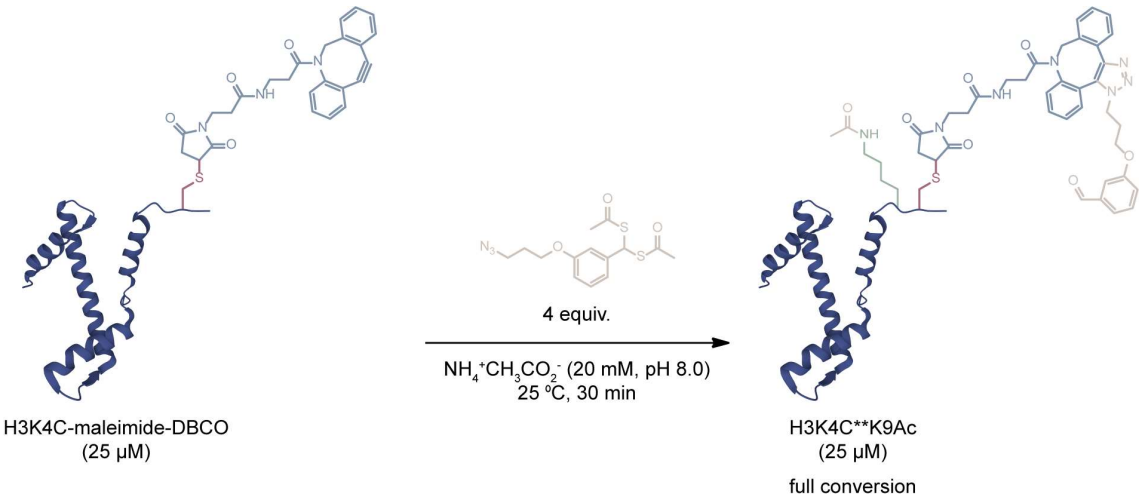
Scheme 7. Site-selective modification of the H3R52C protein with maleimide-DBCO.



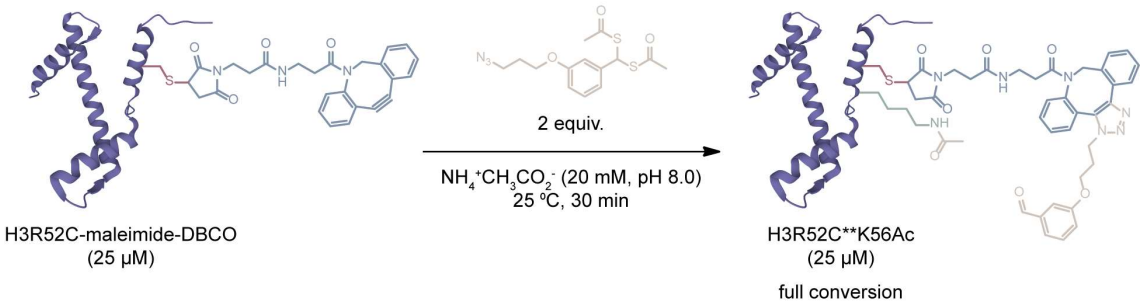
Scheme 8. Site-selective modification of the H3K4CR52C protein with maleimide-DBCO.

2.5.3 Site-specific acetylation of H3K4C, H3R52C and H3K4CR52C proteins conjugated to maleimide-DBCO following SPAAC reaction with compound 4

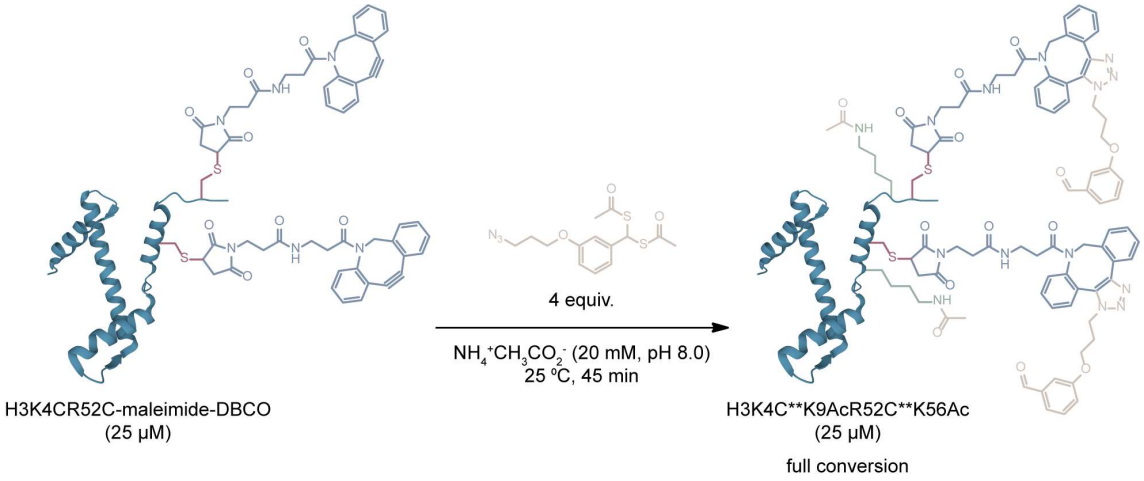
Initially, the SPAAC reaction was performed under basic pH conditions. In this context, independent aliquots of the H3K4C, H3R52C and H3K4CR52C proteins containing the maleimide-DBCO clickable handle were diluted to 25 μM in ammonium acetate buffer (20 mM, pH 8.0). An aliquot of 4 equiv. of compound **4** was added to the diluted H3K4C-maleimide-DBCO solution, while an aliquot of 2 equiv. of the same compound was added to the solution containing the diluted H3R52C-maleimide-DBCO protein. The reactions were shaken at 400 rpm and 25 °C. After 30 min, the reaction mixtures were analyzed by LC-MS and complete conversion to the respective mono-acetylated products was observed (H3K4C**K9Ac: calculated mass = 15889 DA; observed mass = 15890 Da; H3R52C**K56Ac: calculated mass = 15861 Da; observed mass = 15862 Da). An aliquot of 4 equiv. of compound **4** was added to the diluted H3K4CR52C-maleimide-DBCO solution and the reaction was shaken at 400 rpm and 25 °C. After 45 min, the reaction mixture was analyzed by LC-MS and complete conversion to the respective di-acetylated product was observed (H3K4C**K9AcR52C**K56Ac: calculated mass = 16510 Da; observed mass = 16513 Da). A schematic representation of these reactions is shown in **Scheme 9, Scheme 10** and **Scheme 11**. The resulting deconvoluted ESI-MS spectra are shown in **Figure 57**, **Figure 58** and **Figure 65**.



Scheme 9. Site-specific acetylation of the H3K4C-maleimide-DBCO protein when the SPAAC reaction with compound **4** is performed at a basic pH.

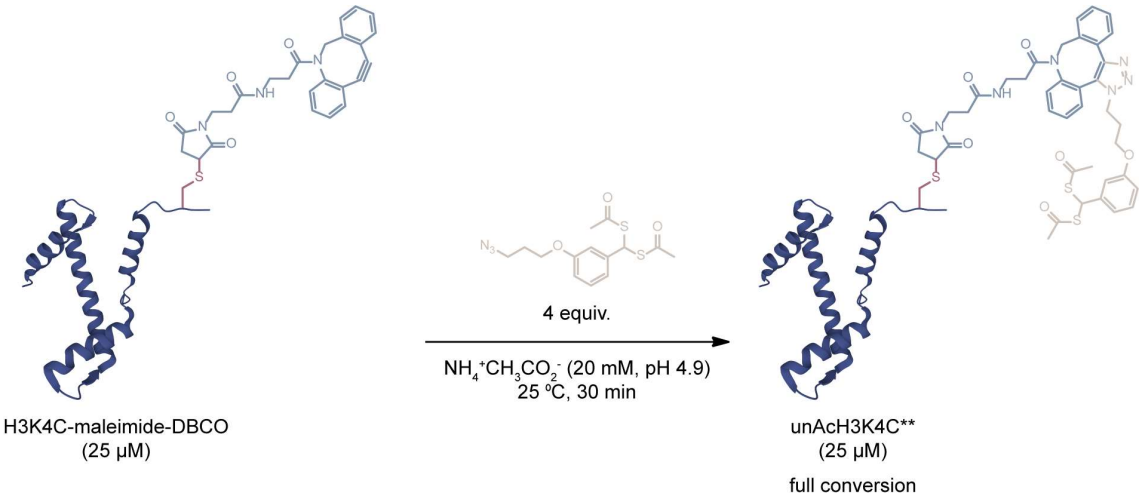


Scheme 10. Site-specific acetylation of the H3R52C-maleimide-DBCO protein when the SPAAC reaction with compound **4** is performed at a basic pH.

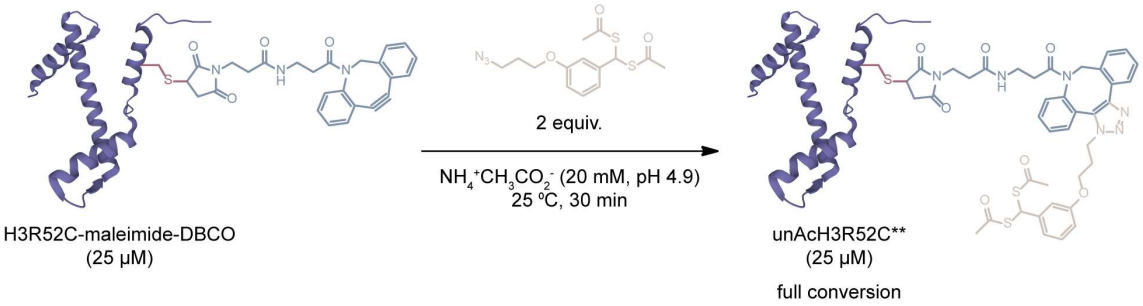


Scheme 11. Site-specific acetylation of the H3K4CR52C-maleimide-DBCO protein when the SPAAC reaction with compound **4** is performed at a basic pH.

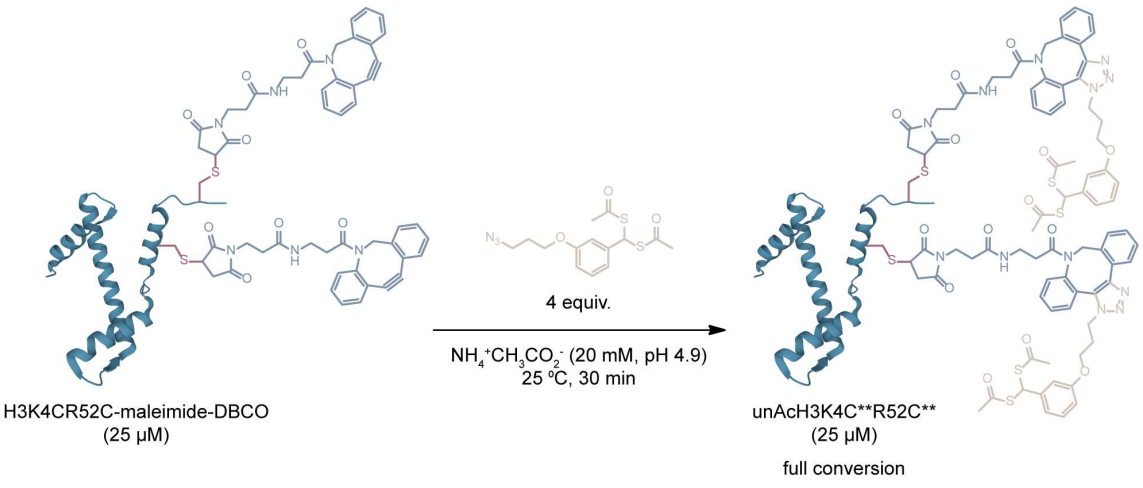
Later, the SPAAC reaction was also performed at an acidic pH. To this end, independent aliquots of the H3K4C, H3R52C and H3K4CR52C proteins containing the maleimide-DBCO clickable handle were diluted to 25 µM in ammonium acetate buffer (20 mM, pH 4.9). An aliquot of 4 equiv. of compound **4** was added to the diluted H3K4C-maleimide-DBCO and H3K4CR52C-maleimide-DBCO solutions, while an aliquot of 2 equiv. of the same compound was added to the solution containing the diluted H3R52C-maleimide-DBCO protein. The reactions were shaken at 400 rpm and 25 ºC. After 30 min, the reaction mixtures were analyzed by LC-MS and full conversion to the respective unacetylated H3K4C, H3R52C and H3K4CR52C SPAAC-containing products with the *gem*-dithioacetate group still left to react was observed (unAcH3K4C**: calculated mass = 15981 Da; observed mass = 15983 Da; unAcH3R52C**: calculated mass = 15953 Da; observed mass = 15956 Da; unAcH3K4C**R52C**: calculated mass = 16695 Da; observed mass = 16699 Da). A schematic representation of these reactions is shown in **Scheme 12**, **Scheme 13** and **Scheme 14**. The resulting ESI-MS spectra are shown in **Figure 59**, **Figure 60** and **Figure 66**.



Scheme 12. SPAAC reaction between the DBCO moiety attached to the H3K4C-maleimide-DBCO protein and the azide counterpart of compound **4**, with the *gem*-dithioacetate still left to react at an acidic pH.



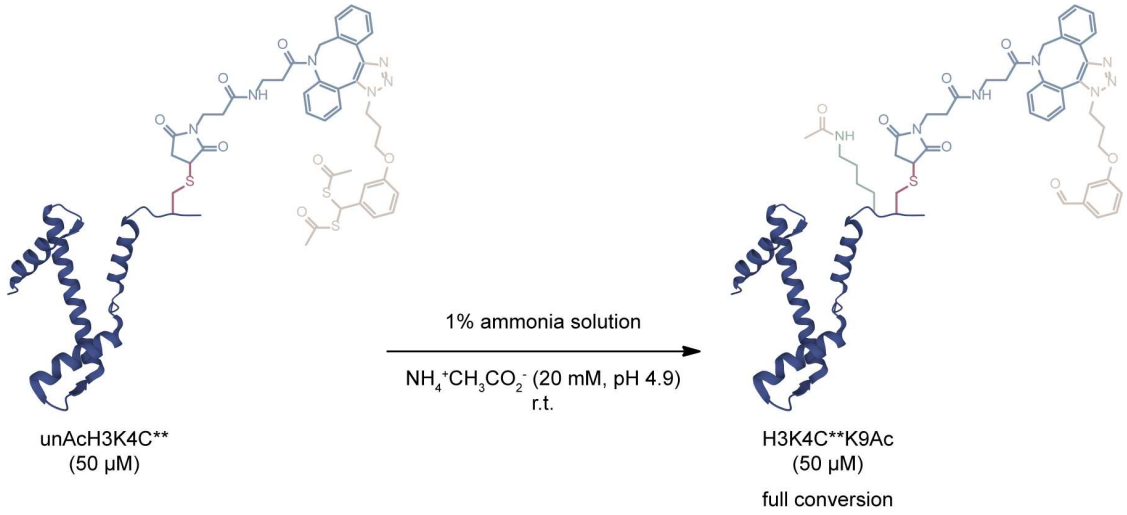
Scheme 13. SPAAC reaction between the DBCO moiety attached to the H3R52C-maleimide-DBCO protein and the azide counterpart of compound **4**, with the *gem*-dithioacetate still left to react at an acidic pH.



Scheme 14. SPAAC reaction between the DBCO moiety attached to the H3K4CR52C-maleimide-DBCO protein and the azide counterpart of compound **4**, with the *gem*-dithioacetate still left to react at an acidic pH.

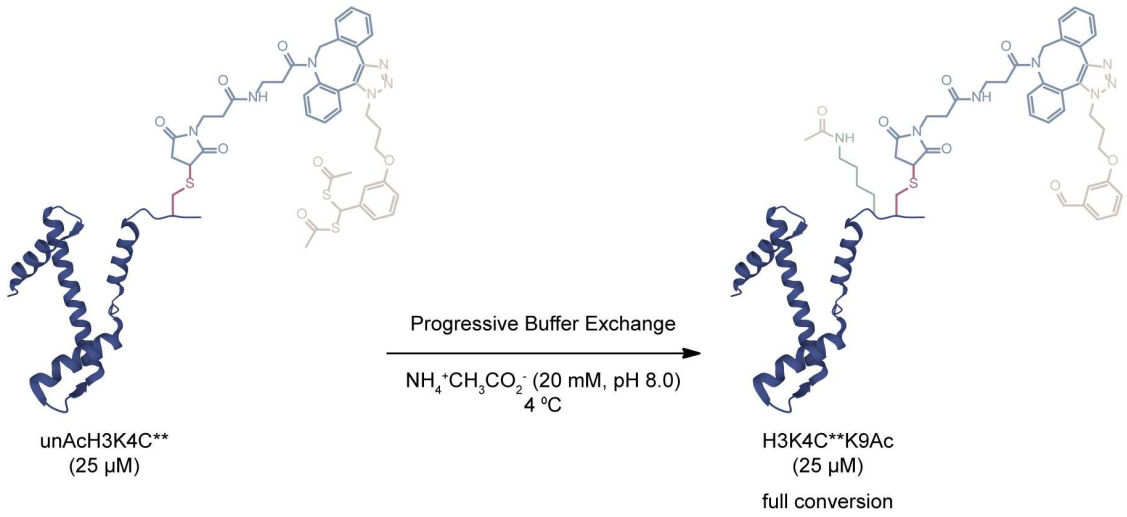
At this point, three approaches were pursued to promote acetylation at K9. In the first approach, the reaction buffer containing unAcH3K4C** was increased to approximately pH 8.0 with a 1% ammonia solution. A schematic representation of this reaction is shown in **Scheme 15**.

The resulting deconvoluted ESI-MS spectrum of the H3K4C**K9Ac protein is shown in **Figure 29** (H3K4C**K9Ac: calculated mass = 15889 Da; observed mass = 15889 Da).



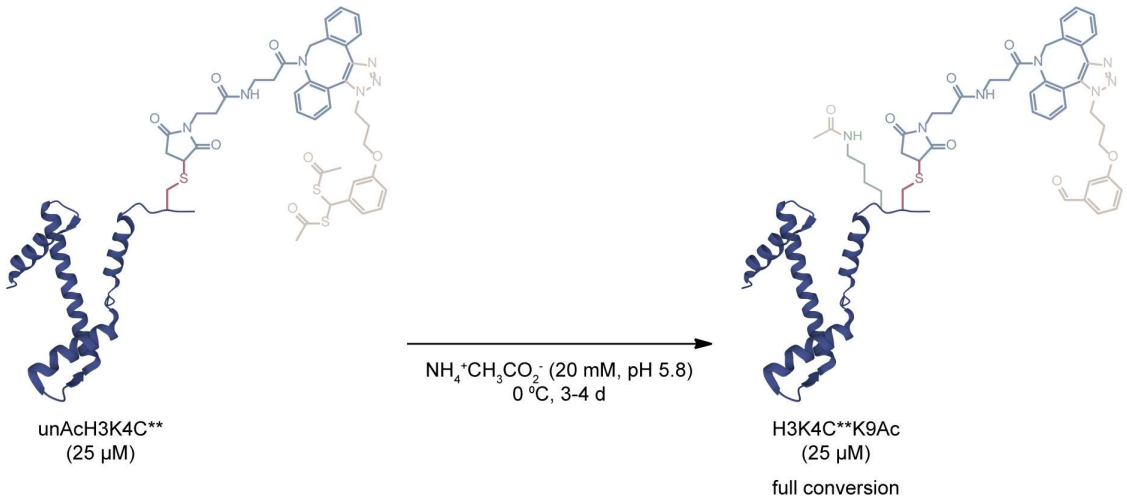
Scheme 15. Site-specific acetylation at K9 when the buffer pH of the unAcH3K4C** protein is increased from 4.9 to approximately 8.0 with a 1% ammonia solution.

In the second approach, an aliquot of unAcH3K4C** was progressively buffer exchanged to ammonium acetate buffer at a basic pH (20 mM, pH 8.0) and then concentrated using Amicon centrifugal filter units (Merck Millipore). Protein concentration was then determined by the Bradford assay and conversion to the corresponding acetylated product was complete as monitored by LC-MS. A schematic representation of this reaction is shown in **Scheme 16**. The resulting deconvoluted ESI-MS spectrum of the H3K4C**K9Ac protein is shown in **Figure 30** (H3K4C**K9Ac: calculated mass = 15889 Da; observed mass = 15882 Da).

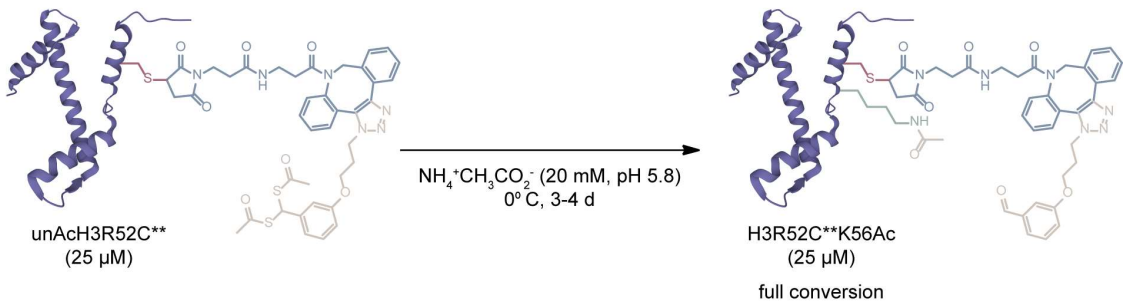


Scheme 16. Site-specific acetylation at K9 when the unAcH3K4C** protein is progressively buffer-exchanged from a pH of 4.9 to 8.0.

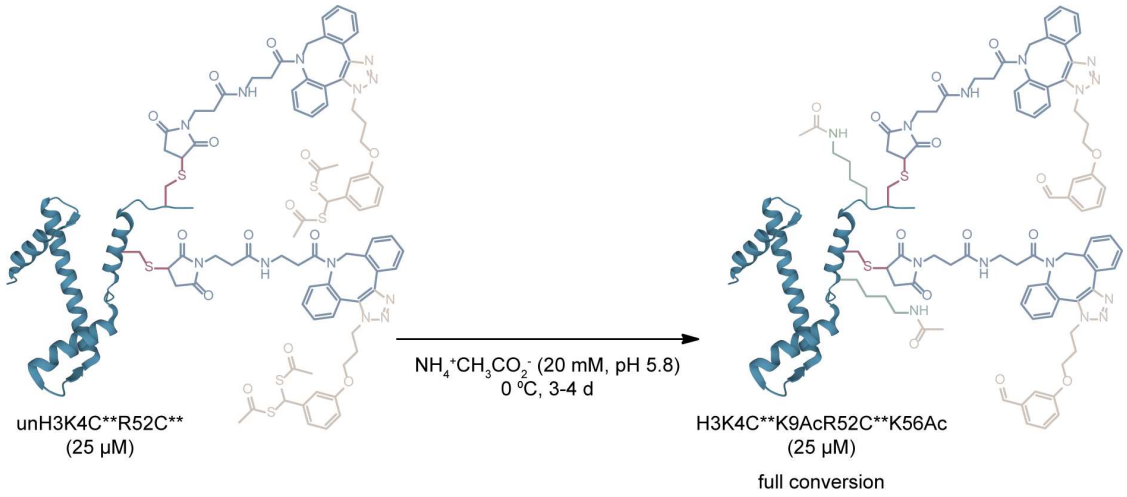
In the third approach, which was also used to promote acetylation at K56 in addition to K9, independent aliquots of unAcH3K4C**, unAcH3R52C** and unAcH3K4C**R52C** were buffer exchanged to ammonium acetate buffer at a slightly less acidic pH (20 mM, pH 5.8) and concentrated using Amicon centrifugal filter units (Merck Millipore). Protein concentrations were then determined by the Bradford assay. The unAcH3K4C**, unAcH3R52C** and unAcH3K4C**R52C** protein samples were placed on ice for three to four days until the spontaneous conversion to their corresponding acetylated products was complete as monitored by LC-MS. A schematic representation of these reactions is shown in **Scheme 17**, **Scheme 18** and **Scheme 19**. The resulting deconvoluted ESI-MS spectra of the H3K4C**K9Ac, H3R52C**K56Ac and H3K4C**K9AcR52C**K56Ac proteins are shown in **Figure 61**, **Figure 62** and **Figure 67**, respectively (H3K4C**K9Ac: calculated mass = 15889 Da; observed mass = 15882 Da; H3R52C**K56Ac: calculated mass = 15861 Da; observed mass = 15862 Da; H3K4C**K9AcR52C**K56Ac: calculated mass = 16510 Da; observed mass = 16513 Da). The acetylated proteins were then desalted into water and quantified by the Bradford assay. Finally, the aliquots of H3K4C**K9Ac, H3R52C**K56Ac and H3K4C**K9AcR52C**K56Ac were lyophilized and frozen at -80 °C for further studies.



Scheme 17. Site-specific acetylation at K9 when the unAcH3K4C** protein is left on ice for 3 to 4 days at a pH of 5.8.



Scheme 18. Site-specific acetylation at K56 when the unAcH3R52C** protein is left on ice for 3 to 4 days at a pH of 5.8.

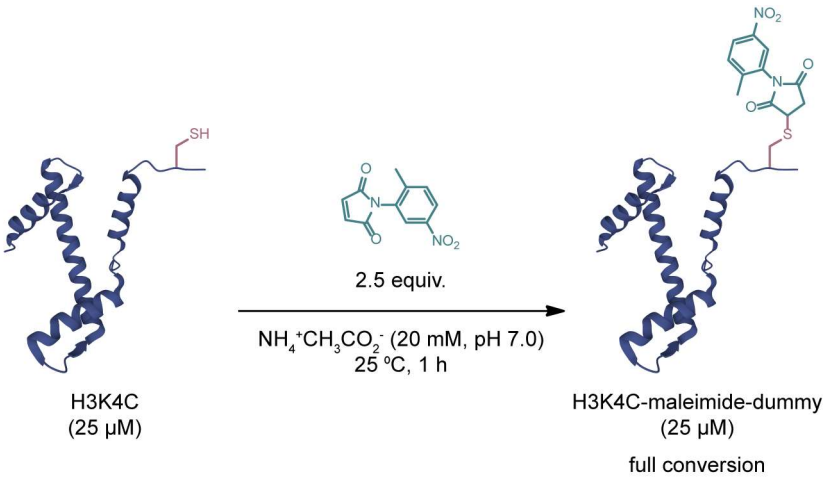


Scheme 19. Site-specific acetylation at K9 and K56 when the unAcH3K4C**R52C** protein is left on ice for 3 to 4 days at a pH of 5.8.

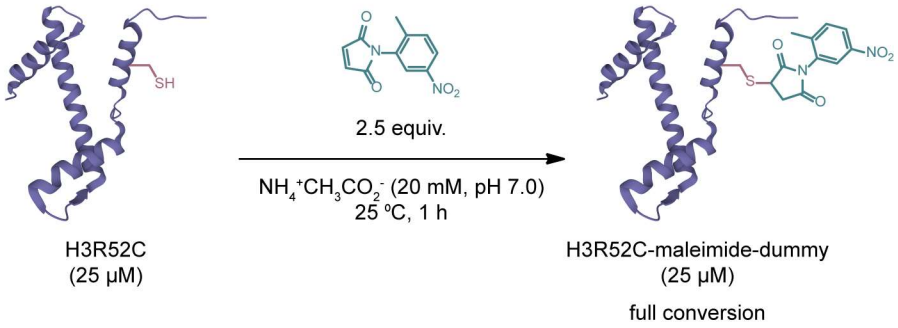
2.6 Control reactions (1) with a maleimide-dummy

2.6.1 Conjugation of H3K4C, H3R52C and H3K4CR52C with maleimide-dummy

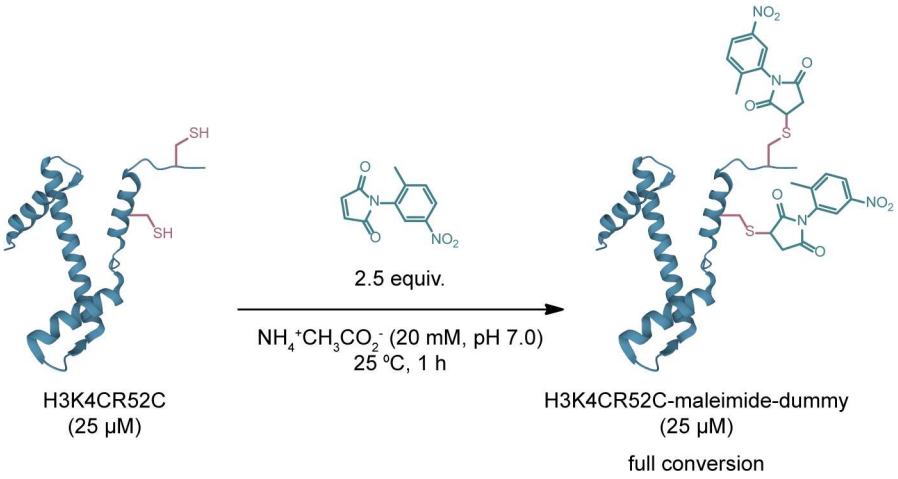
To demonstrate that the SPAAC reaction is essential in promoting acetylation at the nearest lysine residue, the H3K4C, H3R52C and H3K4CR52C proteins were also conjugated to a maleimide reagent without a SPAAC clickable handle. For this purpose, N-(5-nitro-ortho-tolyl)-maleimide was used since it was readily available in the laboratory and will be henceforth referred to as maleimide-dummy. Aliquots of histones H3K4C, H3R52C and H3K4CR52C were diluted to 25 µM with ammonium acetate buffer (20 mM, pH 7.0). An aliquot of 2.5 equiv. of maleimide-dummy was added to each mixture and the reactions were shaken at 400 rpm and 25 ºC. After 1 h, the reaction mixtures were analyzed by LC-MS and complete conversion to the H3K4C, H3R52C and H3K4CR52C proteins containing the maleimide-dummy was observed (H3K4C-maleimide-dummy: calculated mass = 15446 Da; observed mass = 15450 Da; H3R52C-maleimide-dummy: calculated mass = 15418 Da; observed mass = 15415 Da; H3K4CR52C-maleimide-dummy: calculated mass = 15625 Da; observed mass = 15632 Da). A schematic representation of these reactions is shown in **Scheme 20**, **Scheme 21** and **Scheme 22**. The resulting deconvoluted ESI-MS spectra are shown in **Figure 74**, **Figure 75** and **Figure 76**. The independent aliquots of H3K4C-maleimide-dummy, H3R52C-maleimide-dummy and H3K4CR52C-maleimide-dummy were desalted and concentrated using Amicon centrifugal filter units (Merck Millipore). Protein concentrations were then determined by the Bradford assay.



Scheme 20. Site-selective modification of the H3K4C protein with the maleimide-dummy reagent.



Scheme 21. Site-selective modification of the H3R52C protein with the maleimide-dummy reagent.



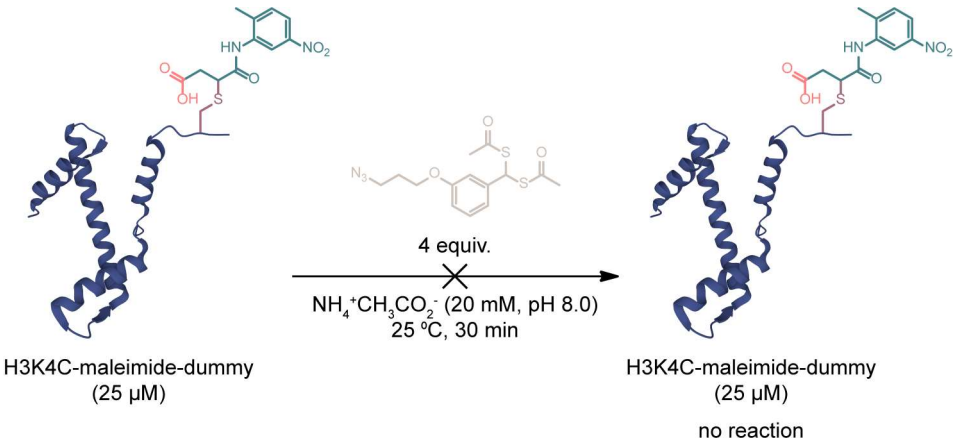
Scheme 22. Site-selective modification of the H3K4CR52C protein with the maleimide-dummy reagent.

2.6.2 Treatment of H3K4C, H3R52C and H3K4CR52C proteins conjugated to maleimide-dummy with compound 4

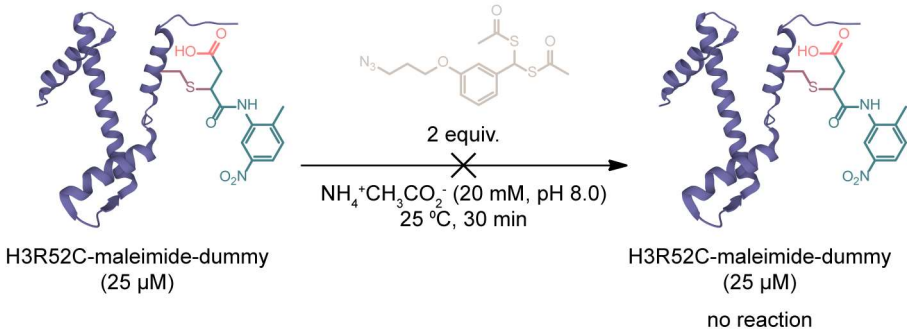
Subsequently, these modified proteins were incubated with compound **4** at a basic pH using the same reaction parameters described previously for the SPAAC reaction. Thus,

independent aliquots of H3K4C, H3R52C and H3K4CR52C modified with maleimide-dummy were diluted to 25 μM in ammonium acetate buffer (20 mM, pH 8.0). The resulting deconvoluted ESI-MS spectra of the diluted H3K4C-maleimide-dummy, H3R52C-maleimide-dummy and H3K4CR52C-maleimide-dummy proteins in ammonium acetate buffer (20 mM, pH 8.0) are shown in **Figure 77**, **Figure 78** and **Figure 79**. The mass difference observed in these spectra compared to those in **Figure 74**, **Figure 75** and **Figure 76** likely results from maleimide hydrolysis promoted by the strong electron withdrawing nature of the nitro group (H3K4C-hydrolyzed-maleimide-dummy: calculated mass = 15464 Da; observed mass = 15458 Da; H3R52C-hydrolyzed-maleimide-dummy: calculated = 15436 Da; observed = 15430 Da; H3K4CR52C-hydrolyzed-maleimide-dummy: calculated mass = 15661 Da; observed mass = 15663 Da).

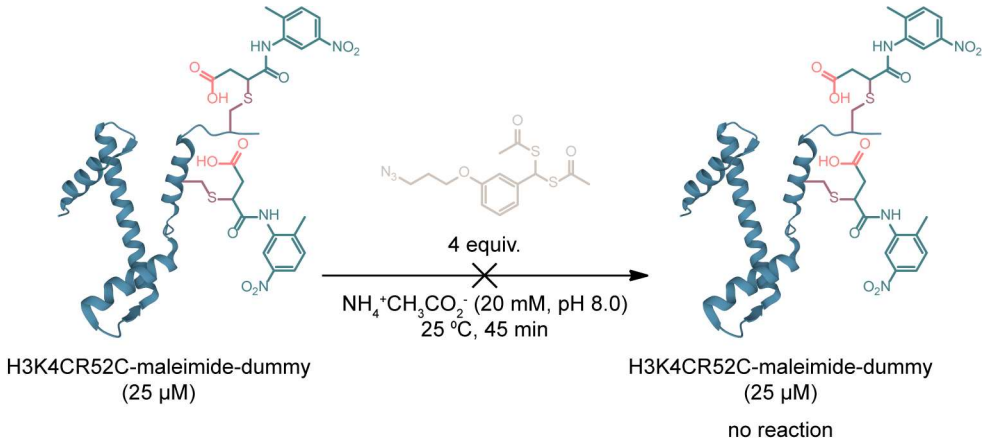
An aliquot of 4 equiv. of compound **4** was added to the diluted H3K4C-maleimide-dummy solution, while an aliquot of 2 equiv. of the same compound was added to the solution containing the diluted H3R52C-maleimide-dummy protein. The reactions were shaken at 400 rpm and 25 ºC. After 30 min, the reaction mixtures were analyzed by LC-MS and acetylation was not detected in any of the two modified proteins (H3K4C-hydrolyzed-maleimide-dummy: calculated mass = 15464 Da; observed mass = 15458 Da; H3R52C-hydrolyzed-maleimide-dummy: calculated mass = 15436 Da; observed mass = 15430 Da). An aliquot of 4 equiv. of compound **4** was added to the diluted H3K4CR52C-maleimide-dummy solution and the reaction was shaken at 400 rpm and 25 ºC. After 45 min, the reaction mixture was analyzed by LC-MS and acetylation was also not observed (H3K4CR52C-hydrolyzed-maleimide-dummy: calculated mass = 15661 Da; observed mass = 15664 Da). A schematic representation of these reactions is shown in **Scheme 23, Scheme 24** and **Scheme 25**. The resulting deconvoluted ESI-MS spectra for the H3K4C-hydrolyzed-maleimide-dummy, H3R52C-hydrolyzed-maleimide-dummy and H3K4CR52C-hydrolyzed-maleimide-dummy proteins are shown in **Figure 80**, **Figure 81** and **Figure 82**, respectively.



Scheme 23. Acetylation of H3K4C-maleimide-dummy does not occur following incubation with compound **4** due to the lack of a SPAAC counterpart in this protein.



Scheme 24. Acetylation of H3R52C-maleimide-dummy does not occur following incubation with compound **4** due to the lack of a SPAAC counterpart in this protein.

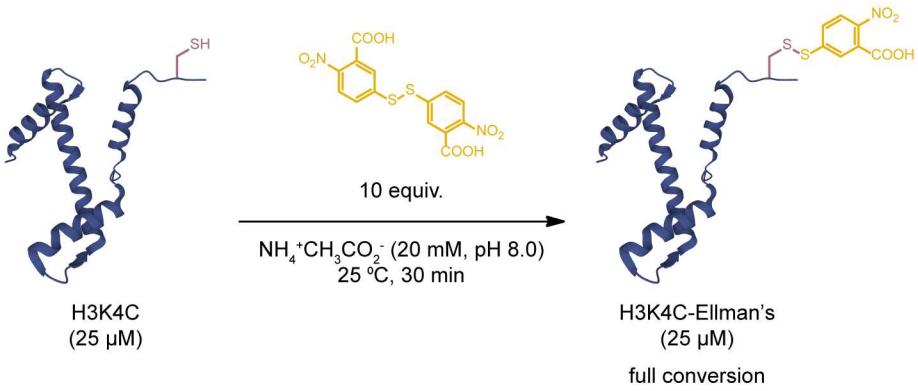


Scheme 25. Acetylation of H3K4CR52C-maleimide-dummy does not occur following incubation with compound **4** due to the lack of a SPAAC counterpart in this protein.

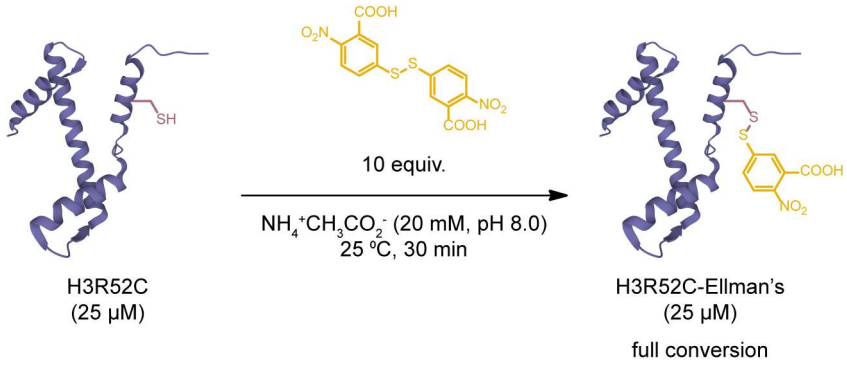
2.7 Control reactions (2) with Ellman's reagent to confirm cysteine availability before and after maleimide conjugation

2.7.1 Evaluation of cysteine availability before maleimide conjugation

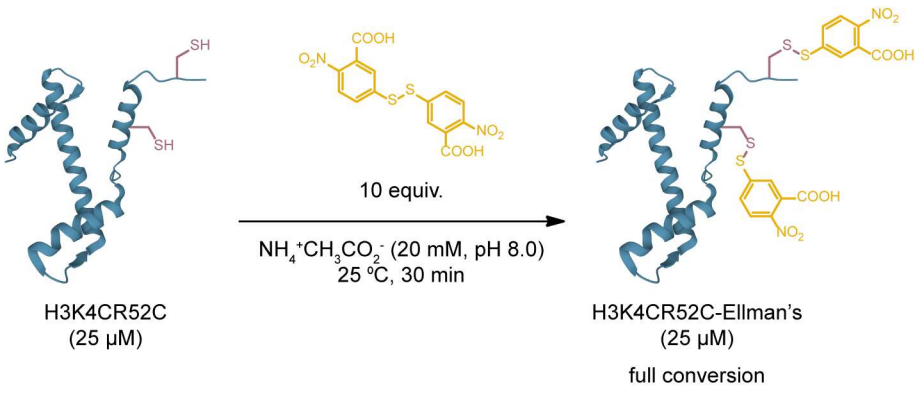
Independent aliquots of reduced H3K4C, H3R52C and H3K4CR52C were diluted to 25 µM with ammonium acetate buffer (20 mM, pH 8.0). Then, 10 equiv. of Ellman's reagent were added to each mixture and the reactions were shaken at 400 rpm and 25 °C. After 30 min, the reaction mixtures were analyzed by LC-MS and complete conversion to the respective Ellman disulfides of H3K4C (calculated mass = 15411 Da; observed mass = 15409 Da), H3R52C (calculated mass = 15383 Da; observed mass = 15387 Da) and H3K4CR52C (calculated mass = 15555 Da; observed mass = 15560 Da) was observed. A schematic representation of these reactions is shown in **Scheme 26, Scheme 27** and **Scheme 28**. The resulting deconvoluted ESI-MS spectra are shown in **Figure 83**, **Figure 84** and **Figure 85**, respectively.



Scheme 26. Site-selective modification of the H3K4C protein with Ellman's reagent.



Scheme 27. Site-selective modification of the H3R52C protein with Ellman's reagent.

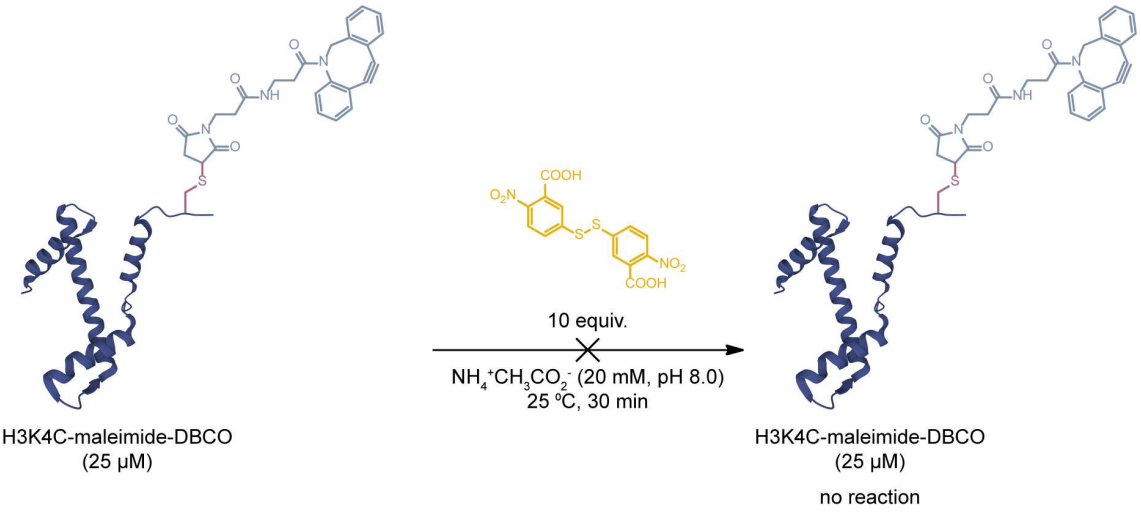


Scheme 28. Site-selective modification of the H3K4CR52C protein with Ellman's reagent.

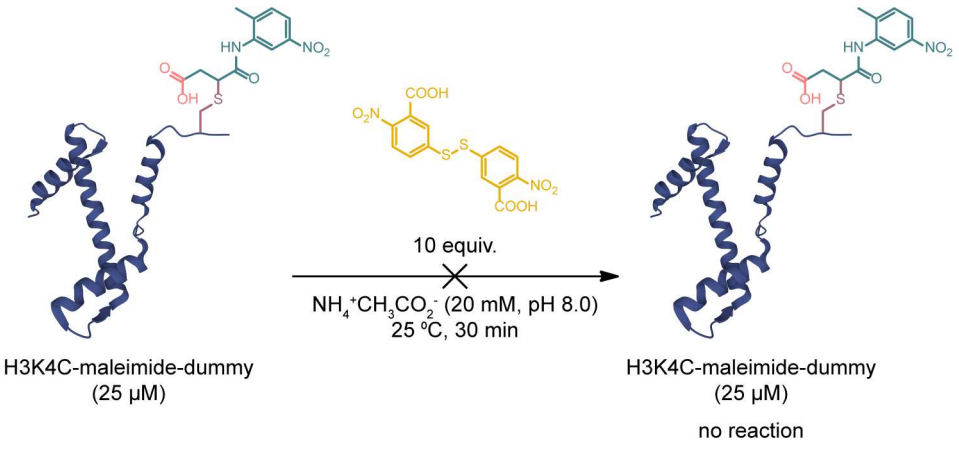
2.7.2 Evaluation of cysteine availability after maleimide conjugation

Independent aliquots of the H3K4C, H3R52C and H3K4CR52C proteins conjugated to the two maleimide reagents described in this thesis (maleimide-DBCO and maleimide-dummy) were diluted to 25 µM with ammonium acetate buffer (20 mM, pH 8.0). Then, 10 equiv. of Ellman's reagent were added to each mixture and the reactions were shaken at 400 rpm and 25 ºC. After 30 min, the reaction mixtures were analyzed by LC-MS and conversion to the corresponding Ellman disulfides was not observed, indicating that maleimide conjugation had occurred at the cysteine residues of H3K4C (H3K4C-maleimide-DBCO: calculated mass = 15641 Da; observed

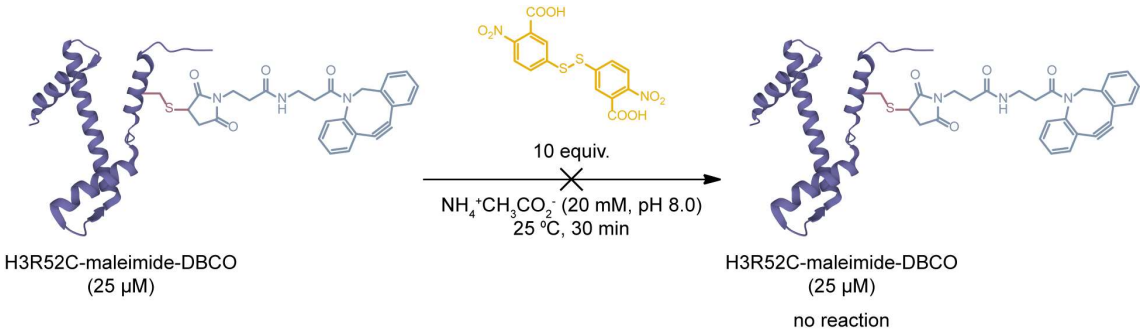
mass= 15640 Da; H3K4C-hydrolyzed-maleimide-dummy: calculated mass = 15464 Da; observed mass = 15459 Da), H3R52C (H3R52C-maleimide-DBCO: calculated mass = 15613 Da; observed mass = 15611 Da; H3R52C-hydrolyzed-maleimide-dummy: calculated mass = 15436 Da; observed mass = 15430 Da) and H3K4CR52C (H3K4CR52C-maleimide-DBCO: calculated mass = 16016 Da; observed mass = 16021 Da; H3K4CR52C-hydrolyzed-maleimide-dummy: calculated mass = 15661 Da; observed mass = 15664 Da). A schematic representation of these reactions is shown in **Scheme 29, Scheme 30, Scheme 31, Scheme 32, Scheme 33** and **Scheme 34**. The resulting deconvoluted ESI-MS spectra are shown in **Figure 86**, **Figure 87**, **Figure 88**, **Figure 89**, **Figure 90** and **Figure 91**.



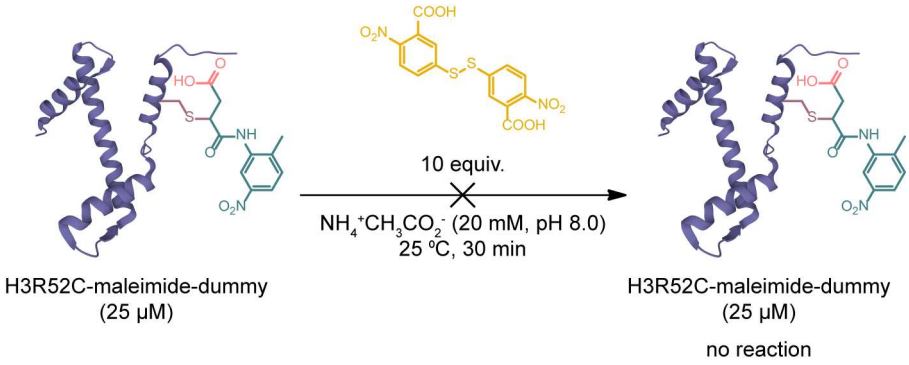
Scheme 29. Site-selective modification with Ellman's reagent does not occur following conjugation of the H3K4C protein with maleimide-DBCO.



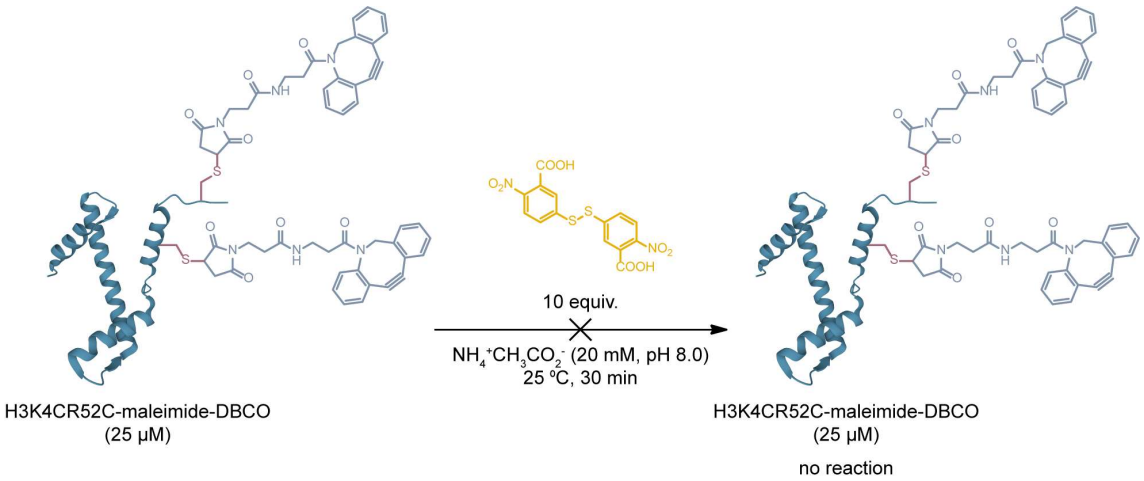
Scheme 30. Site-selective modification with Ellman's reagent does not occur following conjugation of the H3K4C protein with maleimide-dummy.



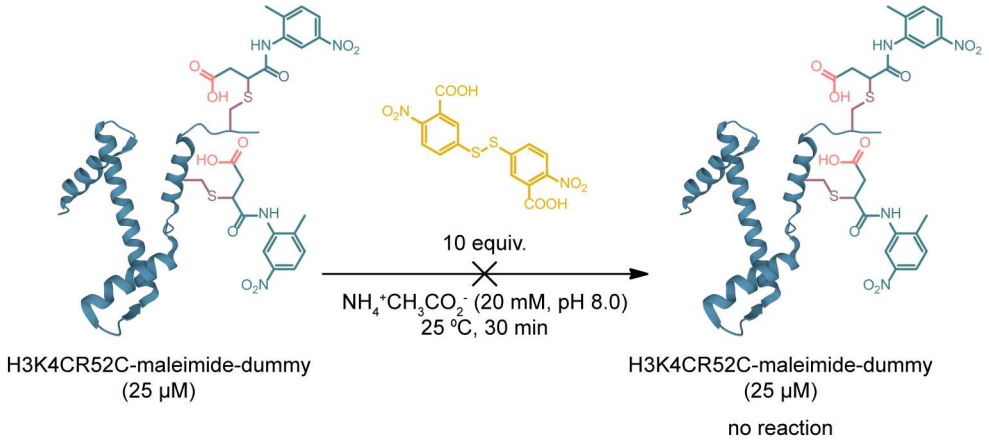
Scheme 31. Site-selective modification with Ellman's reagent does not occur following conjugation of the H3R52C protein with maleimide-DBCO.



Scheme 32. Site-selective modification with Ellman's reagent does not occur following conjugation of the H3R52C protein with maleimide-dummy.



Scheme 33. Site-selective modification with Ellman's reagent does not occur following conjugation of the H3K4CR52C protein with maleimide-DBCO.



Scheme 34. Site-selective modification with Ellman's reagent does not occur following conjugation of the H3K4CR52C protein with maleimide-dummy.

2.8 Control reactions (3) using the ubiquitin UbK63C mutant protein

The ubiquitin mutant UbK63C protein was expressed and purified as previously described and was a kind gift of Dr. Ester Jiménez-Moreno.[285] The amino acid sequence of UbK63C is shown below and it was used to compute the expected mass of this protein using ExPASy.[284] This was essential to determine whether the UbK63C sample was suitable for conjugation reactions and later evaluate if this protein was successfully modified by comparing the expected mass with the observed one in MS experiments.

Ubiquitin UbK63C

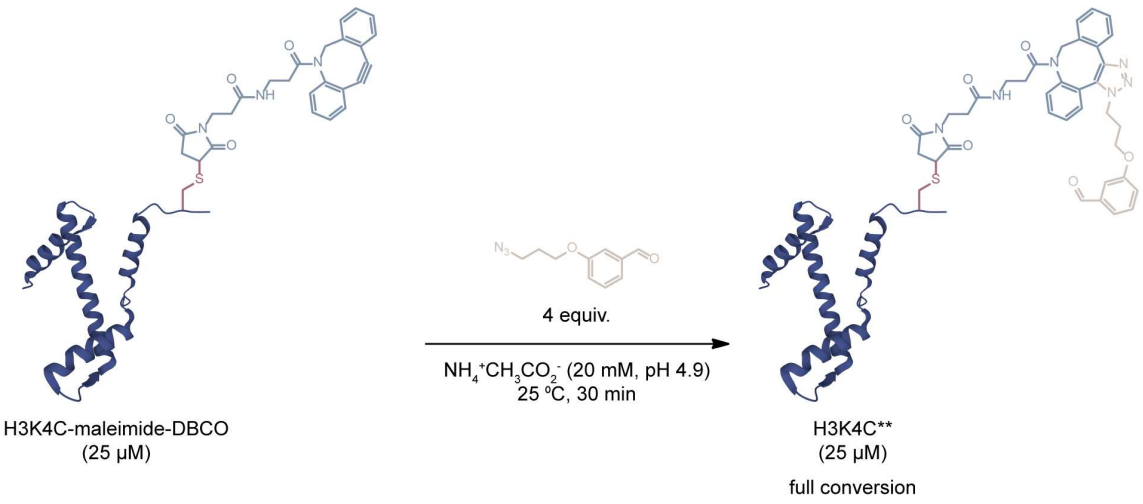
SAQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNI QCESTLHLVLRLRGG

UbK63C was diluted to 50 μM with ammonium acetate buffer (20 mM, pH 8.0). Then, an aliquot of 4 equiv. of TCEP was added to the mixture and the reaction was shaken at 400 rpm and 25°C. After 30 min, the reaction mixture was added to a solution containing 0.12 mg of maleimide-DBCO dissolved in 1 μL of DMF. An aliquot of the reaction mixture was centrifuged and analyzed by LC-MS. Full conversion to the UbK63C protein containing the maleimide-DBCO clickable handle was observed (calculated mass = 8994 Da; observed mass = 8995 Da). The resulting deconvoluted ESI-MS spectrum is shown in **Figure 93**. Following centrifugation, the modified UbK63C-maleimide-DBCO protein was purified by size-exclusion chromatography using a PD MiniTrap G-25 column (Cytiva). Concentration of the modified UbK63C with the maleimide-DBCO clickable handle was determined by Nanodrop.

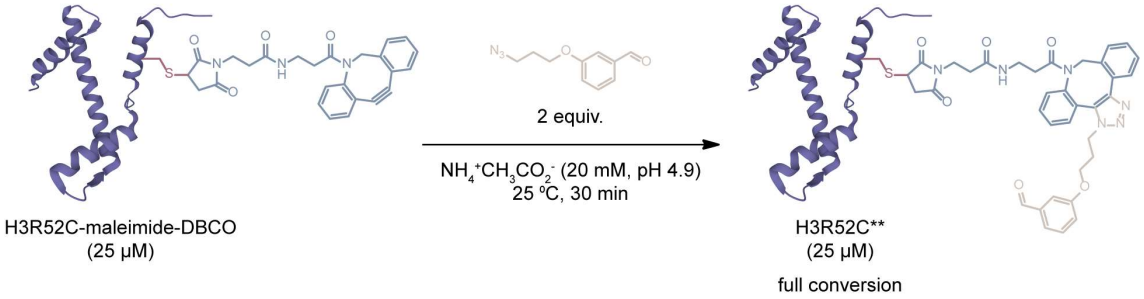
An aliquot of the UbK63C protein containing the maleimide-DBCO clickable handle was diluted to 25 μM in ammonium acetate buffer (20 mM, pH 8.0). An aliquot of 1 equiv. of compound **4** was then added to the diluted UbK63C-maleimide-DBCO solution. The reaction was shaken at 400 rpm and 25 °C. After 30 min, the reaction mixtures were analyzed by LC-MS and complete conversion to the respective unacetylated UbK63C SPAAC-containing product with the *gem*-dithioacetate group still left to react was observed (unAcUbK63C**: calculated mass = 9333 Da; observed mass = 9336 Da). The resulting deconvoluted ESI-MS spectrum is shown in **Figure 94**.

2.9 General procedure to generate unacetylated histones containing the SPAAC by-product at the cysteine

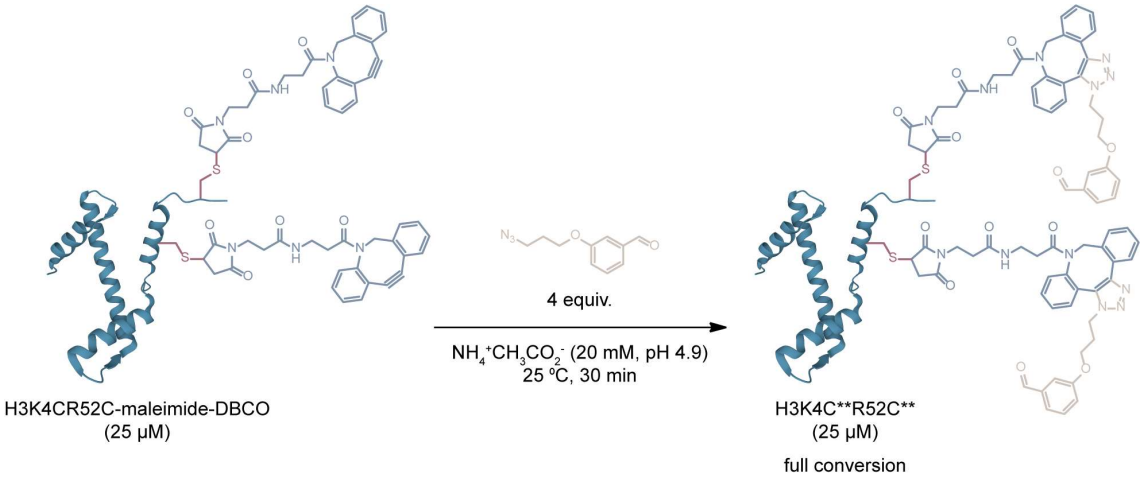
To serve as control proteins in further studies, unacetylated H3K4C, H3R52C and H3K4CR52C histones bearing the SPAAC by-product that remains attached at the cysteine were also prepared. Thus, H3K4C, H3R52C and H3K4CR52C proteins were reduced and conjugated to maleimide-DBCO according to the protocols already described. Subsequently, the SPAAC reaction was performed by incubating the modified proteins with compound **2** using the same reaction parameters reported previously for compound **4**. More specifically, independent aliquots of the H3K4C, H3R52C and H3K4CR52C proteins containing the maleimide-DBCO clickable handle were diluted to 25 µM in ammonium acetate buffer (20 mM, pH 4.9). An aliquot of 4 equiv. of compound **2** was added to the diluted H3K4C-maleimide-DBCO and H3K4CR52C-maleimide-DBCO solutions, while an aliquot of 2 equiv. of the same compound was added to the solution containing the diluted H3R52C-maleimide-DBCO protein. The reactions were shaken at 400 rpm and 25 ºC. After 30 min, the reaction mixtures were analyzed by LC-MS and complete conversion to the respective H3K4C**, H3R52C** and H3K4C**R52C** proteins containing the SPAAC by-product was observed (H3K4C**: calculated mass = 15847 Da; observed mass = 15842 Da; H3R52C**: calculated mass = 15819 Da; observed mass = 15816 Da; H3K4C**R52C**: calculated mass = 16426 Da; observed mass = 16431 Da). A schematic representation of these reactions is shown in **Scheme 35, Scheme 36** and **Scheme 37**. The resulting deconvoluted ESI-MS spectra are shown in **Figure 95**, **Figure 96** and **Figure 97**. The proteins were buffer-exchanged to ammonium acetate buffer (20 mM, pH 5.8), desalted into water and concentrated using Amicon centrifugal filter units (Merck Millipore). Protein concentrations were then determined by the Bradford assay. Finally, the aliquots of H3K4C**, H3R52C** and H3K4C**R52C** were lyophilized and frozen at -80 ºC for further studies.



Scheme 35. SPAAC reaction between the DBCO moiety attached to the H3K4C-maleimide-DBCO protein and the azide counterpart of compound **2**.



Scheme 36. SPAAC reaction between the DBCO moiety attached to the H3R52C-maleimide-DBCO protein and the azide counterpart of compound **2**.



Scheme 37. SPAAC reaction between the DBCO moiety attached to the H3K4CR52C-maleimide-DBCO protein and the azide counterpart of compound **2**.

2.10 Confirmation of site-specific acetylation in the H3K4C**K9Ac protein by tandem mass spectrometry

MS/MS was performed in collaboration with Professor Carlos Cordeiro (Faculdade de Ciências, Universidade de Lisboa) to confirm that acetylation was site-specific for position K9 of the modified H3K4C protein. To this end, a solution of the H3K4C**K9Ac protein was diluted to 5 μM in a 50:50 mixture of water:acetonitrile with 0.1% formic acid and injected. Data was acquired in a high-resolution Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer, the SolariX XR, at 7 Tesla (25 1MB transients added per spectrum). For MS/MS spectra, ions were accumulated for 2 s in the external collision cell before collision-induced dissociation (CID) using Ar (15V) or electron-capture dissociation (ECD, cathode current 1.5 A, 0.02 s pulse). Data was processed in Data Analysis 5.0 and monoisotopic masses calculated with the SNAP algorithm (Bruker Daltonics). MS/MS data analysis was performed with Prosight Lite (The Kelleher Research Group, Pacific Northwest University).

2.11 Circular dichroism spectroscopy experiments

Circular dichroism (CD) spectroscopy was performed to establish (1) whether the H3K4C protein maintains its overall secondary structure following cysteine conjugation and subsequent

SPAAC reaction, and (2) if differences exist in the secondary structure of the various H3 mutants generated in this work. To this end, lyophilized H3K4C, H3K4C**, H3K4C**K9Ac, H3R52C and H3K4CR52C proteins were resuspended in Milli-Q water and quantified using the Bradford assay. Aliquots of these proteins were diluted in filtered sodium phosphate buffer (20 mM, pH 7.4) immediately prior to CD analysis. To answer the first question, aliquots of H3K4C, H3K4C** and H3K4C**K9Ac were diluted to 20 μM. Aliquots of H3K4C, H3R52C and H3K4CR52C were diluted to 10 μM to answer the second question.

Measurements were recorded using a JASCO spectropolarimeter J-815 (Tokyo, Japan) equipped with a PTC-423S/15 Peltier temperature control unit (at 25 °C or 20 °C for samples belonging to the first or second question, respectively). The data was acquired at a scanning speed of 200 nm per min in a quartz cell with a 0.1 cm path length. The data integration time was set to 2 s, the data pitch to 0.5 nm and the bandwidth to 1 nm. Spectra were averaged over five scans, in a wavelength range from 190 nm to 300 nm. The spectrum from a blank sample containing only buffer was subtracted from the averaged ellipticity data.

CD raw data in ellipticity (θ) was normalized to the molar concentration of the samples (molar ellipticity, $[\theta]_{molar,\lambda}$) and number of protein residues (mean residue ellipticity, $[\theta]_{mrw,\lambda}$), according to the following equations:

$$[\theta]_{molar,\lambda} = 100 \times \frac{\theta_\lambda}{m \times d}$$

$$[\theta]_{mrw,\lambda} = MRW \times \frac{\theta_\lambda}{10 \times d \times c}$$

where θ_λ is the observed ellipticity (in degrees) at wavelength λ, *m* is the molar concentration, *d* is the pathlength (cm) and *c* is the concentration (g/ml). *MRW* represents the mean residue weight and is equal to $M/(N-1)$, where *M* is the molecular weight (in Da) of the protein and *N* is the number of residues.[286]

2.12 Western blot analysis of K9 and K56 acetylation in the chemically modified histones

Western blot experiments were performed to establish whether the chemically acetylated H3K4C**K9Ac, H3R52C**K56Ac and H3K4C**K9AcR52C**K56Ac proteins can be recognized by antibodies raised against the natural acetyl-lysine at positions 9 and 56 of histone H3.

Sample preparation and SDS-PAGE

Lyophilized proteins were resuspended in Milli-Q water and quantified using the Bradford assay. Separate SDS-PAGE gels were prepared for the Coomassie staining and Western Blot experiments indicated in **Table 1** and **Table 2**. Protein aliquots were diluted in Tris-HCl (20 mM, pH 7.5) and 5 μL of each was mixed with equal volume of 2x Laemmli sample buffer. BME (0.5 μL) was added as the reducing agent and samples were boiled at 96 °C for 4 min. Samples were loaded into a 15% SDS-PAGE gel, together with the molecular weight marker (PageRuler™ Plus

Prestained Protein Ladder, 10 to 250 kDa, ThermoFisher). Gels for both Coomassie staining and Western blot experiments were run in Tris-glycine-SDS buffer.

Table 1. Details of the Coomassie staining experiments.

Experiment	Name of the Protein	Concentration (ng/µL)	Amount (µg)	Volume (µL)
1	H3K4C	400	2	5
	H3K4C**			
	H3K4C**K9Ac			
2	H3R52C	400	2	5
	H3R52C**			
	H3R52C**K56Ac			
3	H3K4CR52C	400	2	5
	H3K4C**R52C**			
	H3K4C**K56AcR52C**K56Ac			

Table 2. Details of the Western blot experiments to evaluate K9 and K56 acetylation following the site-specific strategy developed in this thesis.

Experiment	Name of the Protein	Target	Concentration (ng/µL)	Amount (µg)	Volume (µL)
1	H3K4C	H3K9Ac	60	0.3	5
	H3K4C**				
	H3K4C**K9Ac				
2	H3R52C	H3K56Ac	400	2	5
	H3R52C**				
	H3R52C**K56Ac				
3	H3K4CR52C	H3K9Ac	60	0.3	5
	H3K4C**R52C**				
	H3K4C**K56AcR52C**K56Ac				
4	H3K4CR52C	H3K56Ac	400	2	5
	H3K4C**R52C**				
	H3K4C**K56AcR52C**K56Ac				
5	H3K4C**K9Ac	H3K9Ac	60	0.3	5
	H3R52C**K56Ac				

Coomassie staining

Following SDS-PAGE, gels destined for Coomassie staining were fixed for 30 min with 40% methanol in Milli-Q water. These were subsequently stained with InstantBlue Coomassie Stain (Abcam) for 1 h. Excess staining was removed with Milli-Q water and the gel was imaged with Amersham Imager 680 RGB (GE Healthcare).

Membrane transfer and probing with the anti-H3K9Ac and anti-H3K56Ac antibodies

Following SDS-PAGE, the proteins in the gels destined for Western blot experiments were transferred to membranes (Amersham Protran Supported Nitrocellulose 0.2 µm) in wet conditions with 20% methanol under constant current for 2 h at 200 mA. The membranes were then washed (1 x 5 min) in Tris-buffered saline (TBS) (20 mM Tris-HCl, 150 mM NaCl, pH 7.5) and blocked with 5% non-fat dry milk in TBS with 0.05% Tween-20 (TBST-0.05) for 1 h at room temperature. Subsequently, the membranes were washed with TBST-0.05 (3 x 5 min) and incubated overnight at 4 ºC with either anti-H3K9Ac or anti-H3K56Ac primary antibody diluted in 5% BSA in TBST-0.05, according to the experimental design indicated in **Table 2**. After washing with TBST-0.05 (5 x 5 min), the membranes were subsequently incubated at room temperature for 1 h with HRP-conjugated anti-rabbit IgG (H+L) secondary antibody diluted in 5% non-fat dry milk in TBST-0.05. The information pertaining to the primary and secondary antibodies used in Western blot experiments before membrane re-probing and their respective working dilutions are indicated in **Table 3**, **Table 4** and **Table 5**. After washing with TBST-0.05 (5 x 5 min) and TBS (1 x 5 min), the membranes were developed using the Clarity Western ECL Substrate (Bio-Rad) and imaged with Amersham Imager 680 RGB (GE Healthcare).

Table 3. List of primary antibodies used in Western blot experiments.

Primary Antibody	Clonality	Source	Brand
Anti-H3K9Ac (C5B11), cat. # 9649S	Monoclonal	Rabbit IgG	Cell Signaling Technologies
Anti-H3K56Ac, cat. #4243S	Polyclonal	Rabbit	Cell Signaling Technologies
Anti-H3 (1B1B2), cat. #14269S	Monoclonal	Mouse IgG3	Cell Signaling Technologies

Table 4. List of secondary antibodies used in Western blot experiments.

Secondary Antibody	Clonality	Source	Brand
HRP-conjugated anti-rabbit IgG (H+L)	Polyclonal	Goat IgG	Invitrogen
HRP-conjugated anti-mouse IgG (H+L)	Polyclonal	Goat IgG	Invitrogen

Table 5. Combinations of primary and secondary antibodies used in Western blot experiments to probe K9 and K56 acetylation before membrane re-probing.

Target	Primary Antibody	Primary Antibody Dilution	Secondary Antibody	Secondary Antibody Dilution
H3K9Ac	Anti-H3K9Ac	1:6,000 (5% BSA in TBST-0.05)	HRP-conjugated anti-rabbit IgG (H+L)	1:10,000 (5% milk in TBST-0.05)
H3K56Ac	Anti-H3K56Ac	1:1,000 (5% BSA in TBST-0.05)	HRP-conjugated anti-rabbit IgG (H+L)	1:3,000 (5% milk in TBST-0.05)

Membrane stripping and re-probing with the anti-H3 antibody

Following imaging, the membranes were re-probed with the H3 antibody. Membrane stripping (2 x 5 min) was performed with an acidic buffer (200 mM glycine, 0.1% (w/v) SDS, 1%

(v/v) Tween-20, pH 2.2 using HCl). After washing with phosphate-buffered saline (PBS) (2 x 10 min) and TBS (2 x 5 min), the membranes were blocked with 5% non-fat dry milk in TBST-0.05 for 1 h at room temperature. The membranes were then washed with TBST-0.05 (3 x 5 min) and incubated overnight at 4 ºC with anti-H3 primary antibody diluted in 5% BSA in TBST-0.05, according to the experimental design indicated in **Table 2**. After washing with TBS containing 0.1% Tween-20 (TBST-0.1) (6 x 5 min), the membranes were incubated at room temperature for 1 h with HRP-conjugated anti-mouse IgG (H+L) secondary antibody diluted in 5% non-fat dry milk in TBST-0.1. After washing with TBST-0.1 (7 x 5 min) and TBS (1 x 5 min), the membranes were developed using the Clarity Western ECL Substrate (Bio-Rad) and imaged with Amersham Imager 680 RGB (GE Healthcare). The information pertaining to the primary and secondary antibodies used in Western blot experiments after membrane re-probing and their respective working dilutions are indicated in **Table 3**, **Table 4** and **Table 6**.

Table 6. Combinations of primary and secondary antibodies used in Western blot experiments after membrane re-probing.

Prior Target	Primary Antibody	Primary Antibody Dilution	Secondary Antibody	Secondary Antibody Dilution
H3K9Ac	Anti-H3	1:1,000 (5% BSA in TBST-0.05)	HRP-conjugated anti-mouse IgG (H+L)	1:5,000 (5% milk in TBST-0.1)
H3K56Ac	Anti-H3	1:10,000 (5% BSA in TBST-0.05)	HRP-conjugated anti-mouse IgG (H+L)	1:10,000 (5% milk in TBST-0.1)

2.13 Development of an ELISA protocol for the detection and quantitation of H3K9 acetylation

In addition to Western blot analysis, an enzyme-linked immunosorbent assay (ELISA) was also performed to detect K9 acetylation in the H3K4C**K9Ac protein. This two-day assay was developed in-house using the same antibodies as those previously described for Western blot and ultimately served to quantitate the level of K9 acetylation following incubation with sirtuin-3 (Sirt3).

Lyophilized H3K4C, H3K4C** and H3K4C**K9Ac proteins were resuspended in Milli-Q water and quantified using the Bradford assay. Two-fold serial dilutions of each protein were made from 8 µg/mL to 0.5 µg/mL in carbonate buffer (100 mM, pH 9.6). A Nunc MaxiSorp 96 well ELISA plate was coated with duplicate serial dilutions of H3K4C, H3K4C** and H3K4C**K9Ac and incubated overnight at 4 ºC. After washing the plate (3 x 5 min) with PBS containing 0.05% Tween-20 (PBST-0.05), blocking buffer (200 µL per well) was added to each well (3% and 5% milk in PBST-0.05 for anti-H3 and anti-H3K9Ac probing, respectively). After incubating for 2 h at room temperature, the plate was washed (1 x 5 min). Primary antibody dilutions (100 µL per well) of anti-H3 and anti-H3K9Ac were then added. Following incubation at room temperature for 2 h, the plate was washed with PBST-0.05 (3 x 5 min) and secondary antibody dilutions (100 µL per well) were added. The information pertaining to the primary and secondary antibodies used in ELISA experiments, as well as their respective combinations and working dilutions are indicated in **Table**

7, **Table 8** and **Table 9**. After incubating for 1 h at room temperature, the plate was washed (5 x 5 min) with PBST-0.05. TMB substrate (90 µL per well, Invitrogen) was added and the plate was left to incubate at room temperature. After sufficient color development, the colorimetric reaction was stopped with a 1M sulfuric acid solution (50 µL per well). Absorbance at 450 nm was then recorded on a Tecan Infinite M200 plate reader.

ELISA results were analyzed using GraphPad Prism software (version 6.01) by fitting the data of each combination of target protein (H3K4C, H3K4C** and H3K4C**K9Ac) and probe (anti-H3 and anti-H3K9Ac) to a four-parameter logistic regression of a standard curve, with the logarithm of the concentration plotted on the x-axis and the absorbance plotted on the y-axis.

Table 7. List of primary antibodies used in ELISA experiments.

Primary Antibody	Clonality	Source	Brand
Anti-H3K9Ac (C5B11), cat. # 9649S	Monoclonal	Rabbit IgG	Cell Signaling Technologies
Anti-H3 (1B1B2), cat. #14269S	Monoclonal	Mouse IgG3	Cell Signaling Technologies

Table 8. List of secondary antibodies used in ELISA experiments.

Secondary Antibody	Clonality	Source	Brand
HRP-conjugated anti-rabbit IgG (H+L)	Polyclonal	Goat IgG	Invitrogen
HRP-conjugated anti-mouse IgG (H+L)	Polyclonal	Goat IgG	Invitrogen

Table 9. Combinations of primary and secondary antibodies used in ELISA experiments.

Target	Primary Antibody	Primary Antibody Dilution	Secondary Antibody	Secondary Antibody Dilution
H3K9Ac	Anti-H3K9Ac	1:2,000 (5% BSA in PBST-0.05)	HRP-conjugated anti-rabbit IgG (H+L)	1:4,000 (5% milk in TBST-0.05)
H3	Anti-H3	1:2,000 (3% BSA in PBST-0.05)	HRP-conjugated anti-mouse IgG (H+L)	1:2,000 (3% milk in TBST-0.05)

2.14 Deacetylation assays of H3K4C**K9Ac and H3R52C**K56Ac with Sirt3

To establish whether the H3K4C**K9Ac and H3R52C**K56Ac proteins can be recognized by histone-modifying enzymes, these chemically acetylated histones were used as substrates in NAD^+-dependent deacetylation assays with Sirt3.

Lyophilized H3K4C**K9Ac and H3R52C**K56Ac proteins were resuspended in Milli-Q water and quantified using the Bradford assay. The control and reaction samples were then prepared with a final volume of 45 µL for H3K4C**K9Ac and 20 µL for H3R52C**K56Ac. For both control and reaction samples, the 10x deacetylation buffer (250 mM Tris, 1.37 M NaCl, 27 mM KCl, 10 mM $MgCl_2$, pH 8.0) was diluted 1:10 in Milli-Q water. DTT (90 mM in Milli-Q water) and NAD^+ (50 mM in 50 mM Tris-HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$ from Enzo Life Sciences) were added to a final concentration of 1 mM and 3.5 mM, respectively. Aliquots of

H3K4C**K9Ac and H3R52C**K56Ac were added at a final concentration of 60 ng/µL and 200 ng/µL, respectively, to their corresponding control and reaction samples. For the reaction samples, a 0.5 µL aliquot containing 20 units of recombinant Sirt3 (1 µg/µL in 25 mM Tris pH 7.5, 100 mM NaCl, 5 mM DTT, 10% glycerol from Enzo Life Sciences) was also added. In the case of the deacetylation assay of H3K4C**K9Ac, the control and reaction samples were incubated at 37 ºC and 400 rpm for 2 h. For the deacetylation assay of H3R52C**K56Ac, the control and reaction samples were incubated at 37 ºC and 400 rpm for 4 h. Following incubation, samples were spun down and placed on ice. K9 and K56 deacetylation were subsequently evaluated by Western blot. In addition, K9 deacetylation was also evaluated by ELISA.

Western blot analysis to evaluate K9 and K56 deacetylation

Sample preparation and SDS-PAGE

Two separate SDS-PAGE gels were prepared for the Western blot experiments indicated in **Table 10**. Protein aliquots from the deacetylation experiments were mixed with equal volume of 2x Laemmli sample buffer. BME (0.5 µL) was added as the reducing agent and samples were boiled at 96 ºC for 4 min. Samples were loaded into a 15% SDS-PAGE gel, together with the molecular weight marker (PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, ThermoFisher). The two gels were run in Tris-glycine-SDS buffer.

Table 10. Details of the Western blot experiments to evaluate K9 and K56 deacetylation following incubation of the H3K4C**K9Ac and H3R52C**K56Ac proteins with Sirt3.

Experiment	Name of the Protein	Sample	Target	Amount (µg)	Volume (µL)
1	H3K4C**K9Ac	Control	H3K9Ac	0.3	5
		Reaction			
2	H3R52C**K56Ac	Control	H3K56Ac	2	10
		Reaction			

Membrane transfer and probing with the anti-H3K9Ac and anti-H3K56Ac antibodies

Following SDS-PAGE, the proteins in the gels were transferred to membranes (Amersham Protran Supported Nitrocellulose 0.2 µm) in wet conditions with 20% methanol under constant current for 2 h at 200 mA. The membranes were then washed (1 x 5 min) in TBS and blocked with 5% non-fat dry milk in TBST-0.05 for 1 h at room temperature. Subsequently, the membranes were washed with TBST-0.05 (3 x 5 min) and incubated overnight at 4 ºC with either anti-H3K9Ac or anti-H3K56Ac primary antibody diluted in 5% BSA in TBST-0.05, according to the experimental design indicated in **Table 10**. After washing with TBST-0.05 (5 x 5 min), the membranes were subsequently incubated at room temperature for 1 h with HRP-conjugated anti-rabbit IgG (H+L) secondary antibody diluted in 5% non-fat dry milk in TBST-0.05. After washing with TBST-0.05 (5 x 5 min) and TBS (1 x 5 min), the membranes were developed using the Clarity Western ECL Substrate (Bio-Rad) and imaged with Amersham Imager 680 RGB (GE Healthcare). The primary

and secondary antibodies used for these experiments, as well as their respective combinations and working dilutions were the same as those mentioned in subsection 2.12 and this information can be found in **Table 3**, **Table 4** and **Table 5**.

Membrane stripping and re-probing with the anti-H3 antibody

Following imaging, the membranes were re-probed with the H3 antibody. Membrane stripping (2 x 5 min) was performed with an acidic buffer (200 mM glycine, 0.1% (w/v) SDS, 1% (v/v) Tween-20, pH 2.2 using HCl). After washing with PBS (2 x 10 min) and TBS (2 x 5 min), the membranes were blocked with 5% non-fat dry milk in TBST-0.05 for 1 h at room temperature. The membranes were then washed with TBST-0.05 (3 x 5 min) and incubated overnight at 4 ºC with anti-H3 primary antibody diluted in 5% BSA in TBST-0.05, according to the experimental design indicated in **Table 10**. After washing with TBST-0.1 (6 x 5 min), the membranes were incubated at room temperature for 1 h with HRP-conjugated anti-mouse IgG (H+L) secondary antibody diluted in 5% non-fat dry milk in TBST-0.1. After washing with TBST-0.1 (7 x 5 min) and TBS (1 x 5 min), the membranes were developed using the Clarity Western ECL Substrate (Bio-Rad) and imaged with Amersham Imager 680 RGB (GE Healthcare). The primary and secondary antibodies used for these experiments, as well as their respective combinations and working dilutions were the same as those mentioned in subsection 2.12. and this information can be found in **Table 3**, **Table 4** and **Table 6**.

ELISA

An ELISA was also performed to confirm the Western blot results showing K9 deacetylation following incubation with Sirt3. To this end, a standard curve of the H3K4C**K9Ac protein was prepared by making two-fold serial dilutions from 16 to 0.5 µg/mL in carbonate buffer (100 mM, pH 9.6). Aliquots (28 µL) of the same control and reaction samples as those used for the corresponding Western blot analysis were diluted with 182 µL of carbonate buffer (100 mM, pH 9.6) to a final concentration of 8 µg/mL. These diluted control and reaction samples were plated together with the H3K4C**K9Ac serial dilutions.

A Nunc MaxiSorp 96 well ELISA plate was coated with duplicate serial dilutions of H3K4C**K9Ac and incubated overnight at 4 ºC. After washing the plate (3 x 5 min) with PBST-0.05, blocking buffer (200 µL per well) was added to each well (3% and 5% non-fat dry milk in PBST-0.05 for anti-H3 and anti-H3K9Ac probing, respectively). After incubating for 2 h at room temperature, the plate was washed (1 x 5 min). Primary antibody dilutions (100 µL per well) of anti-H3 and anti-H3K9Ac were then added. Following incubation at room temperature for 2 h, the plate was washed with PBST-0.05 (3 x 5 min) and secondary antibody dilutions (100 µL per well) were added. After incubating for 1 h at room temperature, the plate was washed (5 x 5 min) with PBST-0.05. TMB substrate (90 µL per well, Invitrogen) was added and the plate was left to incubate at room temperature. After sufficient color development, the colorimetric reaction was stopped with a 1M sulfuric acid solution (50 µL per well). Absorbance at 450 nm was then recorded on a Tecan Infinite M200 plate reader. The primary and secondary antibodies used for

these experiments, as well as their respective combinations and working dilutions were the same as those mentioned in subsection 2.13 and this information can be found in **Table 7**, **Table 8** and **Table 9**.

ELISA results were analyzed using GraphPad Prism software (version 6.01) by fitting the data of each combination of target protein (H3K4C**K9Ac) and probing (anti-H3 and anti-H3K9Ac) to a four-parameter logistic regression of a standard curve, with the logarithm of the concentration plotted on the x-axis and the absorbance plotted on the y-axis. Subsequently, the level of K9 acetylation and H3 concentration were interpolated from the standard curve of H3K4C**K9Ac probed with anti-H3K9Ac and anti-H3 antibodies, respectively.

2.15 Deacetylation assays of H3K4C**K9Ac with Sirt6

To establish whether the H3K4C**K9Ac protein can be recognized by histone-modifying enzymes, this chemically acetylated histone was used as a substrate in NAD^+-dependent deacetylation assays with sirtuin-6 (Sirt6).

Lyophilized H3K4C**K9Ac protein was resuspended in Milli-Q water and quantified using the Bradford assay. The control and reaction samples were then prepared with a final volume of 45 µL for ELISA or 15 µL for Western blot experiments. For both samples and experiments, the 10x concentrated deacetylation buffer (250 mM Tris, 1.37 M NaCl, 27 mM KCl, 10 mM $MgCl_2$, pH 8.0) was diluted 1:10 in Milli-Q water. DTT (90 mM in Milli-Q water for ELISA; 30 mM in Milli-Q water for Western blot) and NAD^+ (50 mM in 50 mM Tris-HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$ from Enzo Life Sciences) were added to a final concentration of 1 mM and 3.5 mM, respectively. An aliquot of H3K4C**K9Ac was added to the mixture at a final concentration of 60 ng/µL. For the reaction sample, a 6.7 µL aliquot containing recombinant Sirt6 (1 µg/µL in 50 mM Tris, 100 mM NaCl, pH 8.0 containing 20% glycerol from BioVision) was also added. The control and reaction samples were then incubated at 37 ºC and 400 rpm for 4 h. Following incubation, samples were spun down and placed on ice. K9 deacetylation was subsequently evaluated by Western blot and ELISA.

Western blot analysis

5 µL each of control and reaction samples were mixed with equal volume of 2x Laemmli sample buffer. BME (0.5 µL) was added as the reducing agent and samples were boiled at 96 ^oC for 4 min. Samples were loaded into a 15% SDS-PAGE gel (300 ng per well), together with the molecular weight marker (PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, ThermoFisher). After running the gel in Tris-glycine-SDS buffer, proteins were transferred to a membrane (Amersham Protran Supported Nitrocellulose 0.2 µm) in wet conditions with 20% methanol under constant current for 2 h at 200 mA. The membrane was then washed (1 x 5 min) in TBS and blocked with 5% non-fat dry milk in TBST-0.05 for 1 h at room temperature. The membrane was then washed with TBST-0.05 (3 x 5 min) and incubated overnight at 4 ºC with anti-H3K9Ac primary antibody diluted in 5% BSA in TBST-0.05. After washing with TBST-0.05 (5 x 5 min), the membrane was then incubated at room temperature for 1 h with HRP-conjugated

anti-rabbit IgG (H+L) secondary antibody diluted in 5% non-fat dry milk in TBST-0.05. After washing with TBST-0.05 (5 x 5 min) and TBS (1 x 5 min), the membrane was developed using the Clarity Western ECL Substrate (Bio-Rad) and imaged with Amersham Imager 680 RGB (GE Healthcare). The primary and secondary antibodies used for these experiments, as well as their respective combinations and working dilutions were the same as those mentioned in subsection 2.12 and this information can be found in **Table 3**, **Table 4** and **Table 5**.

Following imaging, the membrane was then re-probed with the H3 antibody. Membrane stripping (2 x 5 min) was performed with an acidic buffer (200 mM glycine, 0.1% (w/v) SDS, 1% (v/v) Tween-20, pH 2.2 using HCl). After washing with PBS (2 x 10 min) and TBS (2 x 5 min), the membrane was blocked with 5% non-fat dry milk in TBST-0.05 for 1 h at room temperature. The membrane was then washed with TBST-0.05 (3 x 5 min) and incubated overnight at 4 ºC with anti-H3 primary antibody diluted in 5% BSA in TBST-0.05. After washing with TBST-0.1 (6 x 5 min), the membrane was incubated at room temperature for 1 h with HRP-conjugated anti-mouse IgG (H+L) secondary antibody diluted in 5% non-fat dry milk in TBST-0.1. After washing with TBST-0.1 (7 x 5 min) and TBS (1 x 5 min), the membrane was developed using the Clarity Western ECL Substrate (Bio-Rad) and imaged with Amersham Imager 680 RGB (GE Healthcare). The primary and secondary antibodies used for these experiments, as well as their respective combinations and working dilutions were the same as those mentioned in subsection 2.12. and this information can be found in **Table 3**, **Table 4** and **Table 6**.

ELISA

An ELISA was also performed to corroborate the Western blot results showing no K9 deacetylation following incubation with Sirt6. To this end, a standard curve of the H3K4C**K9Ac protein was prepared by making two-fold serial dilutions from 16 to 0.5 µg/mL in carbonate buffer (100 mM, pH 9.6). Aliquots (28 µL) of the same control and reaction samples as those used for the corresponding Western blot analysis were diluted with 182 µL of carbonate buffer (100 mM, pH 9.6) to a final concentration of 8 µg/mL. These diluted control and reaction samples were plated together with the H3K4C**K9Ac serial dilutions.

A Nunc MaxiSorp 96 well ELISA plate was coated with duplicate serial dilutions of H3K4C**K9Ac and incubated overnight at 4 ºC. After washing the plate (3 x 5 min) with PBST-0.05, blocking buffer (200 µL per well) was added to each well (3% and 5% non-fat dry milk in PBST-0.05 for anti-H3 and anti-H3K9Ac probing, respectively). After incubating for 2 h at room temperature, the plate was washed (1 x 5 min). Primary antibody dilutions (100 µL per well) of anti-H3 and anti-H3K9Ac were then added. Following incubation at room temperature for 2 h, the plate was washed with PBST-0.05 (3 x 5 min) and secondary antibody dilutions (100 µL per well) were added. After incubating for 1 h at room temperature, the plate was washed (5 x 5 min) with PBST-0.05. TMB substrate (90 µL per well, Invitrogen) was added and the plate was left to incubate at room temperature. After sufficient color development, the colorimetric reaction was stopped with a 1M sulfuric acid solution (50 µL per well). Absorbance at 450 nm was then recorded on a Tecan Infinite M200 plate reader. The primary and secondary antibodies used for

these experiments, as well as their respective combinations and working dilutions were the same as those mentioned in subsection 2.13 and this information can be found in **Table 7**, **Table 8** and **Table 9**.

ELISA results were analyzed using GraphPad Prism software (version 6.01) by fitting the data of each combination of target protein (H3K4C**K9Ac) and probing (anti-H3 and anti-H3K9Ac) to a four-parameter logistic regression of a standard curve, with the logarithm of the concentration plotted on the x-axis and the absorbance plotted on the y-axis. Subsequently, the level of K9 acetylation and H3 concentration were interpolated from the standard curve of H3K4C**K9Ac probed with anti-H3K9Ac and anti-H3 antibodies, respectively.

3. Results

3.1 Experiments in the H3K4C-(1-15) peptide

3.1.1 Confirmation of previous experimental data

Experimental data obtained within the group had shown that the sequential conjugation to maleimide-DBCO followed by a SPAAC reaction with an acetyl donor bearing a click counterpart could be applied to acetylate histone peptides with site-specificity. To confirm these results and evaluate the feasibility of this strategy in proteins, the experiments concerning the site-specific acetylation of peptide H3K4C-(1-15) at K9 were validated. To this end, an HPLC method that enables a clear separation of all peaks of interest throughout the two chemical reactions was first developed (for information, see subsection 2.1). The conjugation of H3K4C-(1-15) (**Figure 16**) to maleimide-DBCO (4 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 37 °C) resulted in the full conversion to the corresponding H3K4C-(1-15)-maleimide-DBCO product (**Figure 17**).

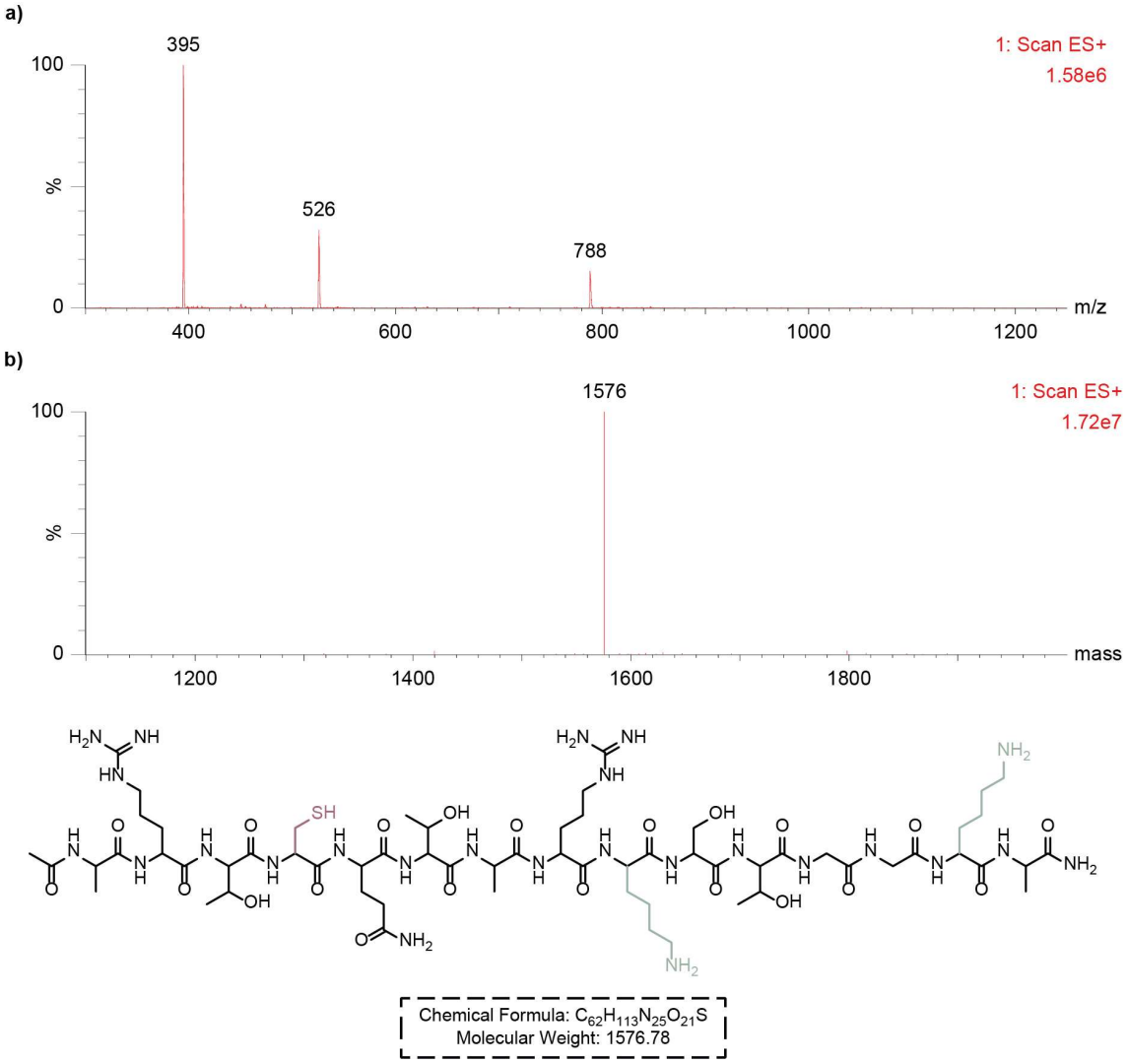


Figure 16. ESI-MS of the H3K4C-(1-15) peptide (50 µM) in ammonium acetate buffer (20 mM, pH 8.0). **a)** ion series, and **b)** deconvoluted mass spectrum with representation of the major product identified: Peptide H3K4C-(1-15) (calculated mass = 1577 Da; observed mass = 1576 Da).

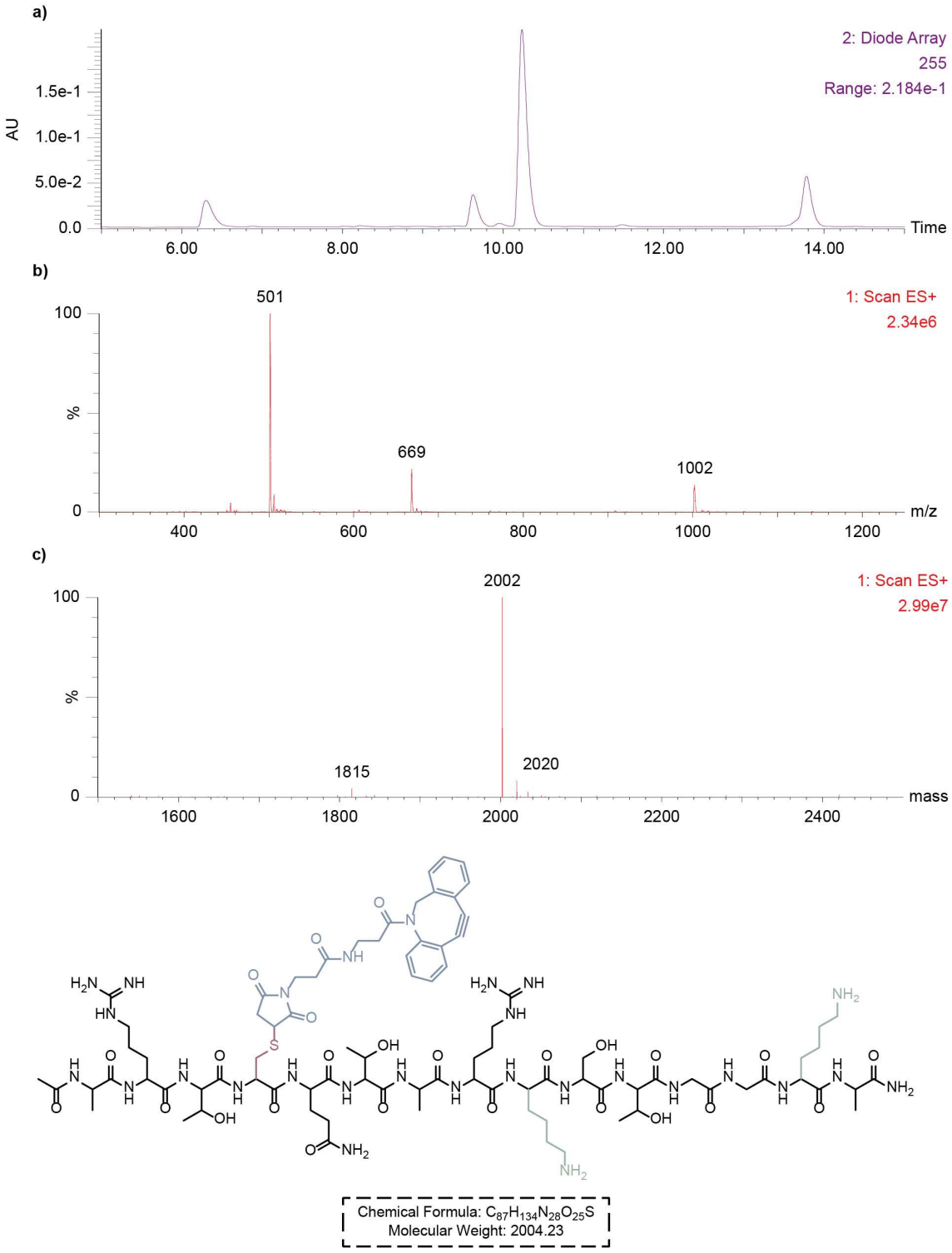


Figure 17. LC-MS of the H3K4C-(1-15) peptide (50 μM) following incubation with maleimide-DBCO (200 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 37 °C and 600 rpm. **a)** UV chromatogram at 255 nm, **b)** ion series, and **c)** deconvoluted mass spectrum with representation of the major product identified: Peptide H3K4C-(1-15)-maleimide-DBCO (calculated mass = 2004 Da; observed mass = 2002 Da).

With the H3K4C-(1-15)-maleimide-DBCO peptide, the directed acetylation of K9 was then attempted. Incubation with compound **4** (16 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 37 °C) afforded the peptide containing the click by-product from the SPAAC and a single acetylation (**Figure 18**), henceforth referred to as H3K4C**-(1-15)-K9Ac peptide.

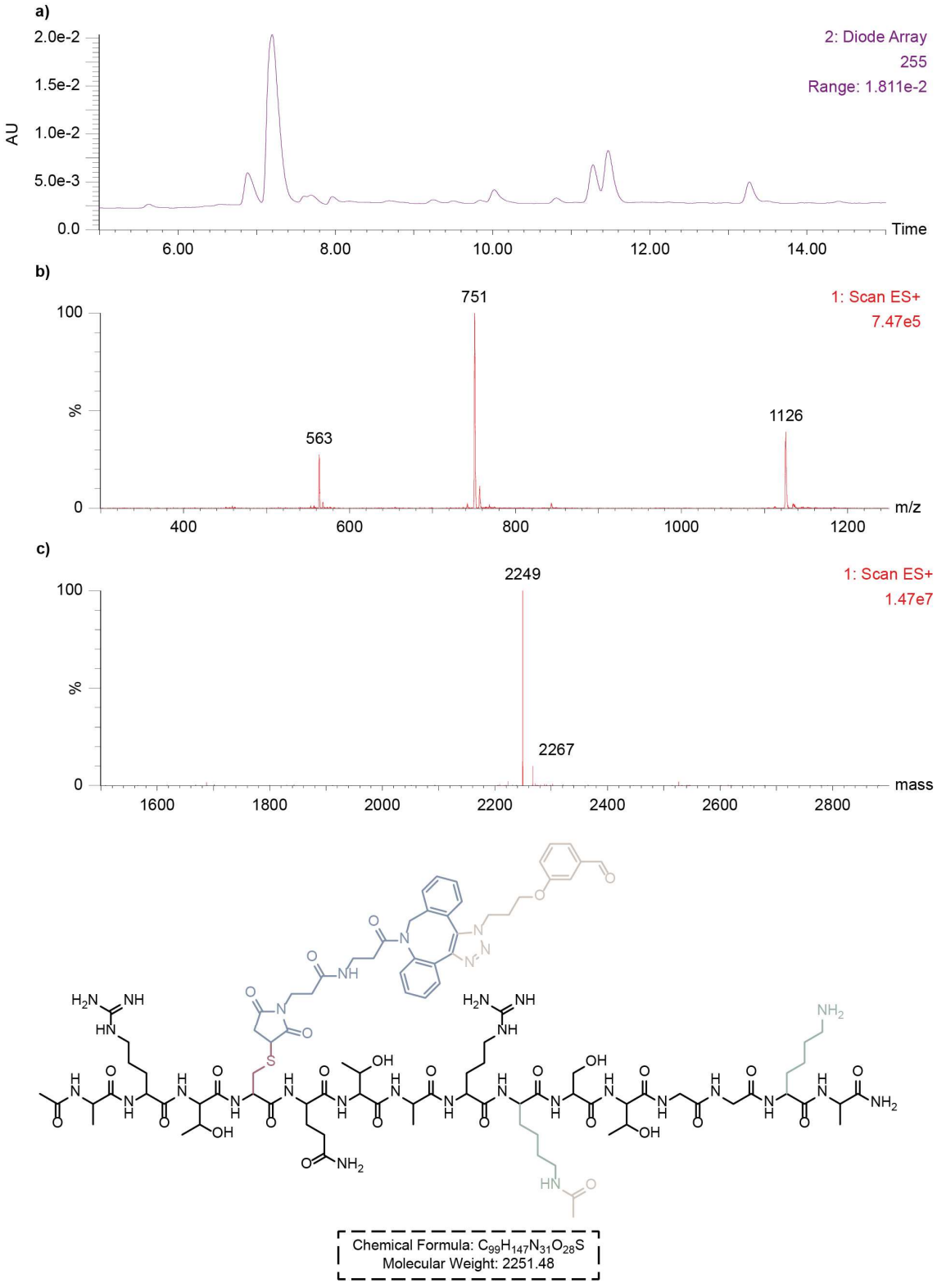


Figure 18. LC-MS of the H3K4C-(1-15)-maleimide-DBCO peptide (25 µM) following incubation with compound **4** (400 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 37 °C and 600 rpm. **a)** UV chromatogram at 255 nm, **b)** ion series, and **c)** deconvoluted mass spectrum with representation of the major product identified: Peptide H3K4C**-(1-15)-K9Ac (calculated mass = 2251 Da; observed mass = 2249 Da).

3.1.2 Desulfurization attempts in the H3K4C**-(1-15)-K9Ac peptide using a mild system

With the modified peptide in hand, the possibility of removing the SPAAC by-product that remains attached at the cysteine in the H3K4C**-(1-15)-K9Ac peptide was also evaluated. Previous attempts within the group were performed by reacting the peptide with BME to promote thiol-exchange via a retro-Michael addition (12.5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 23 h at 25 ^oC). Although somewhat successful, the reaction did not reach completion due to irreversible hydrolysis of the maleimide that confers resistance to the retro-Michael addition. Hydrolysis of the acetylated K9 residue to the corresponding unmodified H3K4C-(1-15) peptide was also observed following retro-Michael addition. Furthermore, considering that the unmodified H3K4C-(1-15) peptide was detected, intramolecular rearrangements could have occurred due to the prolonged reaction time and instead of an acetylated amine, one would detect an acetylated cysteine thiol. Therefore, an alternative approach to remove the SPAAC by-product at the cysteine could consist in the desulfurization at this residue to afford the H3K4A-(1-15)-K9Ac peptide.

Several strategies have been described for the desulfurization of peptides and proteins, which is mainly used as a subsequent step following NCL to convert ligation products containing cysteine into native alanine residues. These include (1) metal-based desulfurization using Raney nickel or Pd/Al_2O_3 under hydrogen,[248,287] (2) metal-free desulfurization using the radical initiator VA-044 in the presence of TCEP and *tert*-butyl mercaptan,[288] (3) photochemical desulfurization which depends on the use of metal complexes as photocatalysts and light for radical initiation,[289,290] and (4) P-B desulfurization which uses TCEP in combination with a boron reagent such as $NaBH_4$ or $LiBEt_3H$.[291]

Here, the chosen strategy consisted of desulfurization with TCEP/$NaBH_4$, since this system is inexpensive and mild, thus avoiding the use of metal complexes or other initiators and being potentially suitable for histone proteins in future applications. In this context, the desulfurization reaction was carried out by incubating the H3K4C**-(1-15)-K9Ac peptide with a premixed solution of TCEP and $NaBH_4$ (100 equiv., 1:1 molar ratio, $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0). This reaction resulted in a mixture of products, with the initial H3K4C**-(1-15)-K9Ac peptide and the intended H3K4A-(1-15)-K9Ac product absent from the spectrum (**Figure 19**). Interestingly, the main peak at 1632 Da corresponds to the di-acetylated H3K4A-(1-15) peptide. Since desulfurization was attempted without a prior purification step to remove the unreacted maleimide-DBCO, compound **4** and resulting click product formed between the two reagents, this result can be construed as follows: preceding, concomitantly or following desulfurization of H3K4C**-(1-15)-K9Ac to H3K4A-(1-15)-K9Ac, unreacted compound **4** and/or the click product formed between the latter and maleimide-DBCO promotes acetylation of the remaining free lysine in the peptide. Purification by dialysis was attempted, but most of the modified peptide was not recovered.

While unsuccessful, this preliminary experiment established that desulfurization can be used to generate a mono-acetylated H3 tail peptide without the SPAAC by-product at the cysteine if prior purification is performed to remove the acetylation reagents present in solution.

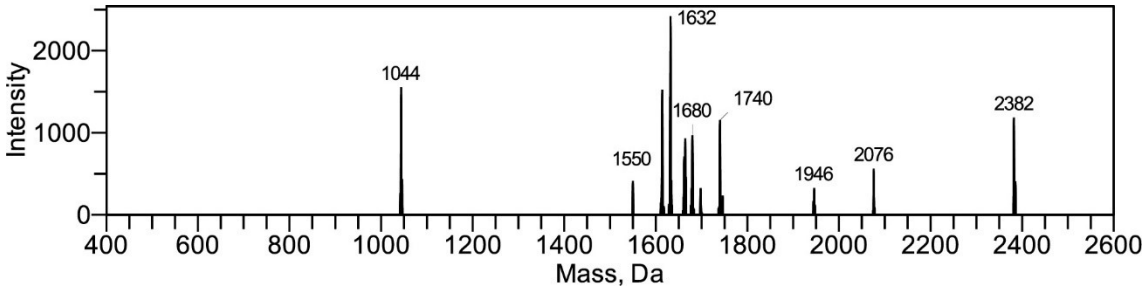


Figure 19. Deconvoluted mass spectrum following incubation of the H3K4C**-(1-15)-K9Ac peptide (10 μM) with a solution of TCEP and $NaBH_4$ (100 equiv., 1:1 molar ratio) in ammonium acetate buffer (20 mM, pH 7.0) for 3 h at 37 °C and 900 rpm. Major product identified: Di-acetylated H3K4A-(1-15) peptide (calculated mass = 1629 Da; observed mass = 1632 Da).

3.2 Application of the cysteine-assisted click chemistry strategy to acetylate histone H3 proteins with site-specificity

Having demonstrated the feasibility of the cysteine-assisted click-chemistry strategy for the proximity-driven, site-specific acetylation of histone H3 tail peptides, its broader applicability was subsequently evaluated using the full-length protein. To this end, various histone H3 mutants were recombinantly expressed in *E. coli* and purified by FPLC.

3.2.1 Expression and purification of the H3K4C protein

To develop a method based on an initial thiol conjugation at a given site of H3, the reactivity of its single-cysteine residue (C110) can be substituted by an alanine, without noticeable effects on nucleosome structure and function.[292] In fact, this substitution is naturally found in yeast.[293] Therefore, the *X. laevis* H3 (C110A) sequence was chosen as the basis for the site-specific acetylation strategy developed in this thesis, since it is commonly used in biophysical studies of chromatin and in the development of chemical biology approaches to histone modification.

Previous results from the research group had established that the H3K4C**-(1-15)-K9Ac peptide is more susceptible to Sirt6-mediated deacetylation in comparison to H3K14C**-(1-15)-K9Ac. Furthermore, the cysteine at position 4 has no lysine residues other than K9 at equal or similar distance in the primary sequence, while the cysteine at position 14 has both K9 and K18. Thus, the single-cysteine mutant H3K4C (C110A) was chosen as the first target on which to test the proximity-driven acetylation strategy in proteins. Its sequence was generated from the wild-type H3 sequence of *X. laevis* by site-directed mutagenesis.

Overexpression of individual histones in *E. coli* is straightforward but produces insoluble aggregates that accumulate in inclusion bodies. Protein purification from inclusion bodies is quite laborious and requires the use of strong denaturants such as dimethyl sulfoxide (DMSO), urea or guanidine that can solubilize them.[294] With the purpose of circumventing the issue of inclusion body formation and subsequent use of denaturing reagents, expression and purification of H3K4C were initially attempted by using a construct with an MBP sequence that had been generated within the group prior to the start of this thesis. The MBP tag has the advantage of enhancing the solubility of the target protein while facilitating its purification by affinity chromatography.[295,296] MBP is more effective at solubilizing a protein when fused to its N-terminus rather than its C-

terminus.[297] For this reason, the MBP tag is typically fused to the N-terminal region of the protein of interest. To add an additional purification option and obtain the full-length protein containing only the histone-specific amino acids, the plasmid backbone also contained a 6xHis tag between the MBP and H3K4C sequences, which was flanked upstream and downstream by a TEV protease and EK recognition sites, respectively.

While expression and purification of the MBP-H3K4C fusion protein proceeded without difficulty, cleavage with either TEV or EK proteases followed by purification of the resulting target protein proved quite challenging. Most of the fusion protein remained uncleaved following digestion, particularly when using the TEV protease (**Figure 20**). In hindsight, the inefficiency of the TEV protease cleavage might be due to the proximity of its recognition site to the MBP sequence, which can hinder protease access. Nevertheless, a cation-exchange chromatography was attempted following EK digestion, which did not yield any quantifiable amount of protein. This unfavorable outcome might be explained by the natural propensity of histones to self-aggregate, particularly when these contain a free cysteine capable of forming disulfide bridges.[298]

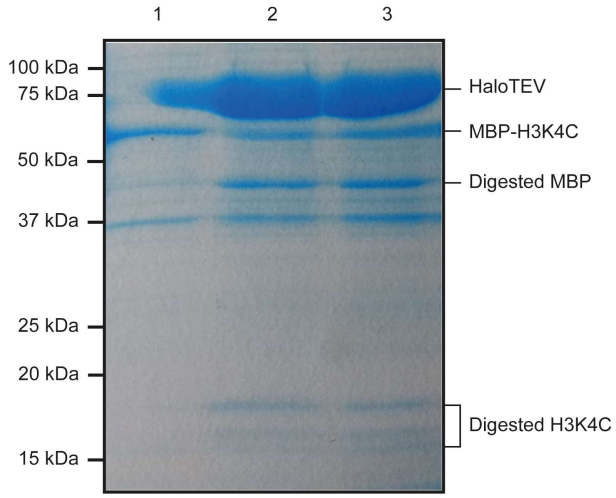


Figure 20. Coomassie staining to evaluate the enzymatic cleavage of the purified MBP-H3K4C protein with the HaloTEV protease. A 15% SDS-PAGE gel was loaded with MBP-H3K4C (lane 1), MBP-H3K4C following incubation with TEV protease for 30 min at 30 ºC and then overnight at 4 ºC with agitation (lane 2), and MBP-H3K4C following incubation with TEV protease for 1 h at room temperature and then overnight at 4 ºC without agitation.

At this stage, the more classical avenue for histone purification using denaturing conditions was pursued. In this context, a plethora of protocols have been described and although a few include sequences with affinity tags to facilitate the purification of histones, their presence is not mandatory. Traditional protocols typically comprise the following sequential steps: (1) inclusion body isolation from the insoluble fraction of bacterial extracts, (2) solubilization of inclusion bodies under denaturing conditions by addition of DMSO followed by incubation with unfolding buffer containing 7 M guanidine hydrochloride, (3) size-exclusion and cation exchange chromatography under denaturing conditions using urea-based buffers, (4) dialysis against water, and finally (5) lyophilization and storage.[123,294] Several reports to shorten the duration of the original purification protocols have been described in which some of these steps, namely the need for size-exclusion

chromatography, are omitted.[299] Recently, an even simpler workflow has been described in which the requirement for inclusion body isolation is also bypassed.[283]

Therefore, purification of H3K4C was subsequently attempted using this streamlined protocol. Here, inclusion bodies are solubilized during cell lysis, which is performed under denaturing conditions to extract the recombinant histone protein. By using an anion- and cation-exchange chromatography in tandem under denaturing conditions, purification of H3K4C was achieved following elution with a sodium chloride gradient from 240 to 300 mM. A representative SDS-PAGE gel of the collected fractions containing the H3K4C protein is shown in **Figure 21**. The deconvoluted ESI-MS spectrum of the purified H3K4C protein is shown in **Figure 22**. As determined by LC-MS, the purity level of the unoxidized H3K4C protein was approximately 80%. One notable contaminant present in the spectrum likely corresponds to the oxidized version of the H3K4C protein. Commercial H3 displays similar purity levels, although these are estimated by SDS-PAGE and thus do not discriminate between unoxidized and oxidized protein, and many have an affinity tag to facilitate its purification. Considering an overall purity level of 80%, a typical H3K4C purification yielded approximately 15 mg of protein per liter of culture.

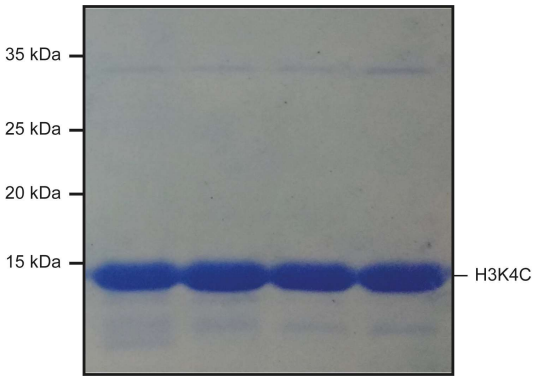


Figure 21. Representative SDS-PAGE gel of the collected fractions during purification of the H3K4C protein. To collect these fractions, cation-exchange chromatography with a sodium chloride gradient from 240 mM to 300 mM in SAU buffer containing 5 mM BME was used.

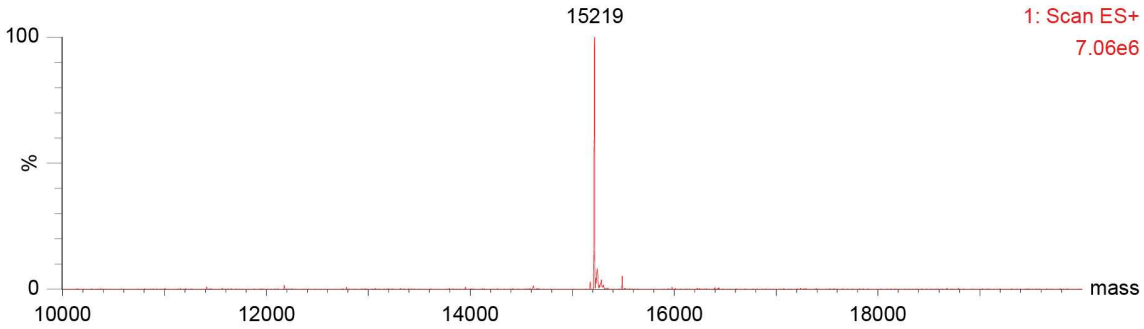


Figure 22. Deconvoluted ESI-MS spectrum of the purified H3K4C protein (25 µM) in Milli-Q water. Major product identified: Protein H3K4C (calculated mass = 15214 Da; observed mass = 15219 Da).

3.2.2 Optimization of reaction conditions in the H3K4C protein

The same set of bioconjugation reactions as those previously described for the H3K4C-(1-15) peptide was performed for the full-length protein. To this end, lyophilized H3K4C protein was

resuspended in a mixture of water and ammonium acetate (20 mM, pH 7.0). It should be noted that resuspension of the protein would sometimes lead to the formation of several visible aggregates. This is likely due to the previously mentioned propensity of histones towards self-aggregation. Two different approaches to deal with this undesirable event were implemented throughout the project. The first approach, more time-consuming and implemented in the beginning, consisted of separating the soluble and insoluble protein fractions by centrifugation and solubilizing the latter in denaturing buffer (6 M guanidine hydrochloride, 20 mM Tris-HCl pH 7.4, 10 mM DTT). Both fractions were then purified by size-exclusion chromatography using PD MiniTrap G-25 columns to remove contaminating reagents and place the protein in a more suitable buffer for chemical modification reactions and LC-MS analysis. The second approach, simpler and implemented much later in the project, consisted of diluting in the same buffer only a fraction of the resuspended protein that is necessary for the immediate chemical reactions, while the remaining protein is lyophilized and stored for future use.

Conjugation of H3K4C with the maleimide-DBCO clickable handle

Initial conjugation efforts to maleimide-DBCO were performed following mild reduction with DTT and removal of this reducing agent by size-exclusion chromatography. In this context, treatment of reduced H3K4C with an excess of 2.5 equiv. of maleimide-DBCO under slightly basic conditions resulted in the full conversion to the corresponding H3K4C-maleimide-DBCO product ($NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 °C, **Figure 23**).

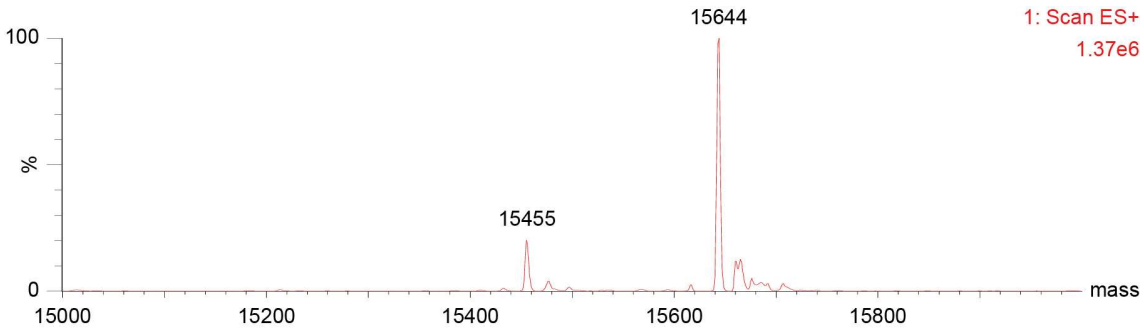


Figure 23. Deconvoluted ESI-MS spectrum obtained following size-exclusion purification of the DTT-reduced H3K4C protein (50 μM) and incubation with maleimide-DBCO (125 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 °C and 400 rpm. Major product identified: Protein H3K4C-maleimide-DBCO (calculated mass = 15641 Da; observed mass = 15644 Da).

Later conjugation attempts of H3K4C to maleimide-DBCO were performed following mild reduction with TCEP, with and without prior purification by size-exclusion chromatography to remove this reducing reagent. TCEP was chosen as the preferred reductant in this work, due to its increased stability as well as stronger and faster reducing capabilities over a broader pH range. In addition, DTT has been reported to interfere with maleimide conjugation more than TCEP.[300] When H3K4C was reduced with TCEP (5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C) and the latter removed prior to maleimide-DBCO conjugation, full conversion was achieved using the same conditions described previously for DTT (2.5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h

at 25 °C, **Figure 24**). However, when the previously reduced H3K4C protein was again treated with TCEP (4 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C) followed by maleimide-DBCO incubation without a purification step in-between, full conversion was not observed using the same reaction conditions (2.5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 °C, **Figure 25**), with a small percentage of the unmodified protein still visible in the spectrum. This indicates that TCEP interferes to a slight extent in the conjugation of H3K4C to maleimide-DBCO. Evidence for TCEP and maleimide-DBCO interaction will be provided throughout subsection 3.2.

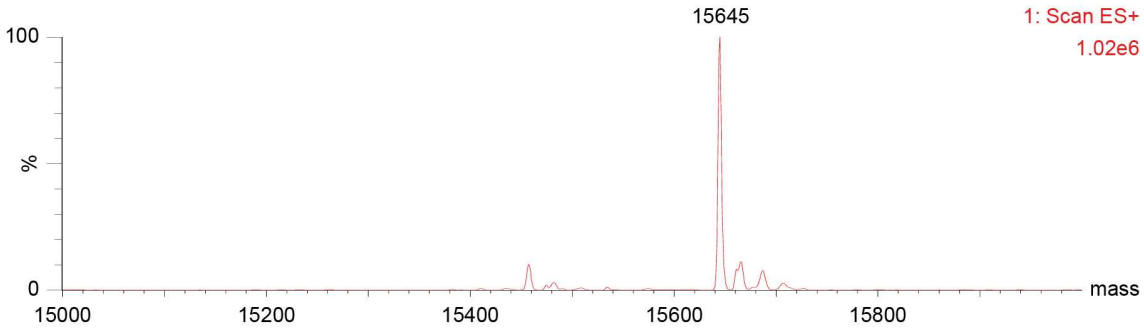


Figure 24. Deconvoluted ESI-MS spectrum obtained following size-exclusion purification of the TCEP-reduced H3K4C protein (50 µM) and incubation with maleimide-DBCO (125 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 °C and 400 rpm. Major product identified: Protein H3K4C-maleimide-DBCO (calculated mass = 15641 Da; observed mass = 15645 Da).

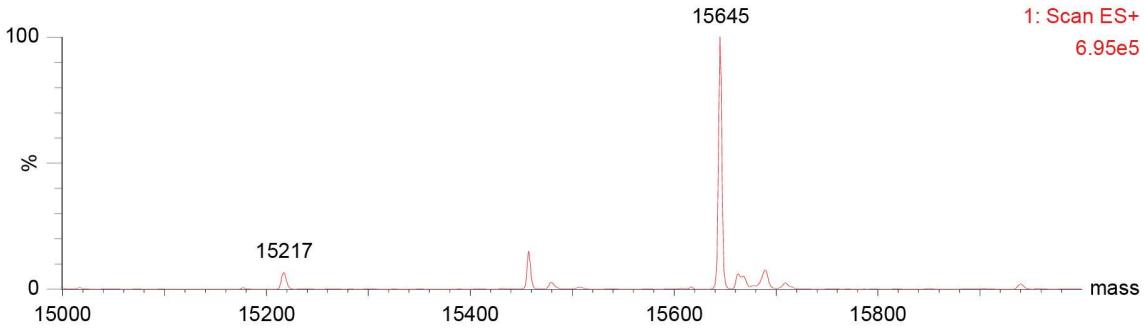


Figure 25. Deconvoluted ESI-MS spectrum obtained following size-exclusion purification of the TCEP-reduced H3K4C protein (50 µM), incubation with TCEP (200 µM) for 30 min at 25 °C and 400 rpm and maleimide-DBCO (125 µM) for 1 h at 25 °C and 400 rpm in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Protein H3K4C-maleimide-DBCO (calculated mass = 15641 Da; observed mass = 15645 Da). An additional peak corresponding to the unmodified H3K4C protein was also observed (calculated mass = 15214 Da; observed mass = 15217 Da).

Site-specific acetylation of H3K4C conjugated to maleimide-DBCO following SPAAC reaction with compound **4**

Following purification by ultrafiltration of the modified protein to remove the maleimide-DBCO reagent in excess, the proximity-driven, site-specific acetylation was attempted. Initially, the SPAAC reaction with compound **4** was attempted at a basic pH to afford in one single step the protein containing the click by-product from the SPAAC and a single acetylation, henceforth referred to as H3K4C**K9Ac. In the beginning of the project, full conversion to the H3K4C**K9Ac product was obtained using only 1 equiv. of compound **4** (**Figure 26**), while 4 equiv. of the same reagent were later required. Two explanations can be put forward to explain this occurrence. First,

purification of H3K4C was optimized as the project progressed, namely in terms of the number and duration of steps during gradient elution with sodium chloride from the cation-exchange column. This was done to minimize the amount of protein contaminants present by increasing the separation between these and H3K4C and thus obtain higher purity levels for the latter. Therefore, it is expected that for protein samples with higher purity levels of H3K4C, the amount of reagent required to achieve full conversion is also increased. The second explanation concerns the possible degradation of compound **4** with time, although this was not observed by LC-MS when performing protein conjugation reactions.

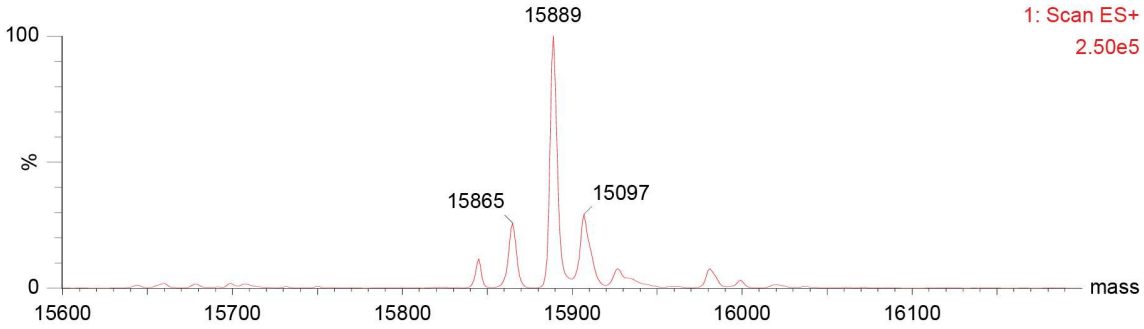


Figure 26. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-maleimide-DBCO protein (50 µM) with compound **4** (50 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 ºC and 400 rpm. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed mass = 15889 Da).

Evaluation of the H3K4C**K9Ac protein obtained following SPAAC reaction with 4 equiv. of compound **4** and stored at -20 ºC for 1 month without prior purification to remove the excess reagent revealed the presence of one and two extra acetylated residues (**Figure 27**). This fact is not surprising, since lysine residues other than K9 exist within the sequence and these can become acetylated in the presence of excess reagent when given sufficient time. In addition, it is worth mentioning that such a result does not diminish the usefulness of the approach developed here, since acetylation was observed to be site-specific for K9 under appropriate reaction conditions. Therefore, purification following the SPAAC reaction is required to remove the excess compound and obtain exclusively the desired H3K4C**K9Ac product.

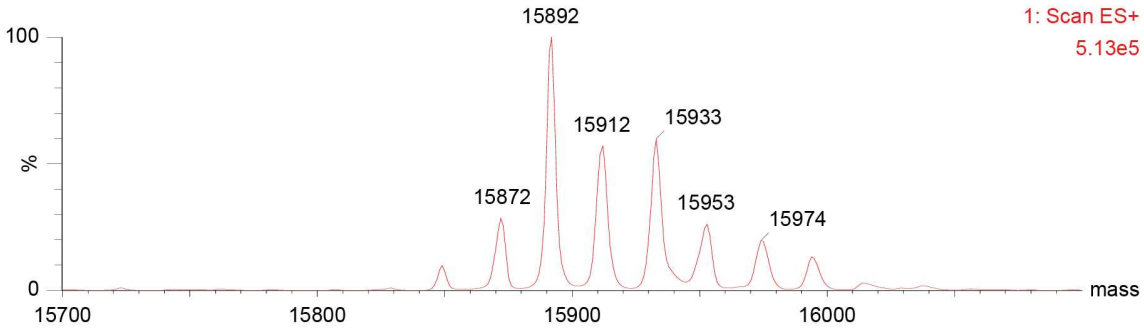


Figure 27. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-maleimide-DBCO protein (25 µM) with compound **4** (100 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 ºC and 400 rpm and subsequent storage at -20 ºC for 1 month. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed mass = 15892 Da). Additional peaks corresponding to the di- (calculated mass = 15931 Da; observed mass = 15933 Da) and tri-acetylated (calculated mass = 15973 Da; observed mass = 15974 Da) products are also observed.

Later, the SPAAC reaction was also performed in acidic conditions ($NH_4^+CH_3CO_2^-$ 20 mM, pH 4.9) to demonstrate the pH dependence of the acetylation reaction and ensure minimal reactivity of the excess acetyl donor still in solution to lysine residues other than K9 before its full removal by ultrafiltration. This afforded the unacetylated H3K4C protein containing the SPAAC product with the *gem*-dithioacetate group still left to react, henceforth referred to as unAcH3K4C** (**Figure 28**). At this stage, three different approaches were pursued to promote acetylation at K9.

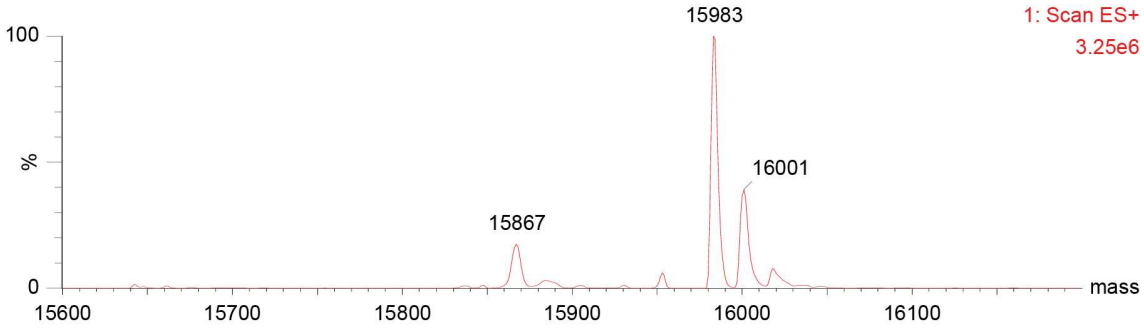


Figure 28. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-maleimide-DBCO protein (25 µM) with compound **4** (100 µM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 ºC and 400 rpm. Major product identified: Protein unAcH3K4C** (calculated mass = 15981 Da; observed mass = 15983 Da). This same spectrum will appear later in subsection 3.2.7 in the context of optimized reaction conditions for the site-specific acetylation at K9.

In the first approach, cruder and implemented at the beginning of the project, the pH of the ammonium acetate buffer in which the SPAAC reaction occurred was increased to approximately 8.0 with a small volume of a 1% ammonia solution to promote acetylation (**Figure 29**).

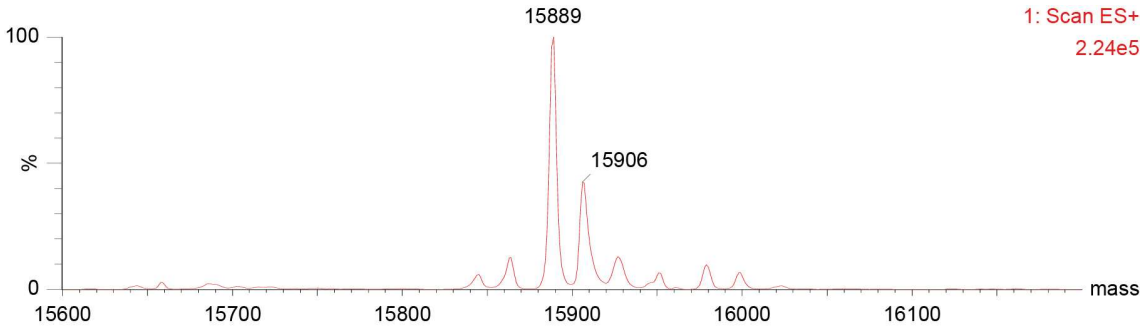


Figure 29. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-maleimide-DBCO-protein (50 µM) with compound **4** (50 µM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 ºC and 400 rpm and subsequent basification of the reaction buffer with a 1% ammonia solution. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed mass = 15889 Da).

In the second approach, the unAcH3K4C** protein was progressively exchanged by ultrafiltration to the same buffer but at a basic pH ($NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0), which afforded the H3K4C**K9Ac protein (**Figure 30**) while removing the excess acetyl donor.

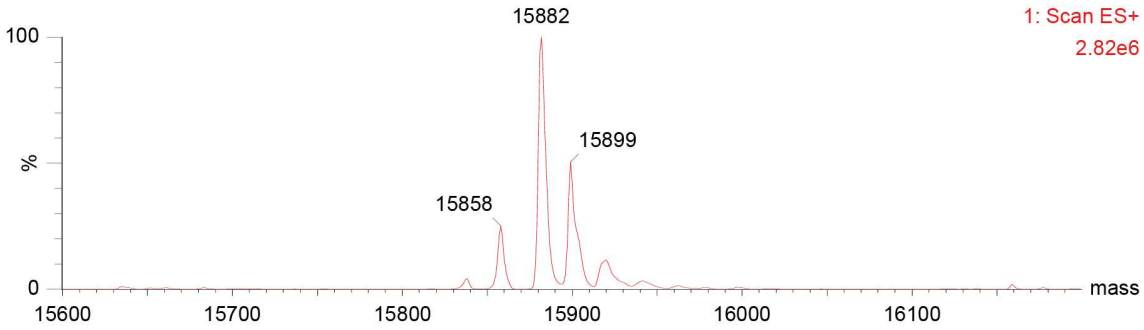


Figure 30. Deconvoluted ESI-MS spectrum of the H3K4C**K9Ac protein (25 μM) obtained following buffer exchange by ultrafiltration of unAcH3K4C** to ammonium acetate buffer at a basic pH (20 mM, pH 8.0). Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed mass = 15882 Da).

In the third approach, the unAcH3K4C** protein was exchanged by ultrafiltration to the same buffer but at a slightly less acidic pH ($NH_4^+CH_3CO_2^-$ 20 mM, pH 5.8), which removed the excess acetyl donor and spontaneously afforded the H3K4C**K9Ac protein after a 3 to 4-day incubation on ice (**Figure 31**). This final approach, more detailed in subsection 3.2.7, was the one chosen to generate the H3K4C**K9Ac protein used for experiments in section 3.7.

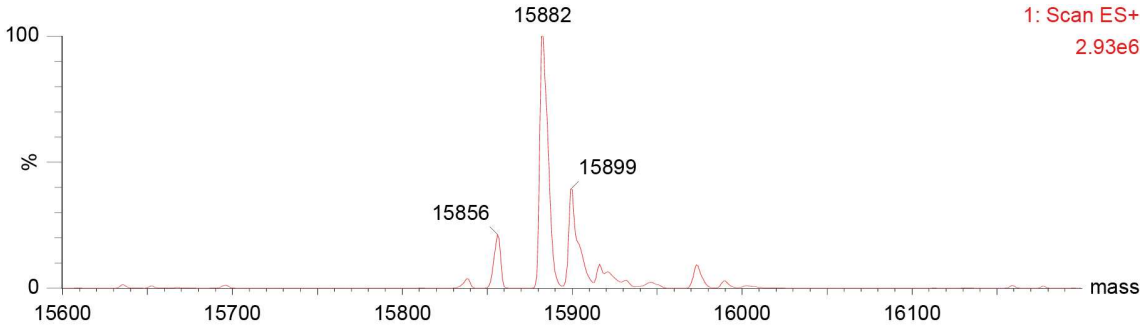


Figure 31. Deconvoluted ESI-MS spectrum of the H3K4C**K9Ac protein (25 μM) obtained following buffer exchange by ultrafiltration of unAcH3K4C** to ammonium acetate buffer (20 mM, pH 5.8) and incubation on ice for 3 to 4 days. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed = 15882 Da). This same spectrum will appear later in subsection 3.2.7 in the context of optimized reaction conditions for the site-specific acetylation at K9.

3.2.3 Attempts to remove the SPAAC by-product via a retro-Michael addition

With the first mono-acetylated histone in hand, the possibility of removing the SPAAC by-product that remains attached at the cysteine was evaluated. As mentioned in subsection 3.1.2, previous attempts within the group to remove this by-product in acetylated peptides had been performed via treatment with BME. Thus, as an initial approach, this same reagent was used to promote thiol-exchange via a retro-Michael addition in the H3K4C**K9Ac protein. Different experimental conditions were tested, namely in terms of the number of equivalents of BME, reaction pH and incubation time. However, the success of these experiments was quite limited. The initial experimental condition in which the retro-Michael reaction proceeded to a greater extent produced only approximately 5% of H3K4CK9Ac (5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 22 h at 25 °C, **Figure 32**), with most of the protein retaining the SPAAC by-product at the cysteine. An undesired side-product with a similar peak intensity to H3K4CK9Ac corresponding to the unmodified H3K4C protein was also observed. At this time, it was hypothesized that by increasing

the reaction pH to 10, this could increase the nucleophilicity of BME by promoting its deprotonation (pK_a = 9.72 at 25 °C) and thus lead to higher yields of H3K4CK9Ac formation. However, this did not occur and under certain experimental conditions (10 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 10.0, 3 h at 25 °C, **Figure 33**), an increase in the yield of H3K4C relative to H3K4CK9Ac was in fact seen. In addition, it was also observed that the yield of H3K4CK9Ac did not increase significantly when a higher number of BME equivalents was used. To explain these results, it was speculated that BME could be forming disulfide bridges at higher concentrations and in basic conditions. Thus, to test this hypothesis, TCEP was added to the reaction mixture to reduce the BME disulfide. While this resulted in an increased yield of the H3K4CK9Ac product, a similar amount of H3K4C protein was also found (5 equiv. BME, 10 equiv. TCEP, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 23 h at 37 °C, **Figure 34**).

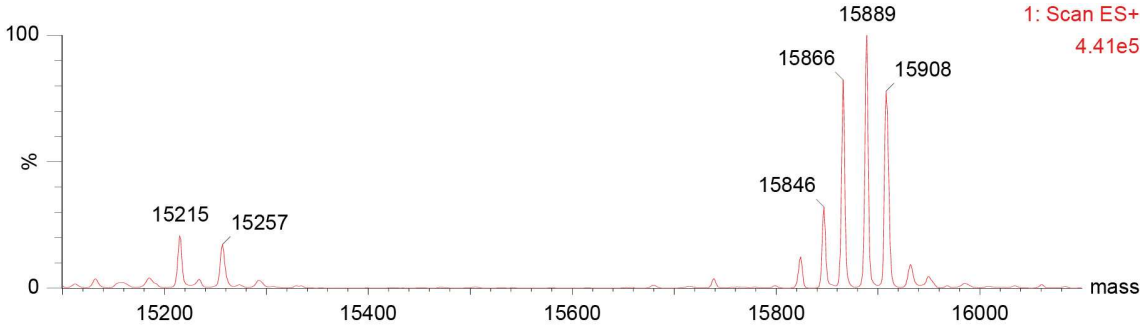


Figure 32. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C**K9Ac protein (50 μM) with BME (250 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 22 h at 25 °C and 400 rpm. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed mass = 15889 Da).

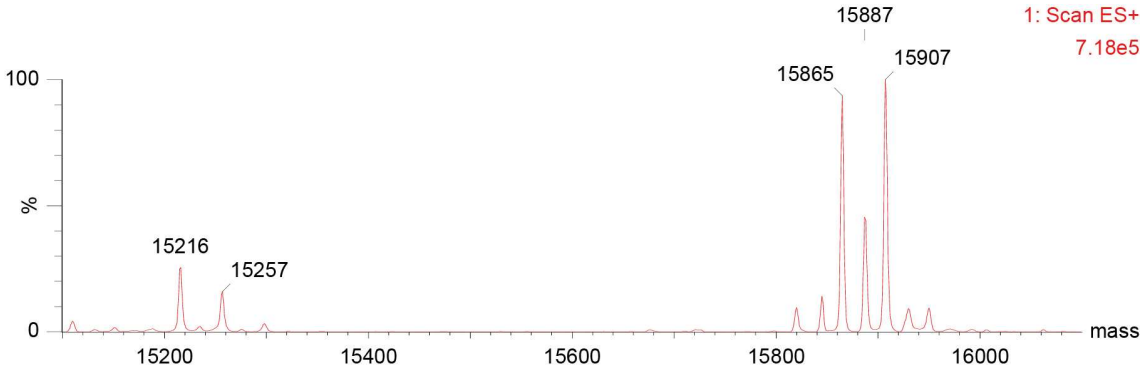


Figure 33. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C**K9Ac protein (50 μM) with BME (500 μM) in ammonium acetate buffer (20 mM, pH 10.0) for 3 h at 25 °C and 400 rpm. Major product identified: Hydrolyzed H3K4C**K9Ac protein (calculated mass = 15907 Da; observed mass = 15907 Da).

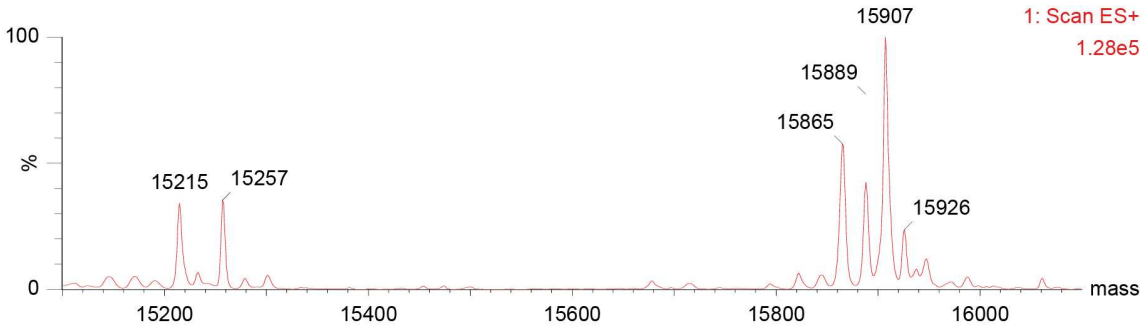


Figure 34. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C**K9Ac protein (50 μM) with BME (250 μM) and TCEP (500 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 23 h at 37 °C and 400 rpm. Major product identified: Hydrolyzed H3K4C**K9Ac protein (calculated mass = 15907 Da; observed mass = 15907 Da).

3.2.4 Confirmation of site-specific acetylation in the H3K4C**K9Ac protein by tandem mass spectrometry

The site-specificity of the cysteine-assisted click chemistry strategy for proximity-driven acetylation at lysine 9 of histone H3K4C was confirmed through tandem MS. By using FT-ICR MS with CID and ECD fragmentation methods and matching all resulting fragment ions to the sequence, a coverage of 67% was achieved comprising the modified C4 and K9 residues (**Figure 35**).

Figure 35. Sequence coverage obtained following tandem MS of the H3K4C**K9Ac protein. Predicted versus calculated mass difference is 0.134 Da or 8.5 ppm. Combining all fragment ions from CID (b, y, blue) and ECD (c, z, red) and matching the sequence with an allowed maximum deviation of 5 ppm, a coverage of 67% was achieved comprising the modified C4 and K9 residues.

3.2.5 Expression and purification of the H3R52C protein

Having demonstrated that the cysteine-assisted click chemistry strategy can be used to acetylate the H3K4C protein with site-specificity, the next step in this project consisted of evaluating whether it could be generalized to other lysine residues for which the biological role of acetylation is also important, less understood or yet unknown. To test this possibility, the more internal K56 residue of histone H3 was chosen as the second target of the directed acetylation strategy outlined here. As mentioned in subsection 1.2.1, H3K56 acetylation is a conserved epigenetic mark found in yeast and higher eukaryotes that plays a critical role in genomic stability by promoting efficient nucleosome assembly following DNA replication. Therefore, the single-cysteine mutant H3R52C (C110A) was generated, recombinantly expressed and purified.

The purification of H3R52C was performed according to the same optimized protocol described previously for the H3K4C protein. The deconvoluted ESI-MS spectrum of the purified

H3R52C protein is shown in **Figure 36**. As determined by LC-MS, the purity level of the unoxidized H3R52C protein was approximately 70%. One notable contaminant present in the spectrum likely corresponds to the oxidized version of the H3R52C protein. Commercial H3 displays slightly higher purity levels, however these are estimated by SDS-PAGE and thus do not discriminate between unoxidized and oxidized protein, and many have an affinity tag to facilitate its purification. Considering an overall purity of 70%, a typical H3R52C purification yielded approximately 13 mg of protein per liter of culture.

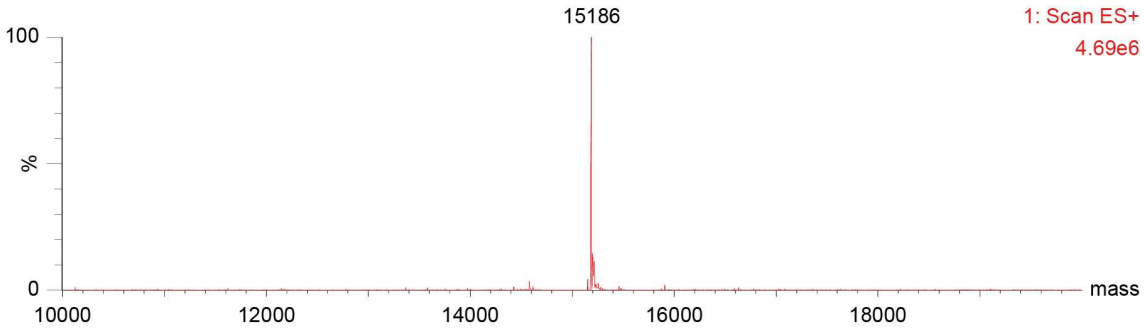


Figure 36. Deconvoluted ESI-MS spectrum of the purified H3R52C protein (25 µM) in Milli-Q water. Major product identified: Protein H3R52C (calculated mass = 15186 Da; observed mass = 15186 Da).

3.2.6 Optimization of reaction conditions in the H3R52C protein

The reactions performed previously in the H3K4C protein served as a starting point for optimizing the conjugation protocol in H3R52C. It should be noted that the insights obtained from H3R52C experiments also contributed to the optimized acetylation workflow in the H3K4C protein.

Conjugation of H3R52C with the maleimide-DBCO clickable handle

Due to the low abundance of its disulfide following H3R52C resuspension (**Figure 37**), conjugation to maleimide-DBCO was initially attempted without prior reduction of the protein with TCEP. However, it was found that the reaction conditions promoted the formation of the H3R52C disulfide during maleimide-DBCO conjugation (5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 ^oC, **Figure 38**). Therefore, this result established that prior reduction of H3R52C with TCEP was required to perform conjugation at the cysteine.

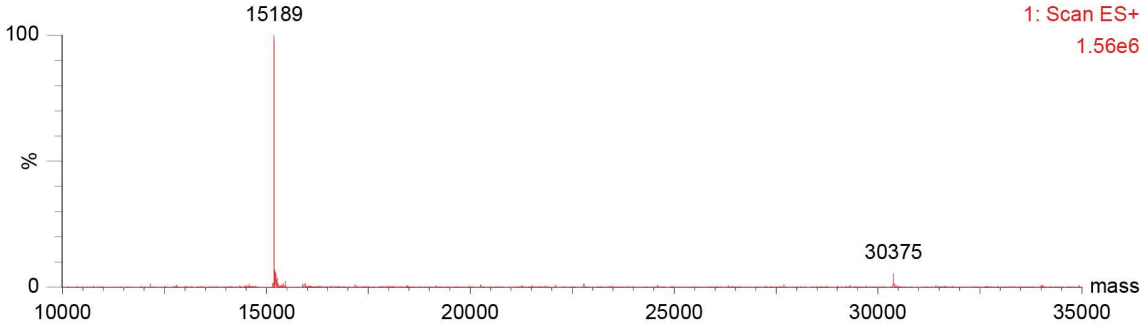


Figure 37. Deconvoluted ESI-MS spectrum obtained following resuspension and size-exclusion purification of the H3R52C protein (50 µM). Major product identified: Protein H3R52C (calculated mass = 15186 Da; observed mass = 15189 Da).

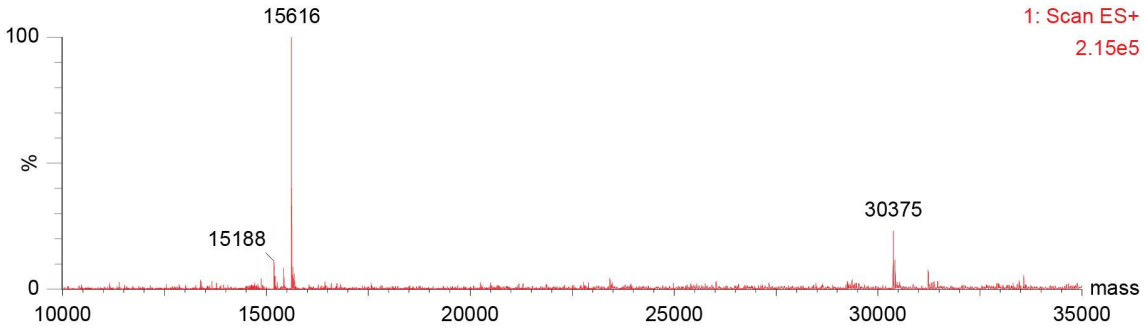


Figure 38. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C protein (50 μM) with maleimide-DBCO (250 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 ºC and 400 rpm. Major product identified: Protein H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15616 Da).

Attempts to conjugate maleimide-DBCO to H3R52C were then performed following mild reduction with TCEP without subsequent purification by size-exclusion chromatography to remove this reducing agent. Using similar conditions to those described previously for the full conversion of the H3K4C protein following maleimide-DBCO incubation, approximately 50% of H3R52C remained unconjugated (4 equiv. TCEP, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 ºC, followed by 2.5 equiv. maleimide-DBCO, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 ºC, **Figure 39**). This result was the first indication of a reactivity difference between H3K4C and H3R52C.

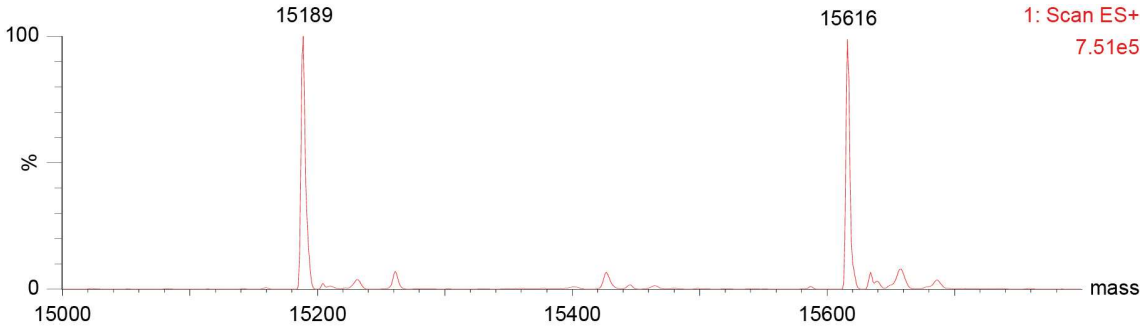


Figure 39. Deconvoluted ESI-MS spectrum obtained following sequential incubation of the H3R52C protein (50 μM) with TCEP (200 μM) for 1 h at 25 ºC and 400 rpm and maleimide-DBCO (125 μM) for 1 h at 25 ºC and 400 rpm in ammonium acetate buffer (20 mM, pH 8.0). Major products identified: Proteins H3R52C (calculated mass = 15186 Da; observed mass = 15189 Da) and H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15616 Da).

To achieve full conversion of the H3R52C protein it was necessary to drastically increase the number of equivalents of maleimide-DBCO (5 equiv. TCEP, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 ºC, followed by 15 equiv. maleimide-DBCO, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 ºC, **Figure 40**). However, under these experimental conditions a minor peak corresponding to a second maleimide-DBCO addition was found, which likely occurred at a slightly deprotonated lysine residue.

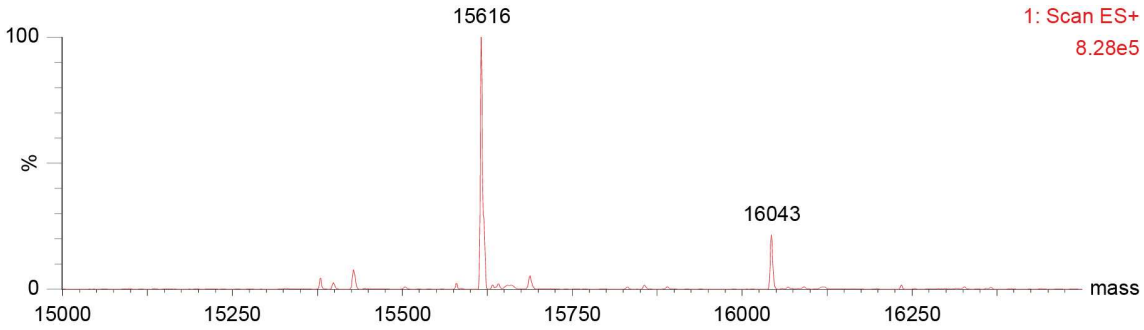


Figure 40. Deconvoluted ESI-MS spectrum obtained following sequential incubation of the H3R52C protein (50 μM) with TCEP (250 μM) for 30 min at 25 °C and 400 rpm and maleimide-DBCO (750 μM) for 1 h at 25 °C and 400 rpm in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Protein H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15616 Da). An additional peak corresponding to the H3R52C protein modified with two maleimide-DBCO reagents was also observed (calculated mass = 16041 Da; observed mass = 16043 Da).

Since the pK_a of a protein cysteine (pK_a = 8.5) is typically lower than that of a lysine (pK_a = 10.5),[301,302] at this stage it was hypothesized that the double addition could be circumvented by decreasing the reaction pH. Indeed, reducing the pH of the reaction to 7.0 resulted in only one addition being observed (5 equiv. TCEP, $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 30 min at 25 °C, followed by 15 equiv. maleimide-DBCO, $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 1 h at 25 °C, **Figure 41**). Halving the duration of maleimide-DBCO incubation also afforded the corresponding modified H3R52C protein with full conversion (5 equiv. TCEP, $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 30 min at 25 °C, followed by 15 equiv. maleimide-DBCO, $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 30 min at 25 °C, **Figure 42**).

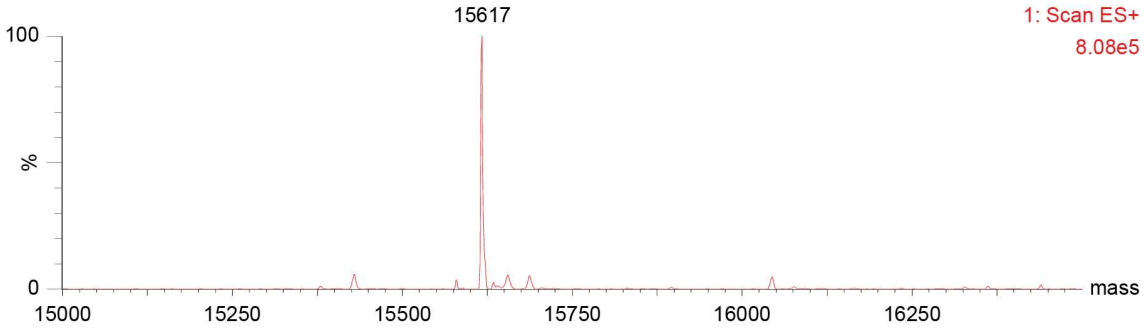


Figure 41. Deconvoluted ESI-MS spectrum obtained following sequential incubation of the H3R52C protein (50 μM) with TCEP (250 μM) for 30 min at 25 °C and 400 rpm and maleimide-DBCO (750 μM) for 1 h at 25 °C and 400 rpm in ammonium acetate buffer (20 mM, pH 7.0). Major product identified: Protein H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15617 Da).

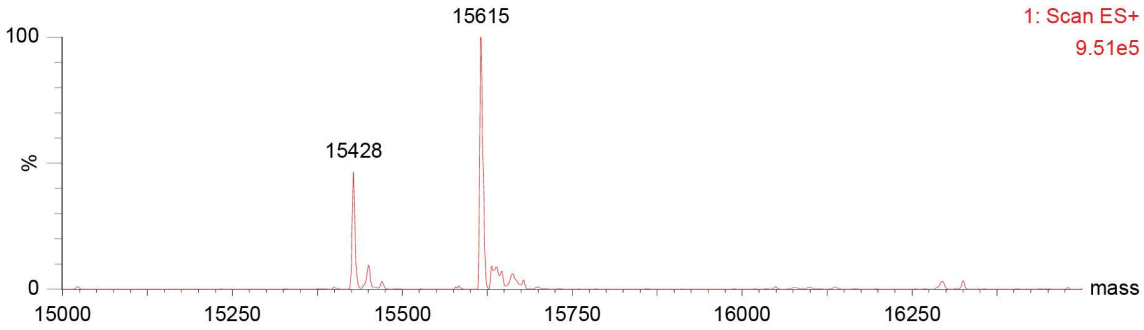


Figure 42. Deconvoluted ESI-MS spectrum obtained following sequential incubation of the H3R52C protein (25 µM) with TCEP (125 µM) for 30 min at 25 ºC and 400 rpm and maleimide-DBCO (375 µM) for 30 min at 25 ºC and 400 rpm in ammonium acetate buffer (20 mM, pH 7.0). Major product identified: Protein H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15615 Da).

Site-specific acetylation of H3R52C conjugated to maleimide-DBCO following SPAAC reaction with compound **4**

Having seemingly optimized the maleimide-DBCO conjugation to H3R52C, the proximity-driven, site-specific acetylation was subsequently tested. Following purification by ultrafiltration of the modified protein to remove the excess reagents, the SPAAC reaction with compound **4** was attempted. Initially, neutral conditions were used to evaluate the possibility of obtaining in one single step the protein containing the click by-product from the SPAAC and a single acetylation, henceforth referred to as H3R52C**K56Ac. However, more than one acetylation was observed (1 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 30 min at 25 ºC, **Figure 43**), even with a significant amount of unacetylated H3R52C protein containing the SPAAC product with the *gem*-dithioacetate group still left to react, henceforth referred to as unAcH3R52C**.

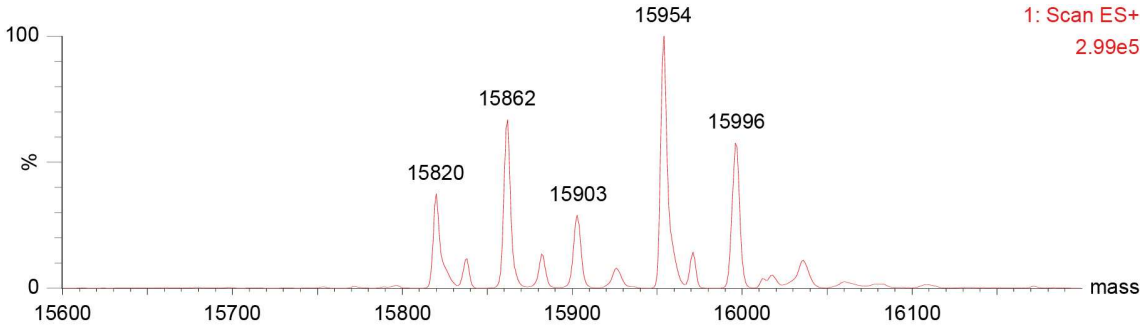


Figure 43. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C-maleimide-DBCO protein (50 µM) with compound **4** (1 equiv.) in ammonium acetate buffer (20 mM, pH 7.0) for 30 min at 25 ºC and 400 rpm. Major product identified: Protein unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15954 Da). Second major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15862 Da). Additional peaks corresponding to the di-acetylated protein with (calculated mass = 15995; observed mass = 15996 Da) and without (calculated mass = 15903 Da; observed mass = 15903 Da) the *gem*-dithioacetate group at the SPAAC moiety still left to react were also observed.

To address this issue, it was hypothesized that performing the SPAAC reaction in acidic conditions to first obtain the unAcH3R52C** product and subsequently increasing the pH to promote the transfer of a single acetyl group could increase the specificity of this strategy and the yield of the desired H3R52C**K56Ac protein. To accomplish this, several approaches similar to

those described previously for the H3K4C protein were pursued. In the first approach, cruder and implemented at the beginning of the project, the pH of the ammonium acetate buffer in which the SPAAC reaction occurred was increased to approximately 9.0 with a small volume of a 1% ammonia solution to promote acetylation. While this approach resulted in the full conversion of unAcH3R52C** and increased the amount of H3R52C**K56Ac formed, two peaks corresponding to the di- and tri-acetylated version of this protein were also observed (**Figure 44**). In addition, a significant peak corresponding to the unacetylated H3R52C protein bearing the SPAAC by-product that remains attached at the cysteine following acetylation, henceforth referred to as H3R52C**, was also observed. The results from this experiment in which the SPAAC reaction is performed at a lower pH, which is subsequently increased with a 1% ammonia solution to promote acetylation, varied on subsequent attempts and this approach was therefore dropped for the H3R52C protein.

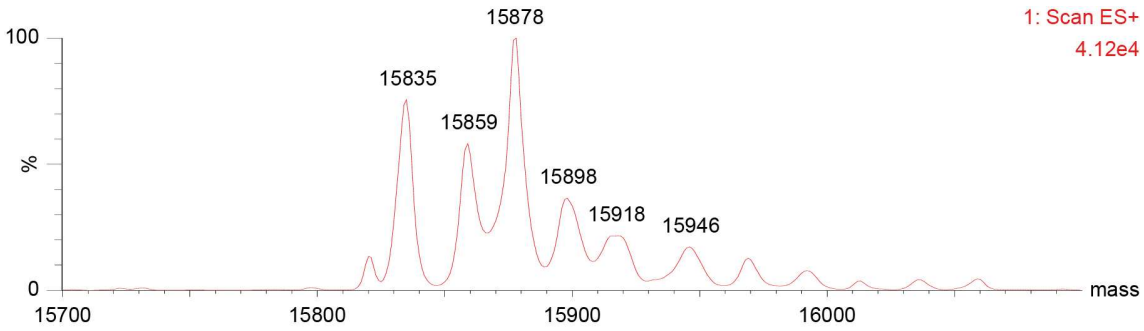


Figure 44. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C-maleimide-DBCO protein (25 μM) with compound **4** (1 equiv.) in ammonium acetate buffer (20 mM, pH 4.9) and subsequent basification with ammonia. Major product identified: Hydrolyzed H3R52C**K56Ac protein (calculated mass = 15879 Da; observed mass = 15878 Da). Second major product identified: Hydrolyzed H3R52C** (calculated mass = 15837 Da; observed mass = 15835 Da). Additional peaks corresponding to the di- (calculated mass = 15903 Da; observed mass = 15898 Da) and tri-acetylated (calculated mass = 15945 Da; observed mass = 15946 Da) H3R52C** products were also observed.

In the second approach, the unAcH3R52C** protein was first exchanged to the same acidic SPAAC reaction buffer to remove the excess acetyl donor. The acetylation reaction was then attempted by incubating the unAcH3R52C** protein in the same buffer but at higher pH values ($NH_4^+CH_3CO_2^-$ 20 mM, pH 6.5, 1 h at 25 ^oC, **Figure 45**; $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 1 h at 25 ^oC, **Figure 46**). However, a considerable amount of unAcH3R52C** remained unreacted, with peaks corresponding to the acetylated protein containing the SPAAC product with the *gem*-dithioacetate group still left to react being observed.

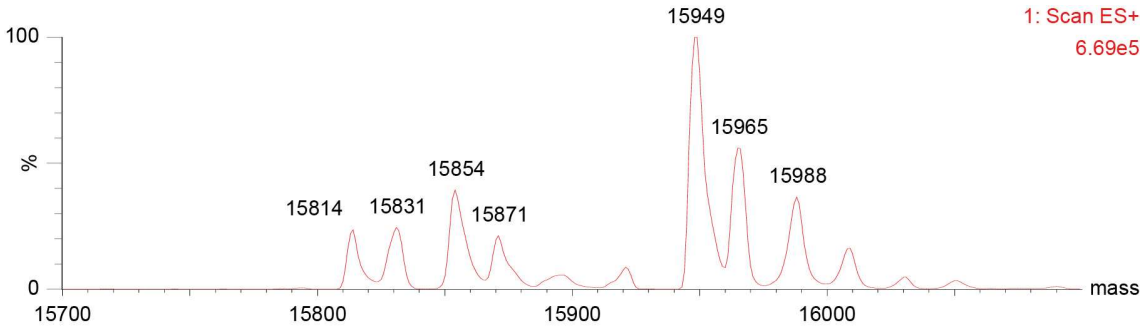


Figure 45. Deconvoluted ESI-MS spectrum obtained following incubation of the unAcH3R52C** protein (20 μM) in ammonium acetate buffer (20 mM, pH 6.5) for 1 h at 25 °C. Major product identified: Protein unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15949 Da). An additional peak corresponding to the acetylated protein with the *gem*-dithioacetate group at the SPAAC moiety still left to react (calculated mass = 15995 Da; observed mass = 15988 Da) was also observed.

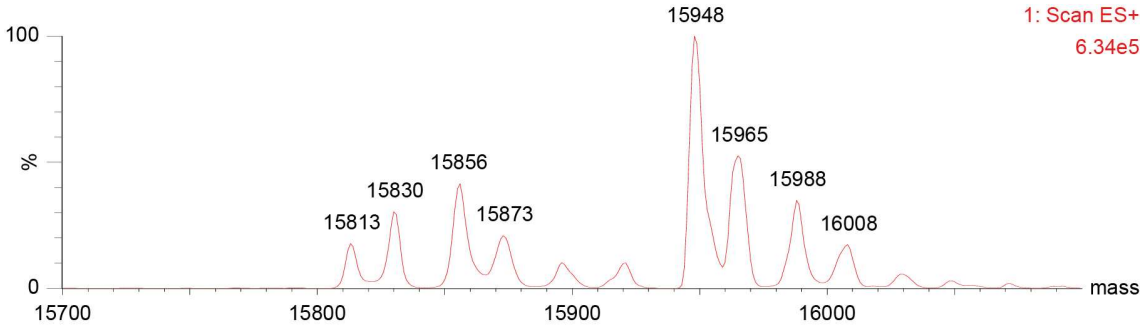


Figure 46. Deconvoluted ESI-MS spectrum obtained following incubation of the unAcH3R52C** protein (20 μM) in ammonium acetate buffer (20 mM, pH 7.0) for 1 h at 25 °C. Major product identified: Protein unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15948 Da). An additional peak corresponding to the acetylated protein with *gem*-dithioacetate group at the SPAAC moiety still left to react (calculated mass = 15995 Da; observed mass = 15988 Da) was also observed.

In a variation of the second approach, the unAcH3R52C** protein was exchanged to the same SPAAC reaction buffer but at a slightly less acidic pH ($NH_4^+CH_3CO_2^-$ 20 mM, pH 5.8) and acetylation was then attempted ($NH_4^+CH_3CO_2^-$ 20 mM, pH 6.5, 1 h at 25 °C, **Figure 47**; $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 1 h at 25 °C, **Figure 48**). While considerably less unAcH3R52C** remained unreacted, dual acetylation was already visible.

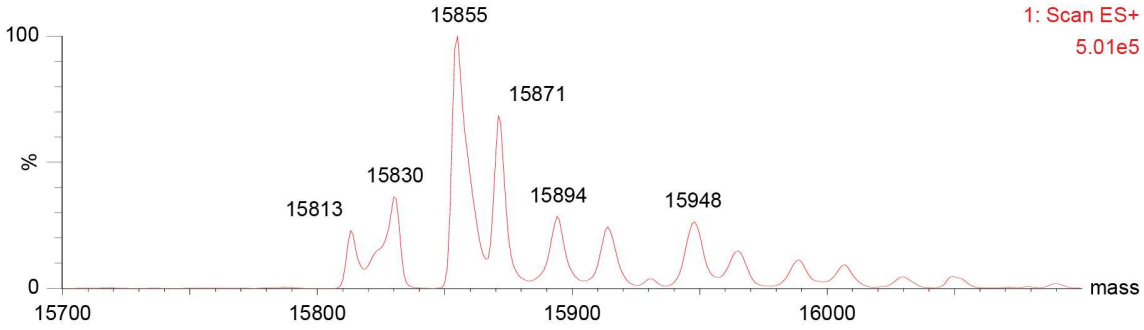


Figure 47. Deconvoluted ESI-MS spectrum obtained following buffer exchange of the unAcH3R52C** protein (20 μM) in ammonium acetate buffer (20 mM, pH 5.8) and incubation in ammonium acetate buffer (20 mM, pH 6.5) for 1 h at 25 °C. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15855 Da). Additional peaks corresponding to di-acetylated H3R52C** (calculated mass = 15903 Da; observed mass = 15894 Da) and unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15948 Da) were also observed.

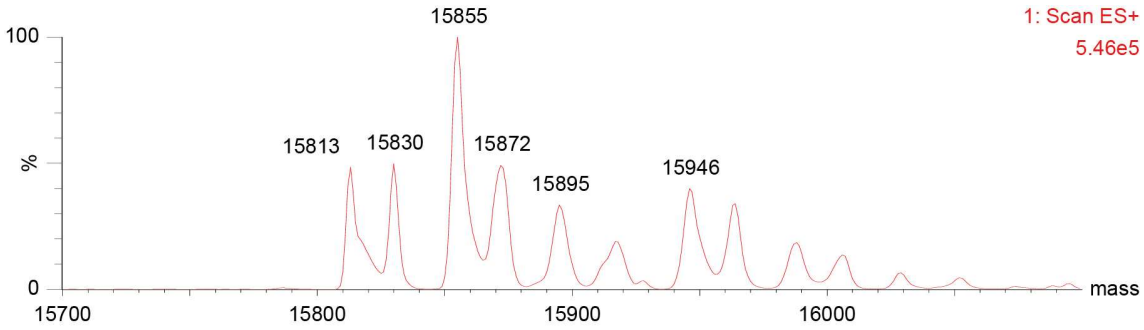


Figure 48. Deconvoluted ESI-MS spectrum obtained following buffer exchange of the unAcH3R52C** protein (20 µM) in ammonium acetate buffer (20 mM, pH 5.8) and incubation in ammonium acetate buffer (20 mM, pH 7.0) for 1 h at 25 °C. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15855 Da). Additional peaks corresponding to di-acetylated H3R52C** (calculated mass = 15903 Da; observed mass = 15895 Da) and unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15946 Da) were also observed.

In another variation of the second approach, the unAcH3R52C** protein was again exchanged to the same SPAAC reaction buffer at a slightly less acidic pH ($NH_4^+CH_3CO_2^-$ 20 mM, pH 5.8) and acetylation was then attempted in this same buffer ($NH_4^+CH_3CO_2^-$ 20 mM, pH 5.8, 5 h at 25 °C, **Figure 49**). This approach reduced the amount of di-acetylated product formed, with no visible peaks corresponding to the unAcH3R52C** protein. To reduce even further the amount of dual acetylation, the reaction was also attempted following exchange of the unAcH3R52C** protein to the same acidic buffer in which the SPAAC was carried out ($NH_4^+CH_3CO_2^-$ 20 mM, pH 4.9). However, some of the unAcH3R52C** protein remained unreacted even after 10 h, with a significant amount of H3R52C** also present ($NH_4^+CH_3CO_2^-$ 20 mM, pH 4.9, 10 h at 25 °C, **Figure 50**).

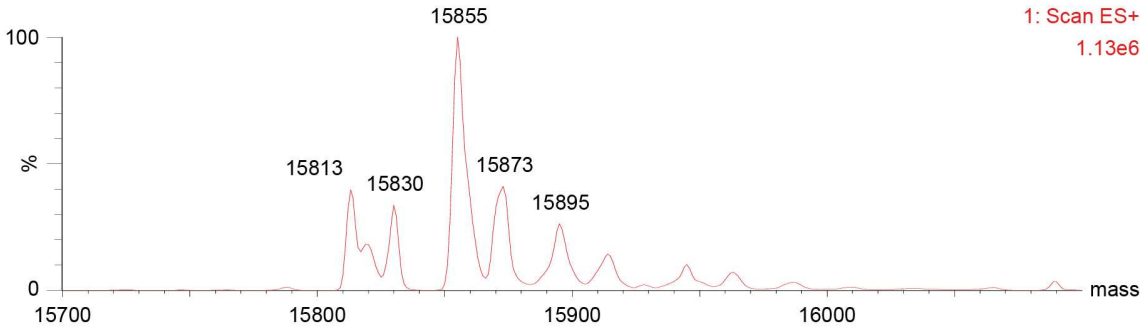


Figure 49. Deconvoluted ESI-MS spectrum obtained following buffer exchange of the unAcH3R52C** protein (20 µM) in ammonium acetate buffer (20 mM, pH 5.8) and incubation in ammonium acetate buffer (20 mM, pH 5.8) for 5 h at 25 °C and 400 rpm. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15855 Da). An additional peak corresponding to di-acetylated H3R52C** (calculated mass = 15903 Da; observed mass = 15895 Da) was also observed.

During this optimization stage of K56 acetylation in the H3R52C protein, it was serendipitously discovered that acetylation also occurs spontaneously when the unAcH3R52C** protein is produced in ammonium acetate (20 mM) at pH 4.9, buffer-exchanged to pH 5.8 and left on ice for 3 to 4 days (**Figure 51**). However, when the protein was exchanged to pH 4.9 instead of pH 5.8, approximately 40-50% of unAcH3R52C** remained unreacted, even after 6 days of incubation on ice (**Figure 52**).

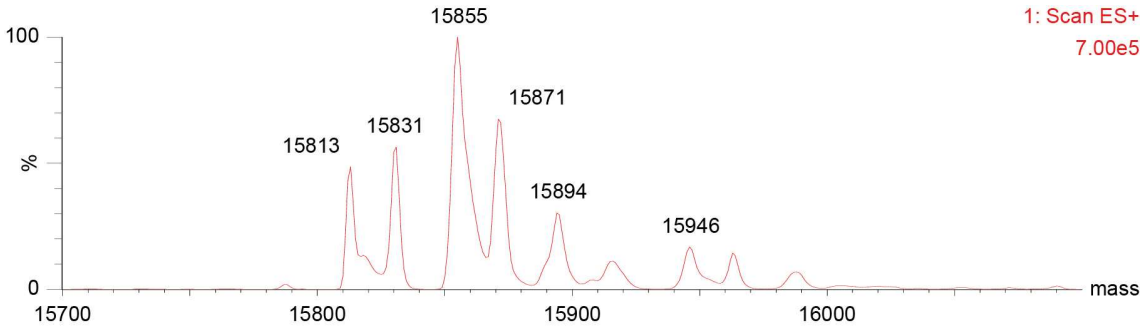


Figure 50. Deconvoluted ESI-MS spectrum obtained following incubation of the unAcH3R52C** protein (20 μM) in ammonium acetate buffer (20 mM, pH 4.9) for 10 h at 25 °C and 400 rpm. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15855 Da). Additional peaks corresponding to H3R52C** (calculated mass = 15819 Da; observed mass = 15813 Da), di-acetylated H3R52C** (calculated mass = 15903 Da; observed mass = 15894 Da) and unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15946 Da) were also observed.

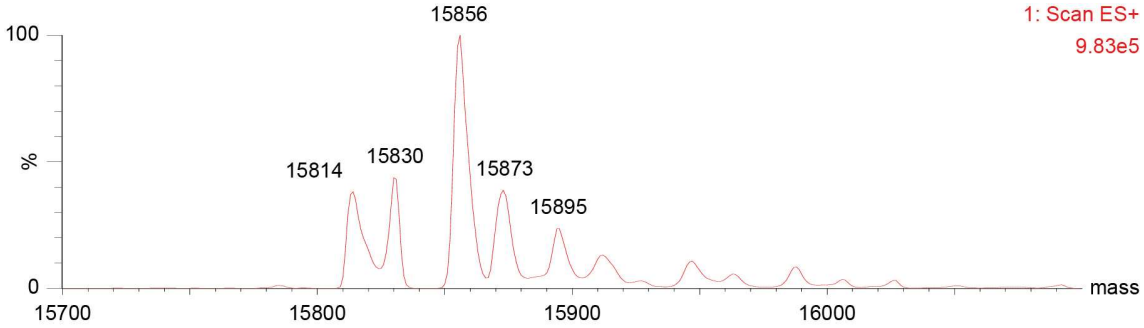


Figure 51. Deconvoluted ESI-MS spectrum obtained following incubation of the unAcH3R52C** protein (20 μM) in ammonium acetate buffer (20 mM, pH 5.8) for 3 to 4 days on ice. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15856 Da). Additional peaks corresponding to H3R52C** (calculated mass = 15819 Da; observed mass = 15814 Da) and di-acetylated H3R52C** (calculated mass = 15903 Da; observed mass = 15895 Da) were also observed.

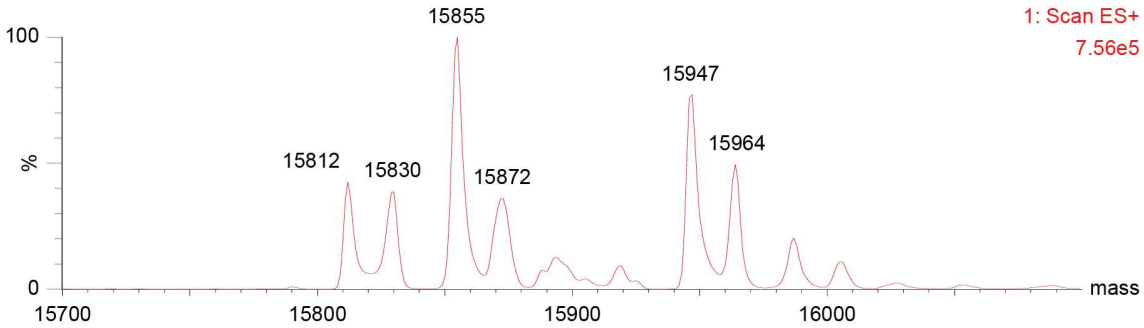


Figure 52. Deconvoluted ESI-MS spectrum obtained following incubation of the unAcH3R52C** protein (20 μM) in ammonium acetate buffer (20 mM, pH 4.9) for 6 days on ice. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15855 Da). Additional peaks corresponding to unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15947 Da) and H3R52C** (calculated mass = 15819 Da; observed mass = 15812 Da) were also observed.

Surprisingly, when replicating the experimental workflow to generate H3R52C**K56Ac from the unmodified H3R52C protein, multi-acetylation was observed. It was discovered that this event was correlated with the appearance of characteristic peaks with m/z values of 678, 769 and 1017 corresponding, respectively, to the [TCEP + maleimide-DBCO + H]$^+$, [maleimide-DBCO + compound **4** + 3H]$^+$ and [TCEP + maleimide-DBCO + compound **4** + H]$^+$ adducts. It should be

mentioned that extensive purification by ultrafiltration with Amicon centrifugal filter units was attempted to remove the [TCEP + maleimide-DBCO + H]$^+$ adduct from the H3R52C-maleimide-DBCO product as thoroughly as possible. Nevertheless, the peak corresponding to this adduct was still observed (**Figure 53**). Furthermore, two additional peaks corresponding to the [maleimide-DBCO + compound **4** + 3H]$^+$ and [TCEP + maleimide-DBCO + compound **4** + H]$^+$ adducts were observed following incubation with compound **4** (**Figure 54**). These results indicated that the [TCEP + maleimide-DBCO + H]$^+$ adduct was interacting with the H3R52C-maleimide-DBCO protein and/or the Amicon membrane and/or forming higher molecular weight aggregates that were impossible to completely remove by ultrafiltration from the H3R52C-maleimide-DBCO product.

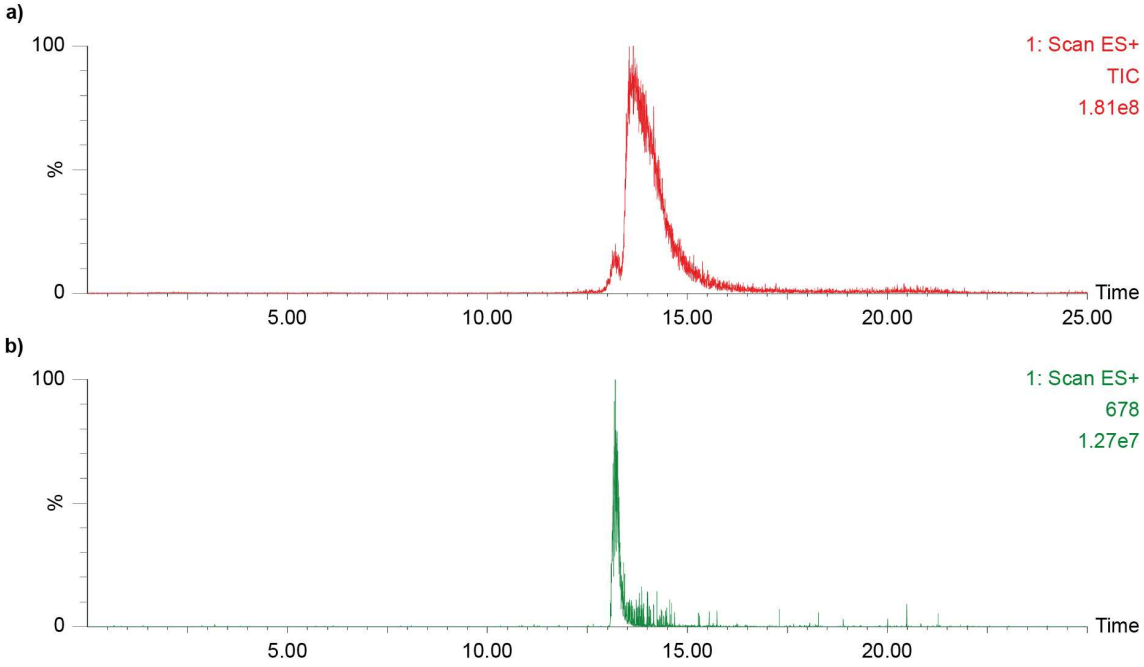


Figure 53. LC-MS of the H3R52C-maleimide-DBCO protein (25 µM) in ammonium acetate buffer (20 mM, pH 4.9) following purification by ultrafiltration to remove the excess maleimide-DBCO reagent. **a)** total ion current chromatogram of the sample, and **b)** ion current chromatogram for the peak with m/z equal to 678 corresponding to the [TCEP + maleimide-DBCO + H]$^+$ adduct.

At this stage, since purification of the TCEP + maleimide-DBCO adduct proved to be unsuccessful, it was hypothesized that the formation of this adduct could be avoided by decreasing the amount of TCEP used and introducing a purification step between reduction and conjugation to maleimide-DBCO. Surprisingly, when this was attempted, full conversion to the modified H3R52C protein was achieved using the same number of equivalents of maleimide-DBCO as those used for H3K4C. By significantly decreasing the amount of TCEP and maleimide-DBCO, the intensity of the adduct peaks mentioned in the above paragraph was considerably reduced and mono-acetylation for the H3R52C protein was consistently replicated. Therefore, the amount of TCEP and maleimide-DBCO is critical for the success of the site-specific acetylation strategy when applied to the more internal K56 residue.

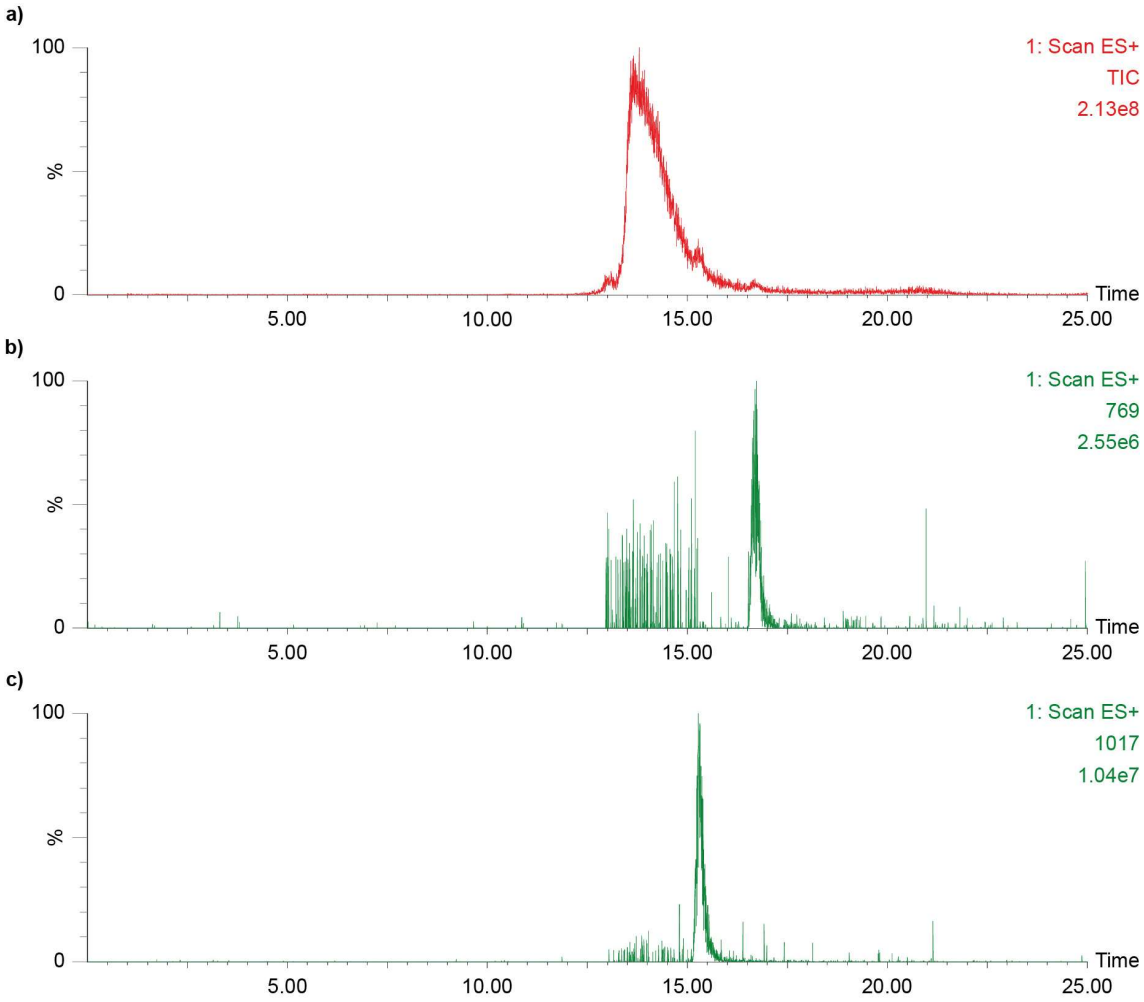


Figure 54. LC-MS obtained following incubation of the H3R52C-maleimide-DBCO protein (25 μM) with compound **4** (100 μM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 ºC and 400 rpm. **a)** total ion current chromatogram, **b)** ion current chromatogram for the peak with m/z equal to 769 corresponding to the [maleimide-DBCO + compound **4** + 3H]$^{+}$ adduct, and **c)** ion current chromatogram for the peak with m/z equal to 1017 corresponding to the [TCEP + maleimide-DBCO + compound **4** + H]$^{+}$ adduct.

3.2.7 Optimized reaction conditions for the site-specific acetylation at K9 of H3K4C and K56 of H3R52C

Considering the results obtained from all the experiments described so far in section 3.2, an optimized workflow for the proximity-driven, site-specific acetylation of the single-cysteine histone mutants H3K4C and H3R52C was envisioned.

The resuspended H3K4C and H3R52C proteins are first treated with a slight excess of TCEP to reduce any contaminant disulfides present and to increase the nucleophilicity of their cysteine residues. This initial step is critical towards the success of the acetylation reactions, particularly those depending on the conjugation to less reactive cysteines, as was the case of the H3R52C protein. For the latter, if incubation with TCEP is not performed, then the conjugation will require a higher number of maleimide-DBCO equivalents to achieve full conversion to the corresponding modified protein. This is important because purification of the resulting modified protein is harder to attain when using a greater excess of the reagent and if maleimide-DBCO is not removed to a large extent, then incubation with compound **4** will result in the formation of the

corresponding click adduct, which will subsequently promote the unspecific transfer of acetyl groups to various nucleophilic lysine residues.

Purification of the reduced protein is then performed by size-exclusion chromatography to remove as much as possible of TCEP. If this step is omitted, then reaction with maleimide-DBCO will also require a higher number of equivalents to achieve full conversion, since TCEP interacts with this reagent and thus interferes in its conjugation to cysteine. In addition, if purification of the protein following its reduction is bypassed, then the corresponding adduct of TCEP and maleimide-DBCO will be formed and its removal by ultrafiltration following cysteine conjugation will be harder to attain than the removal of the two reagents in separate steps.

Following reduction and purification, full conversion to the corresponding modified proteins was achieved using 2.5 equiv. of maleimide-DBCO for both H3K4C (**Figure 55**) and H3R52C (**Figure 56**). Here, purification of the cysteine-conjugated proteins is also important since incubation with compound **4** during the SPAAC reaction will form the corresponding adduct with maleimide-DBCO. As a result, more equivalents of compound **4** will be required to achieve full conversion to the corresponding SPAAC-containing product. This can promote the unspecific transfer of acetyl groups, particularly if the SPAAC reaction occurs under basic conditions. Furthermore, purification by ultrafiltration of the acetylated product from the adduct formed between maleimide-DBCO and compound **4** is harder to attain than purification of the corresponding modified proteins from these two reagents in separate steps. This matters because when increasing the reaction pH or leaving the unAcH3K4C** or unAcH3R52C** proteins on ice to promote acetylation, the maleimide-DBCO + compound **4** adduct that is resistant to purification by ultrafiltration can lead to the unspecific transfer of acetyl groups and thus jeopardize the site-specificity of the approach presented in this thesis.

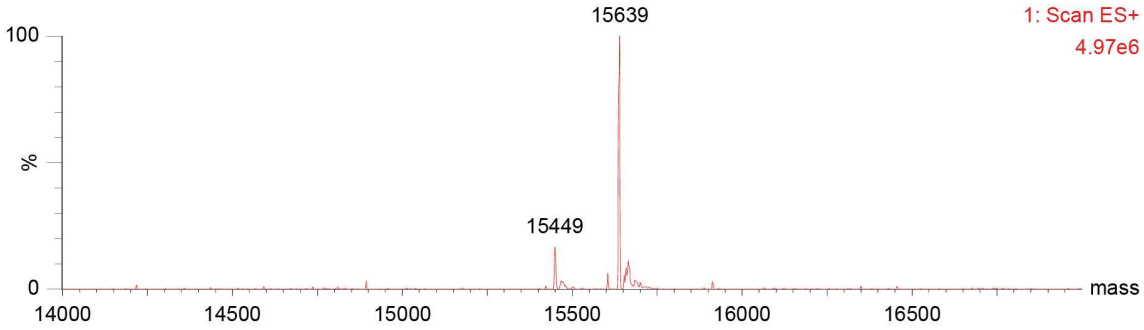


Figure 55. Deconvoluted ESI-MS spectrum obtained following incubation of the reduced and purified H3K4C protein (25 μM) with maleimide-DBCO (62.5 μM) in ammonium acetate buffer (20 mM, pH 7.0) for 1 h at 25 °C and 400 rpm. Major product identified: Protein H3K4C-maleimide-DBCO (calculated mass = 15641 Da; observed mass = 15639 Da).

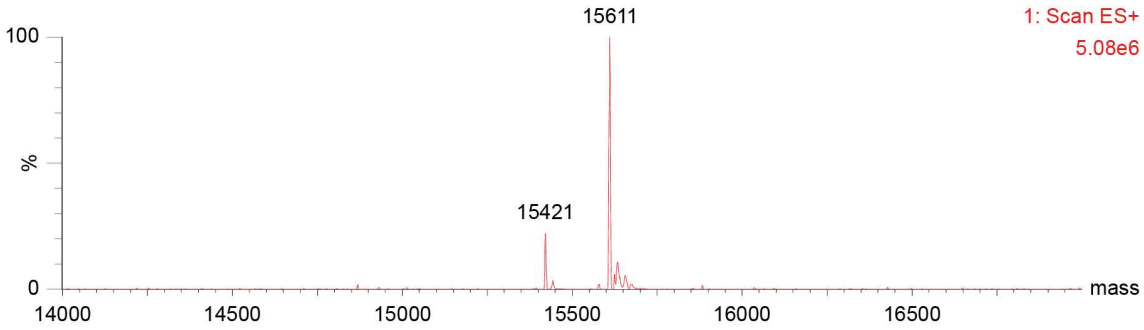


Figure 56. Deconvoluted ESI-MS spectrum obtained following incubation of the reduced and purified H3R52C protein (25 μM) with maleimide-DBCO (62.5 μM) in ammonium acetate buffer (20 mM, pH 7.0) for 1 h at 25 °C and 400 rpm. Major product identified: Protein H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15611 Da).

Treatment of the purified H3K4C-maleimide-DBCO protein with 4 equiv. of compound **4** under basic conditions resulted in the full conversion to the corresponding mono-acetylated product, H3K4C**K9Ac ($NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C, **Figure 57**). Under the same conditions, only half of the number of equivalents of compound **4** was required to achieve full conversion of the purified H3R52C-maleimide-DBCO to H3R52C**K56Ac ($NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C, **Figure 58**).

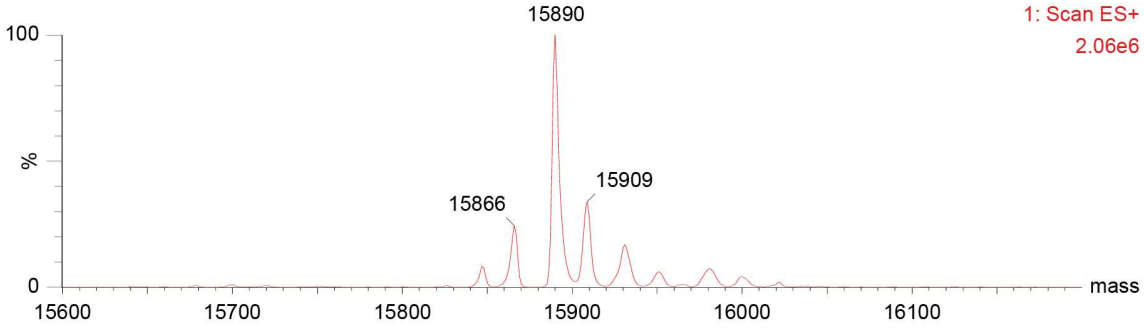


Figure 57. Deconvoluted ESI-MS spectrum obtained following incubation of the purified H3K4C-maleimide-DBCO protein (25 μM) with compound **4** (100 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed mass = 15890 Da).

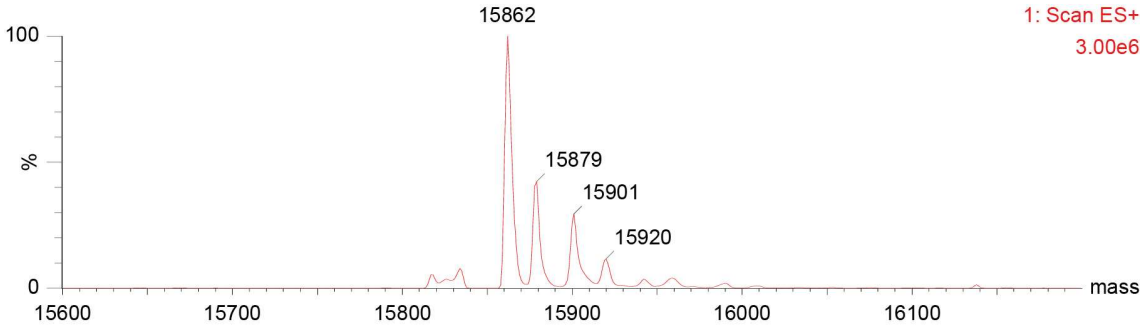


Figure 58. Deconvoluted ESI-MS spectrum obtained following incubation of the purified H3R52C-maleimide-DBCO protein (25 μM) with compound **4** (50 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed mass = 15862 Da).

This difference in the number of equivalents of compound **4** required for full conversion to the mono-acetylated product between the two modified proteins is likely related to the position of the cysteine within the H3 structure. The cysteine in the H3K4C protein is located in the N-terminal tail, which is a highly flexible and intrinsically disordered region. Curiously, histone H3 has the longest N-terminal tail out of all histones.[303] Meanwhile, the cysteine in the H3R52C protein is located in a more structured and fixed region of H3, likely a β-strand.[304] Any chemical reaction depends on the encounter between reagents, with certain orientations and angles of approach being preferred over others. Under the same reaction conditions, if maleimide-DBCO is attached to a protein region that can adopt a multitude of conformational states, as is the case for H3K4C, then it is logical that its encounter with compound **4** has lower probability of occurrence than if the same reagent is attached to another protein region that can only adopt a restricted number of conformational states. Thus, for full conversion to be achieved in the case of H3K4C-maleimide-DBCO, it follows that the number of equivalents of compound **4** (or possibly time and temperature, although this was not tested) will have to be higher than in the case of the H3R52C-maleimide-DBCO protein.

The SPAAC reaction was also performed under acidic conditions, with full conversion achieved by maintaining all the other experimental conditions constant ($NH_4^+CH_3CO_2^-$ 20 mM, pH 4.9, 30 min at 25 °C, **Figure 59** and **Figure 60**).

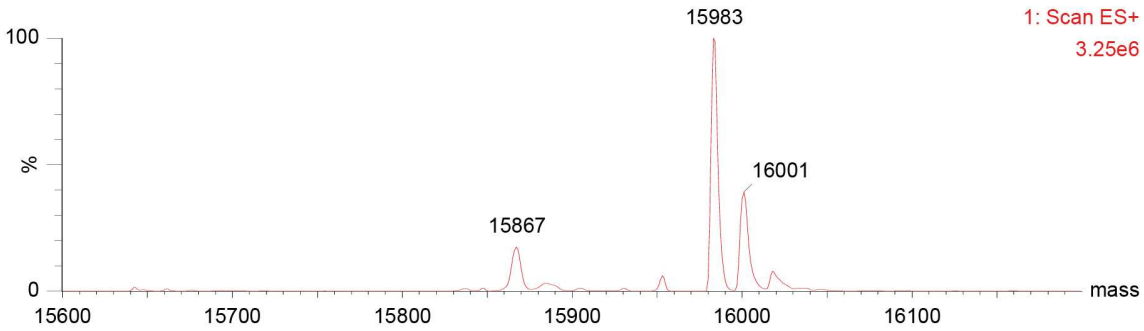


Figure 59. Deconvoluted ESI-MS spectrum obtained following incubation of the purified H3K4C-maleimide-DBCO protein (25 μM) with compound **4** (100 μM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 °C and 400 rpm. Major product identified: Protein unAcH3K4C** (calculated mass = 15981 Da; observed mass = 15983 Da).

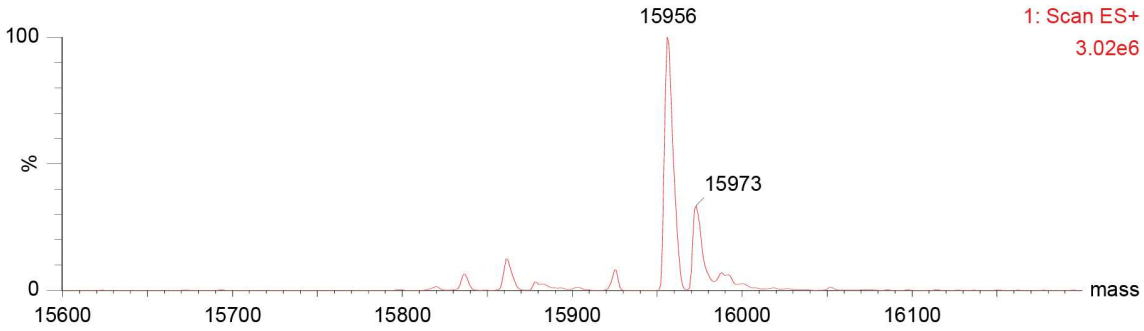


Figure 60. Deconvoluted ESI-MS spectrum obtained following incubation of the purified H3R52C-maleimide-DBCO protein (25 μM) with compound **4** (50 μM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 °C and 400 rpm. Major product identified: Protein unAcH3R52C** (calculated mass = 15953 Da; observed mass = 15956 Da).

Following purification by ultrafiltration to remove excess compound **4**, the unAcH3K4C** and unAcH3R52C** proteins were exchanged to the same buffer but at a slightly less acidic pH ($NH_4^+CH_3CO_2^-$ 20 mM, pH 5.8). This afforded the H3K4C**K9Ac and H3R52C**K56Ac proteins, respectively, after 3 to 4 days of incubation on ice (**Figure 61** and **Figure 62**). These two mono-acetylated proteins were desalted, lyophilized and stored at -80 °C to perform the experiments described in subsections 3.6 and 3.7.

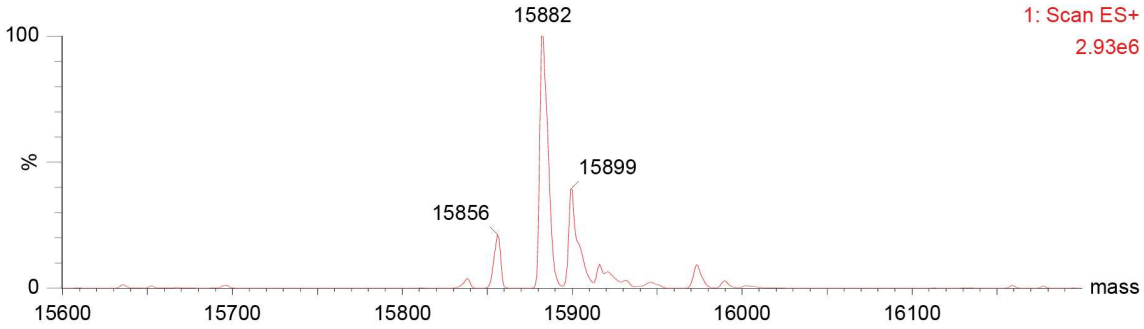


Figure 61. Deconvoluted ESI-MS spectrum of the H3K4C**K9Ac protein (25 µM) obtained following buffer exchange by ultrafiltration of unAcH3K4C** to ammonium acetate buffer (20 mM, pH 5.8) and incubation on ice for 3 to 4 days. Major product identified: Protein H3K4C**K9Ac (calculated mass = 15889 Da; observed = 15882 Da).

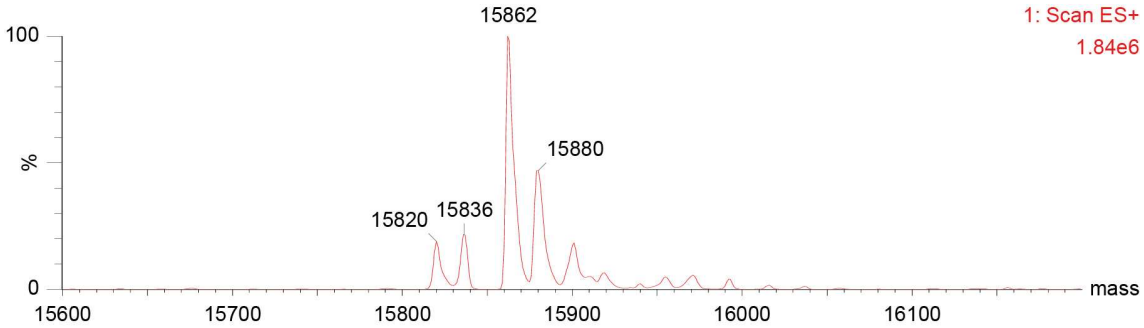


Figure 62. Deconvoluted ESI-MS spectrum of the H3R52C**K9Ac protein (25 µM) obtained following buffer exchange by ultrafiltration of unAcH3R52C** to ammonium acetate buffer (20 mM, pH 5.8) and incubation on ice for 3 to 4 days. Major product identified: Protein H3R52C**K56Ac (calculated mass = 15861 Da; observed = 15862 Da).

3.2.8 Expression and purification of the H3K4CR52C protein

Having fully optimized the reaction conditions for the site-specific acetylation of the N-terminal K9 residue of H3K4C and the more internal K56 residue of H3R52C, the next step in the project consisted of evaluating whether this strategy could also be used to create a dual acetylated histone. To test this possibility, the two residues for which mono-acetylation had previously been achieved were chosen as targets. In this context, the double-cysteine mutant H3K4CR52C (C110A) was generated, recombinantly expressed and purified.

The purification of H3K4CR52C was performed according to the same optimized protocol described previously for the H3K4C and H3R52C proteins. The deconvoluted ESI-MS spectrum of the purified H3K4CR52C protein is shown in **Figure 63**. As determined by LC-MS, the purity level of the unoxidized H3K4CR52C protein was approximately 70%. One notable contaminant present in the spectrum likely corresponds to the oxidized version of the H3K4CR52C protein.

Commercial H3 displays slightly higher purity levels, however these are estimated by SDS-PAGE and thus do not discriminate between unoxidized and oxidized protein, with many also having an affinity tag to facilitate its purification. Considering an overall purity of 70%, a typical H3K4CR52C purification yielded approximately 9.5 mg of protein per liter of culture.

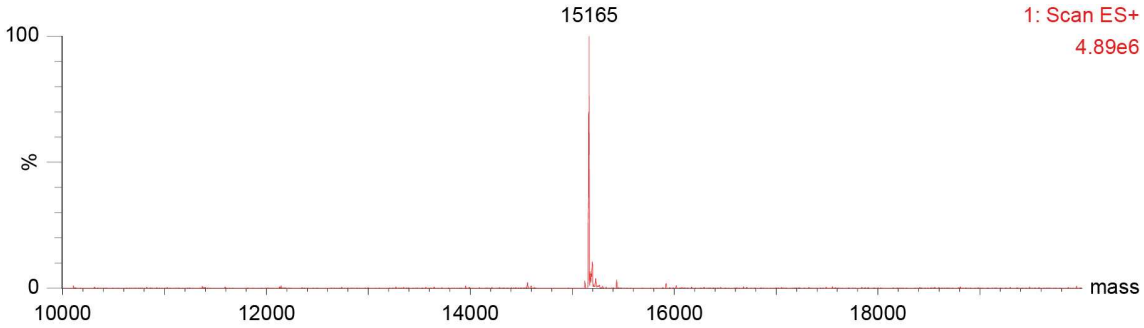


Figure 63. Deconvoluted ESI-MS spectrum of the purified H3K4CR52C protein (25 μM) in Milli-Q water. Major product identified: Protein H3K4CR52C (calculated mass = 15161 Da; observed mass = 15165 Da).

3.2.9 Application of the optimized reaction conditions for the dual site-specific acetylation at K9 and K56 of H3K4CR52C

The dual site-specific acetylation of the H3K4CR52C mutant was attempted using similar reaction conditions to those previously optimized for the mono-acetylation at K9 and K56 of H3K4C and H3R52C, respectively. Following prior reduction with TCEP and removal of this reducing agent by size-exclusion chromatography, conjugation of H3K4CR52C to maleimide-DBCO with full conversion to the corresponding modified protein was achieved using the same optimized conditions as those described for H3K4C and H3R52C (2.5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 1 h at 25 °C, **Figure 64**).

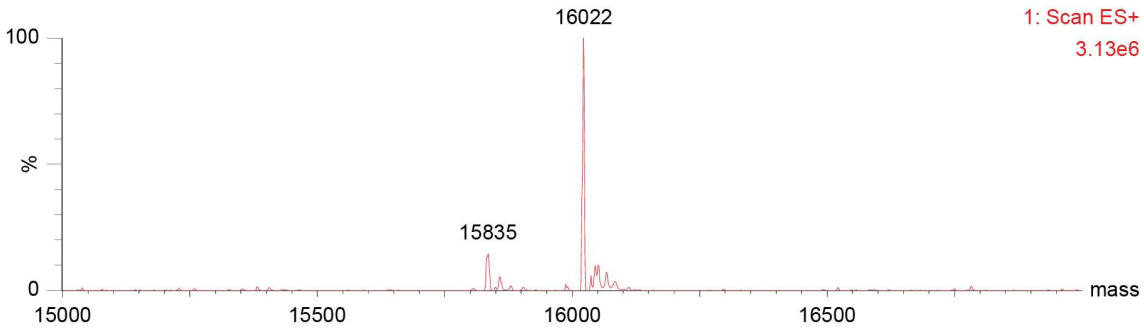


Figure 64. Deconvoluted ESI-MS spectrum obtained following incubation of the reduced and purified H3K4CR52C protein (25 μM) with maleimide-DBCO (62.5 μM) in ammonium acetate buffer (20 mM, pH 7.0) for 1 h at 25 °C and 400 rpm. Major product identified: Protein H3K4CR52C-maleimide-DBCO (calculated mass = 16016 Da; observed mass = 16022 Da).

Following purification by ultrafiltration, the dual site-specific acetylation was attempted by treating the modified protein with compound **4**. Here, a slightly longer incubation of 45 instead of 30 minutes with the reagent under basic conditions was necessary to generate the corresponding

di-acetylated product, henceforth referred to as H3K4C**K9AcR52C**K56Ac (4 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 45 min at 25 °C, **Figure 65**).

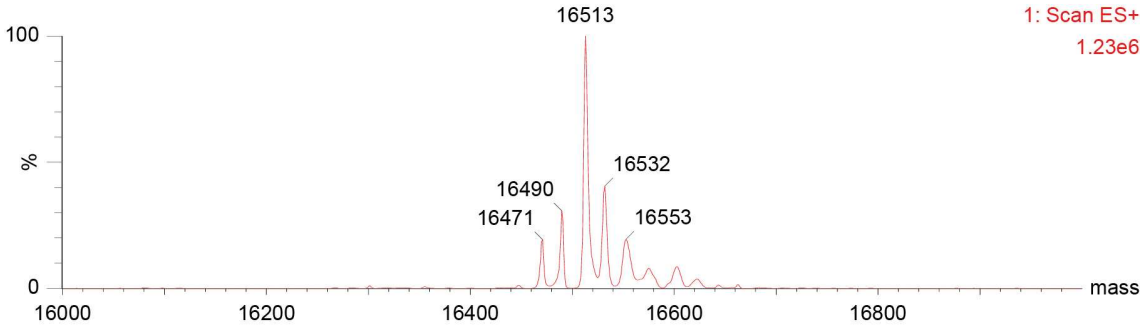


Figure 65. Deconvoluted ESI-MS spectrum obtained following incubation of the purified H3K4CR52C-maleimide-DBCO protein (25 µM) with compound **4** (100 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 45 min at 25 °C and 400 rpm. Major product identified: Protein H3K4C**K9AcR52C**K56Ac (calculated mass = 16510 Da; observed mass = 16513 Da).

Performing the SPAAC reaction first under acidic conditions (4 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 4.9, 45 min at 25 °C, **Figure 66**), removing the acetylating reagent and slightly increasing the pH of the buffer ($NH_4^+CH_3CO_2^-$ 20 mM, pH 5.8) by ultrafiltration also yielded the desired H3K4C**K9AcR52C**K56Ac product after 3 to 4 days of incubation on ice (**Figure 67**).

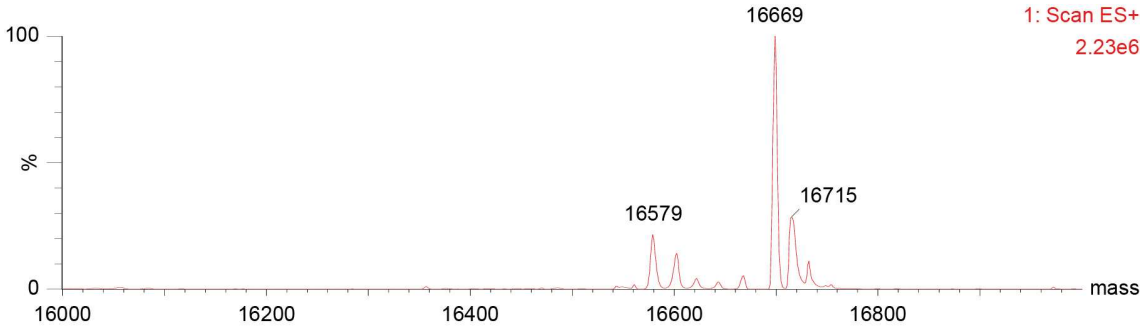


Figure 66. Deconvoluted ESI-MS spectrum obtained following incubation of the purified H3K4CR52C-maleimide-DBCO protein (25 µM) with compound **4** (100 µM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 °C and 400 rpm. Major product identified: Protein unAcH3K4C**R52C** (calculated mass = 16695 Da; observed mass = 16669 Da).

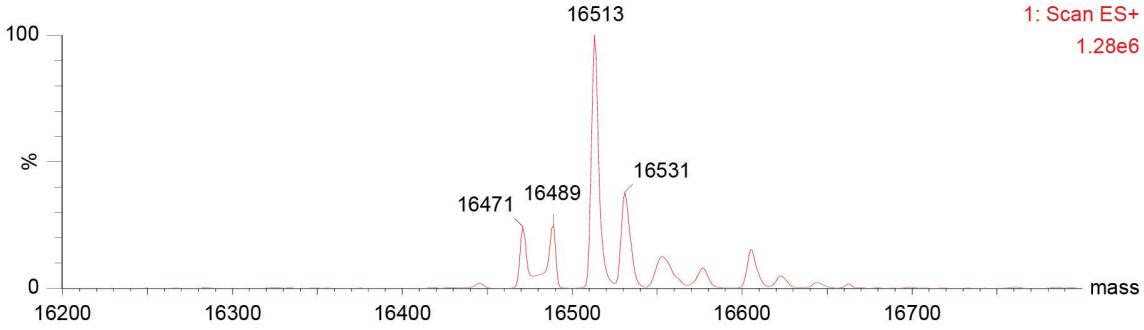


Figure 67. Deconvoluted ESI-MS spectrum of the H3K4C**K9AcR52C**K56Ac protein (25 µM) obtained following buffer exchange by ultrafiltration of unAcH3K4C**R52C** to ammonium acetate buffer (20 mM, pH 5.8) and incubation on ice for 3 to 4 days. Major product identified: Protein H3K4C**K9AcR52C**K56Ac (calculated mass = 16510 Da; observed mass = 16513 Da).

3.3 Oxidation of cysteine-containing histone proteins

Throughout the project, particularly during the optimization phase of the protein chemical reactions, oxidation of the H3K4C and H3R52C histone mutants was observed. While the H3K4C protein preferentially oxidized to form the corresponding +16 and +32 adducts, indicative of the addition of one and two oxygen atoms respectively (**Figure 68**), H3R52C oxidation favored predominantly the formation of its disulfide (**Figure 69**). This distinct oxidation behavior between both proteins might reflect differences in the solvent accessibility of their cysteine residues. The cysteine in the H3K4C protein is closer to the N-terminus, making it relatively more solvent-exposed than its counterpart in H3R52C, where the residue is more internal and thus solvent-protected by the protein's structure.

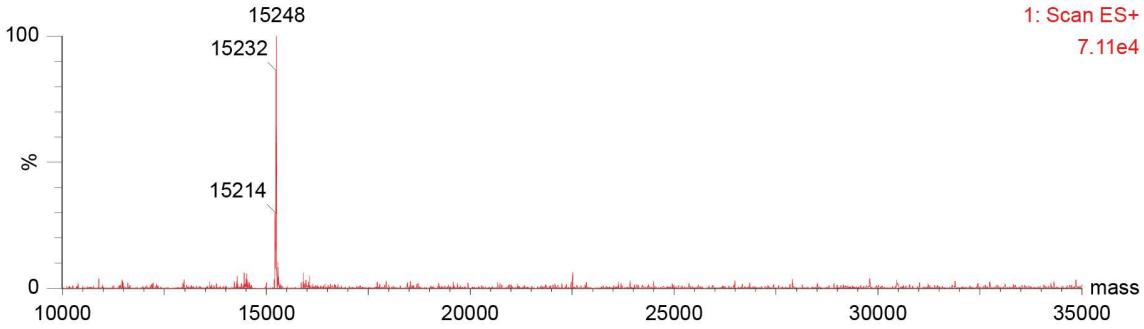


Figure 68. Deconvoluted ESI-MS spectrum showing oxidation of the H3K4C protein (8 µM) in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Double-oxidized H3K4C protein (calculated mass = 15246 Da; observed mass = 15248 Da).

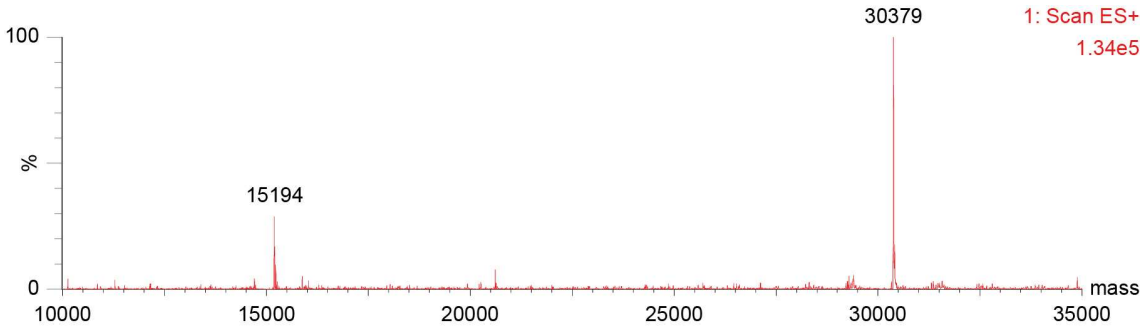


Figure 69. Deconvoluted ESI-MS spectrum showing disulfide formation of the H3R52C protein (25 µM) with time. Major product identified: H3R52C disulfide (calculated mass = 30370 Da; observed mass = 30379 Da).

The addition of oxygen can take place in residues other than cysteine, namely methionine of which two are present in H3. However, on one occasion it was possible to experimentally confirm that this type of oxidation had indeed occurred at the cysteine by observing that the oxidized H3K4C protein did not react with Ellman's reagent (**Figure 70**). Thus, cysteine oxidation occasionally hindered the success of conjugation reactions, particularly in the case of H3K4C.

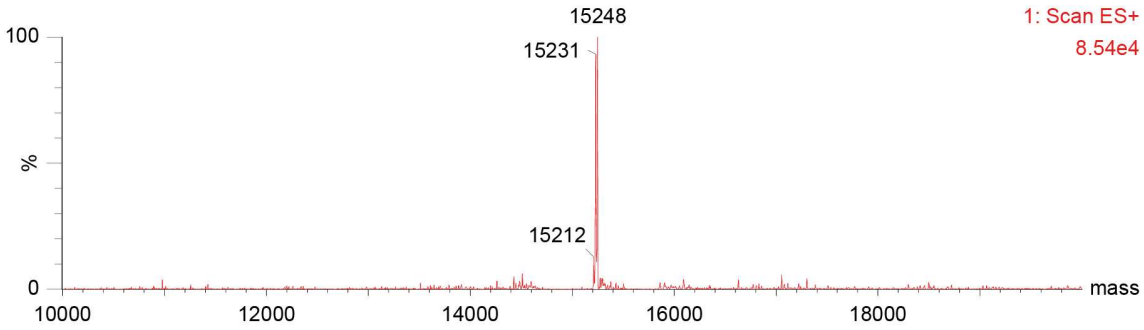


Figure 70. Deconvoluted ESI-MS spectrum confirming cysteine oxidation of the H3K4C protein (8 μM) following incubation with Ellman's reagent (100 equiv.) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 ºC. Major product identified: Double-oxidized H3K4C protein (calculated mass = 15246 Da; observed mass = 15248 Da).

Reduction with TCEP of the oxidized H3K4C (**Figure 71**) and H3R52C proteins was also evaluated, with differing degrees of success. For the H3K4C protein, significant reversal of its oxidation status was achieved by incubating it with a large excess of TCEP (100 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 12 h at 25 ºC, **Figure 72**). In the case of the H3R52C protein, a much milder treatment with TCEP was sufficient to promote disulfide reduction (5 equiv., $NH_4CH_3CO_2^-$ 20 mM, pH 8.0, 30 min, **Figure 73**).

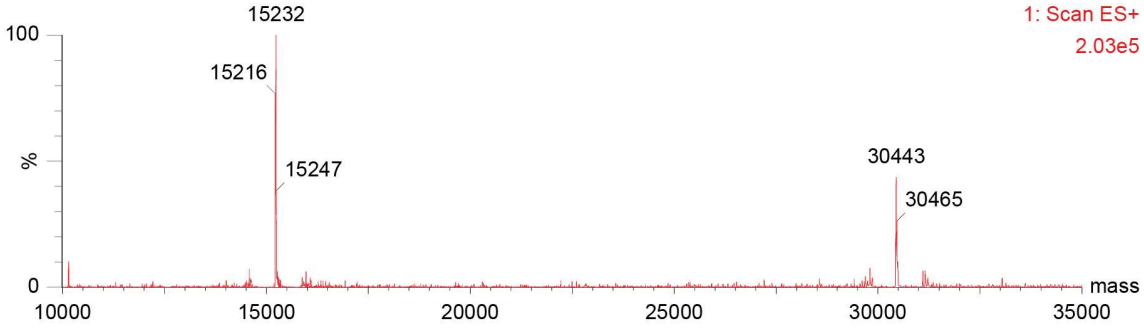


Figure 71. Deconvoluted ESI-MS spectrum showing oxidation of the H3K4C protein (25 μM) in ammonium acetate buffer (20 mM, pH 7.0). Major product identified: Mono-oxidized H3K4C protein (calculated mass = 15230 Da; observed mass = 15232 Da).

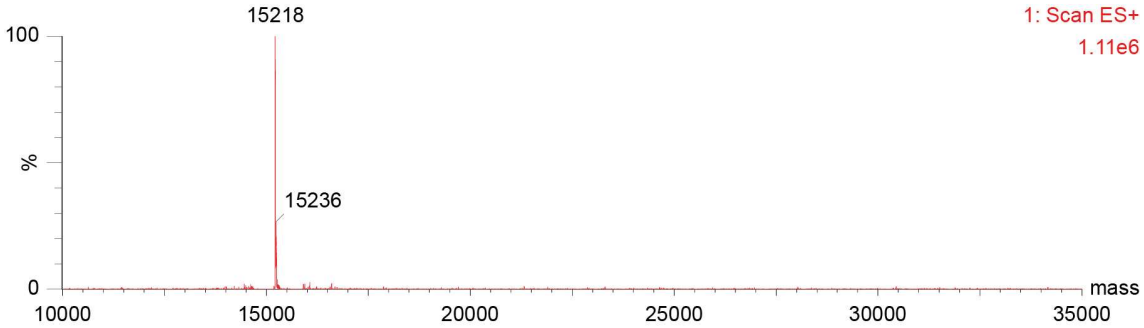


Figure 72. Deconvoluted ESI-MS spectrum obtained following incubation of the oxidized H3K4C protein (25 μM) with TCEP (2.5 mM) in ammonium acetate buffer (20 mM, pH 7.0) for 12 h at 25 ºC and 400 rpm. Major product identified: Protein H3K4C (calculated mass = 15214 Da; observed mass = 15218 Da).

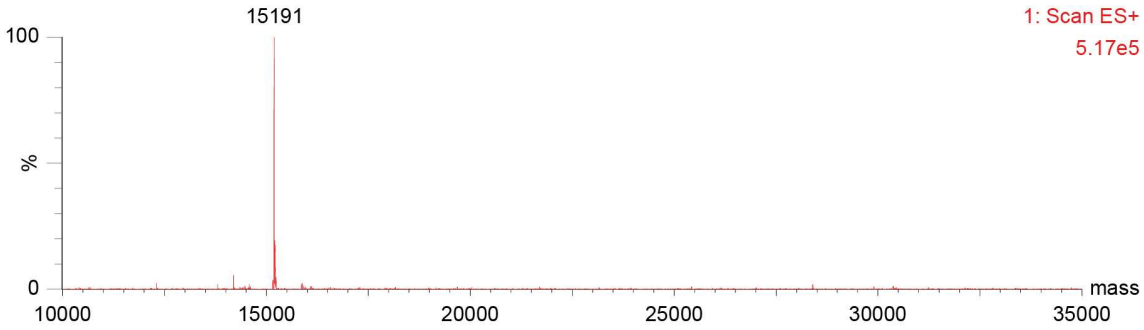


Figure 73. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C disulfide (25 μM) with TCEP (125 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3R52C (calculated mass = 15186 Da; observed mass = 15191 Da).

3.4 Control reactions of the cysteine-assisted click-chemistry strategy

3.4.1 Control reactions (1) with a maleimide-dummy

To establish that the SPAAC is essential to orient the acetyl group towards the nearest lysine and thus for acetylation to occur within the timeframe of the click reaction under basic conditions, the cysteine-assisted strategy was attempted in the various histone mutants using a maleimide reagent without a SPAAC clickable handle. This maleimide-dummy reagent, N-(5-nitro-ortho-tolyl)-maleimide, was chosen because it had previously been used within the group to perform the corresponding control reactions in peptides and was readily available in the laboratory. Thus, the H3K4C, H3R52C and H3K4CR52C proteins were reacted with maleimide-dummy using the optimized conditions described previously for maleimide-DBCO conjugation (2.5 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 1 h at 25 °C). Full conversion to the H3K4C (**Figure 74**), H3R52C (**Figure 75**) and H3K4CR52C (**Figure 76**) proteins containing the maleimide-dummy reagent was observed.

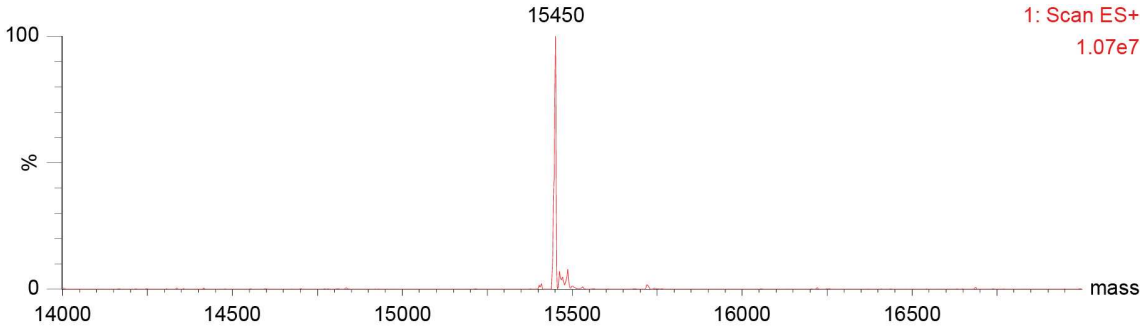


Figure 74. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C protein (25 μM) with maleimide-dummy (62.5 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 °C and 400 rpm. Major product identified: Protein H3K4C-maleimide-dummy (calculated mass = 15446 Da; observed mass = 15450 Da)

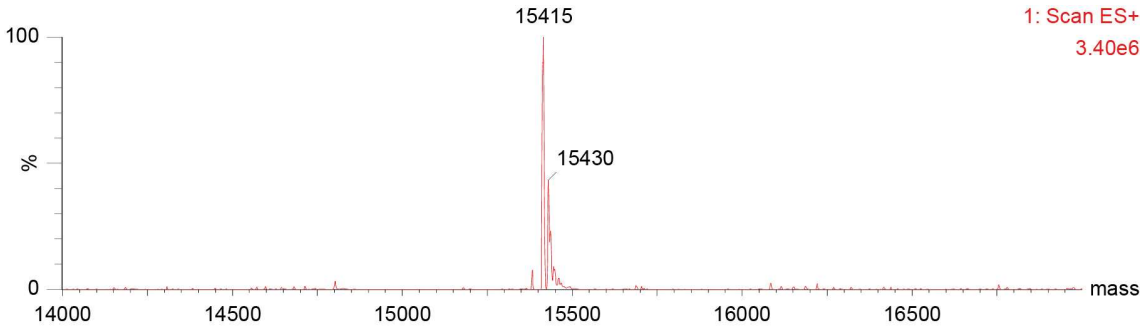


Figure 75. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C protein (25 μM) with maleimide-dummy (62.5 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 ºC and 400 rpm. Major product identified: Protein H3R52C-maleimide-dummy (calculated mass = 15418 Da; observed mass = 15415 Da).

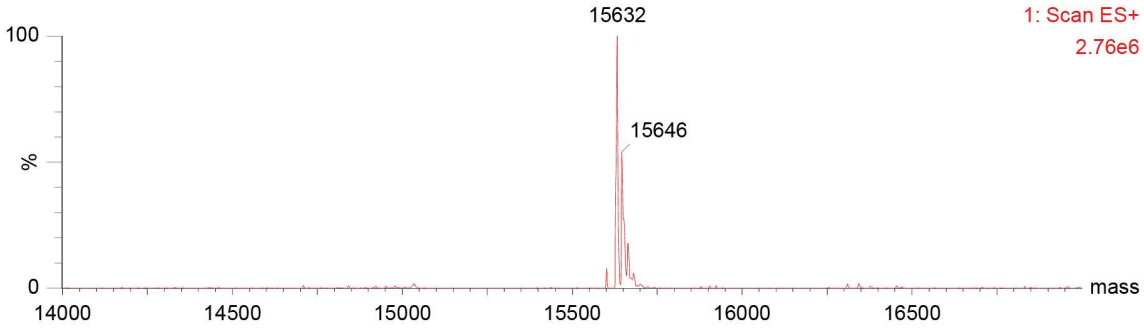


Figure 76. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4CR52C protein (25 μM) with maleimide-dummy (62.5 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 1 h at 25 ºC and 400 rpm. Major product identified: Protein H3K4CR52C-maleimide-dummy (calculated mass = 15625 Da; observed mass = 15632 Da).

Interestingly, following purification to remove the excess reagent and dilution in ammonium acetate buffer (20 mM, pH 8.0), the main peak observed in the spectra of the three modified proteins corresponds to the respective product with an unexpected mass shift, which is likely due to maleimide hydrolysis promoted by the strong electron withdrawing nature of the nitro group (**Figure 77**, **Figure 78** and **Figure 79**).

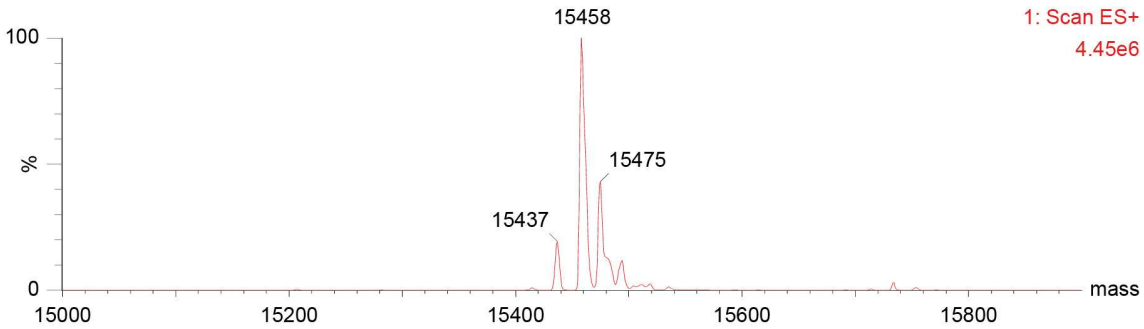


Figure 77. Deconvoluted ESI-MS spectrum of the purified H3K4C-maleimide-dummy protein (25 μM) in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Protein H3K4C-hydrolyzed-maleimide-dummy (calculated mass = 15464 Da; observed mass = 15458 Da).

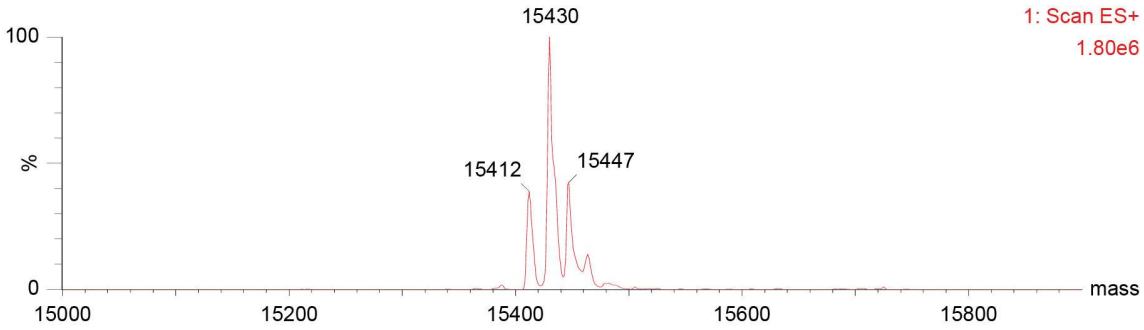


Figure 78. Deconvoluted ESI-MS spectrum of the purified H3R52C-maleimide-dummy protein (25 µM) in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Protein H3R52C-hydrolyzed-maleimide-dummy (calculated mass = 15436 Da; observed mass = 15430 Da).

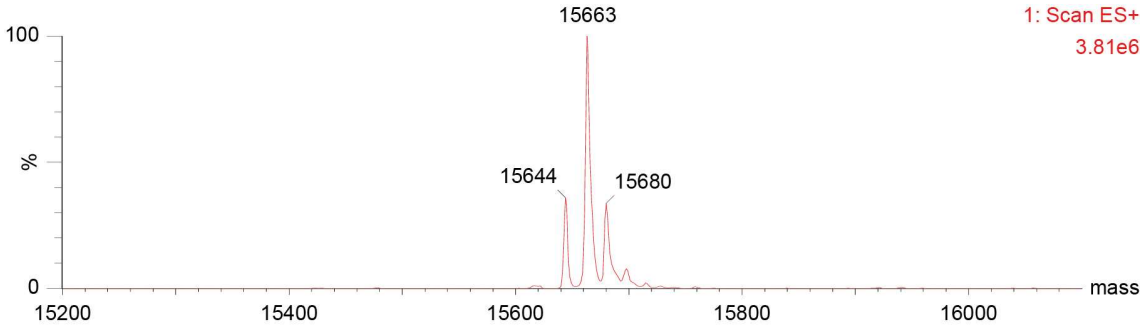


Figure 79. Deconvoluted ESI-MS spectrum of the purified H3K4CR52C-maleimide-dummy protein (25 µM) in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Protein H3K4CR52C-hydrolyzed-maleimide-dummy (calculated mass = 15661 Da; observed mass = 15663 Da).

Subsequently, the products bearing the hydrolyzed maleimide-dummy were incubated with compound **4** at a basic pH using the same optimized SPAAC reaction conditions as those described previously for proteins modified with maleimide-DBCO. Acetylation was not observed in any of the three maleimide-dummy-modified proteins (**Figure 80**, **Figure 81** and **Figure 82**).

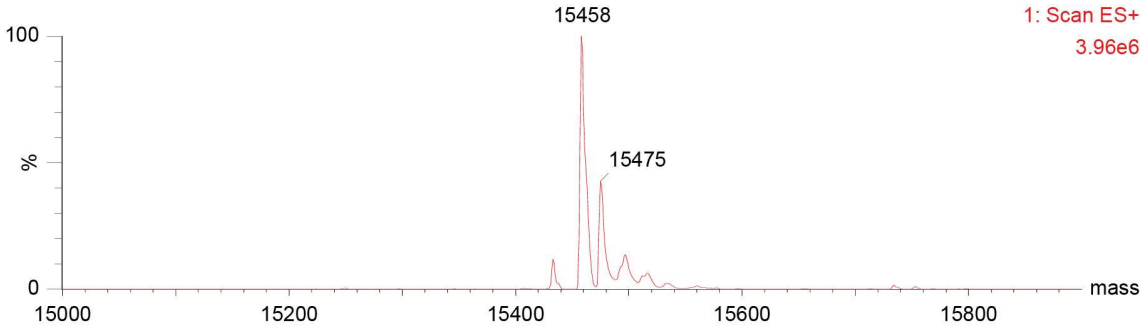


Figure 80. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-hydrolyzed-maleimide-dummy protein (25 µM) with compound **4** (100 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3K4C-hydrolyzed-maleimide-dummy (calculated mass = 15464 Da; observed mass = 15458 Da).

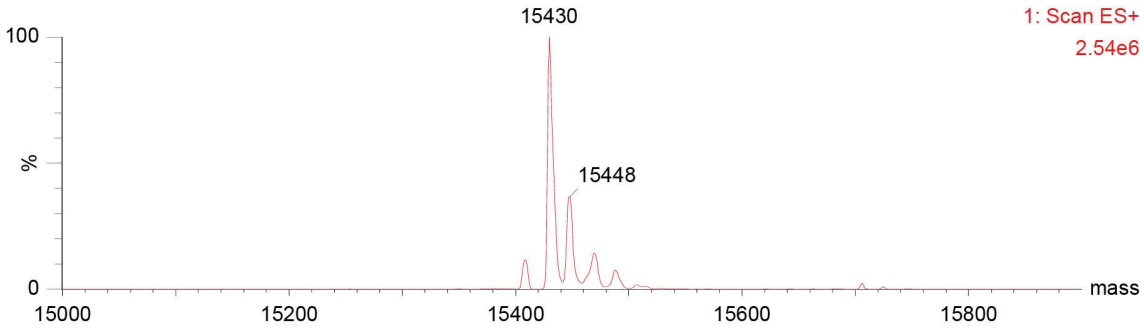


Figure 81. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C-hydrolyzed-maleimide-dummy protein (25 μM) with compound **4** (50 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3R52C-hydrolyzed-maleimide-dummy (calculated mass = 15436 Da; observed mass = 15430 Da).

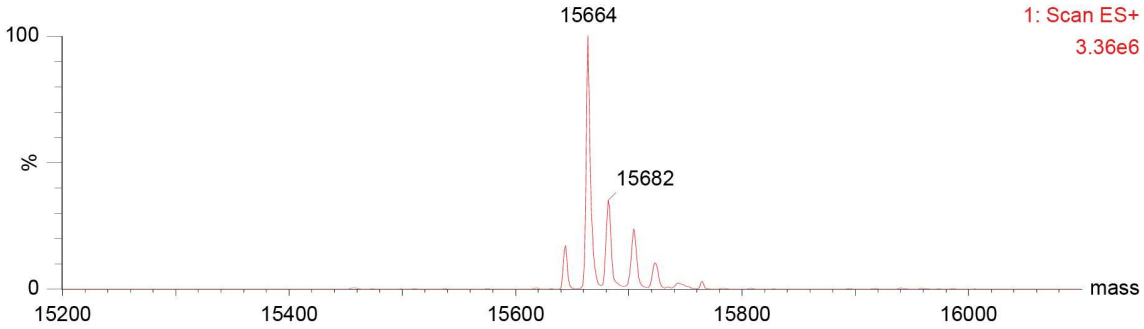


Figure 82. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4CR52C-hydrolyzed-maleimide-dummy protein (25 μM) with compound **4** (100 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 45 min at 25 °C and 400 rpm. Major product identified: Protein H3K4CR52C-hydrolyzed-maleimide-dummy (calculated mass = 15661 Da; observed mass = 15664 Da).

3.4.2 Control reactions (2) with Ellman's reagent to confirm cysteine availability before and after maleimide conjugation

Evaluation of cysteine availability before maleimide conjugation

Ellman's reagent is an aromatic disulfide routinely employed to determine total sulfhydryl content in proteins. Thus, histone mutants were incubated with Ellman's reagent (10 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C) to establish that one (in the case of H3K4C and H3R52C) or two (in the case of H3K4CR52C) free cysteines were present and available for chemical conjugation. Unsurprisingly, full conversion was observed to the corresponding Ellman disulfides of H3K4C (**Figure 83**), H3R52C (**Figure 84**) and H3K4CR52C (**Figure 85**).

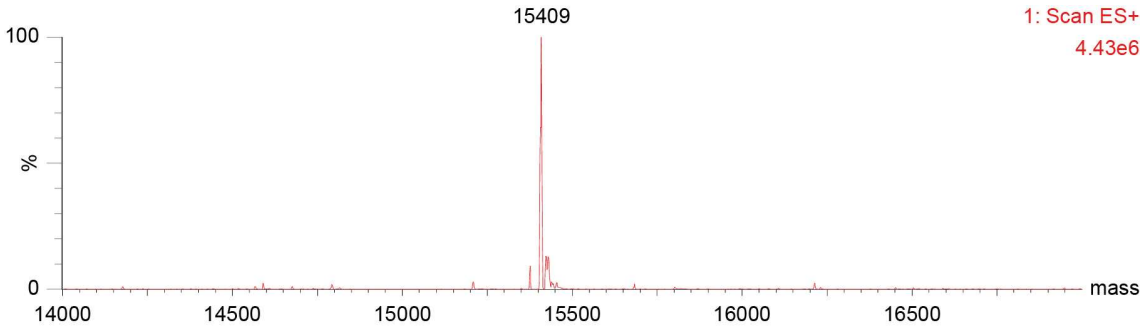


Figure 83. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C protein (25 μM) with Ellman's reagent (250 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3K4C-Ellman's (calculated mass = 15411 Da; observed mass = 15409 Da).

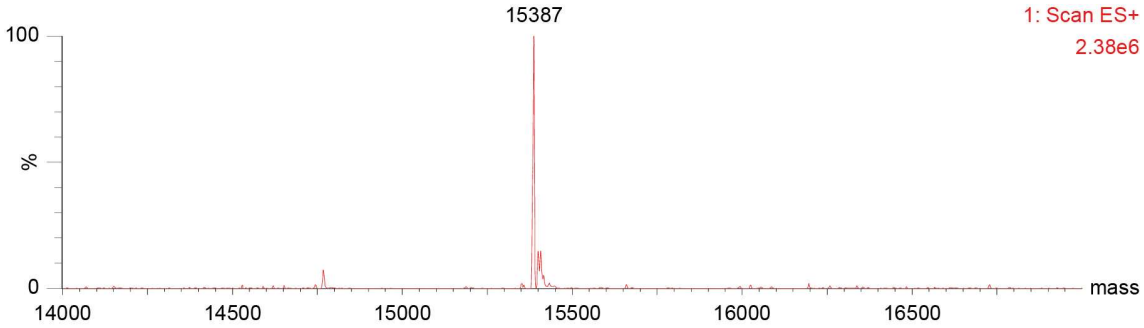


Figure 84. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C protein (25 μM) with Ellman's reagent (250 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3R52C-Ellman's (calculated mass = 15383 Da; observed mass = 15387 Da).

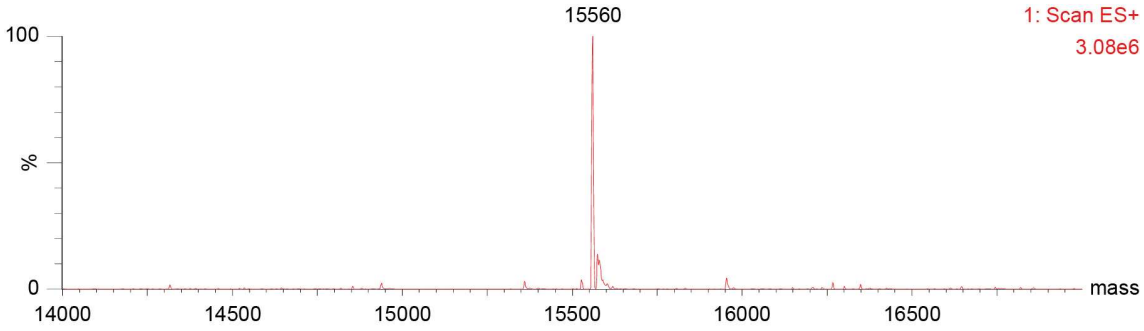


Figure 85. Deconvoluted ESI-MS spectrum obtained following incubation of the H34K4CR52C protein (25 μM) with Ellman's reagent (250 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3K4CR52C-Ellman's (calculated mass = 15557 Da; observed mass = 15560 Da).

Evaluation of cysteine availability after maleimide conjugation

The H3K4C, H3R52C and H3K4CR52C proteins bearing the two maleimides described in this work (maleimide-DBCO and maleimide-dummy) were incubated with Ellman's reagent (10 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C) to confirm that conjugation had occurred at the free cysteine residues of these histones. Expectedly, conversion to the corresponding Ellman disulfides was not observed, indicating that maleimide conjugation had occurred at the cysteine residues of H3K4C (**Figure 86** and **Figure 87**), H3R52C (**Figure 88** and **Figure 89**) and H3K4CR52C (**Figure 90** and **Figure 91**).

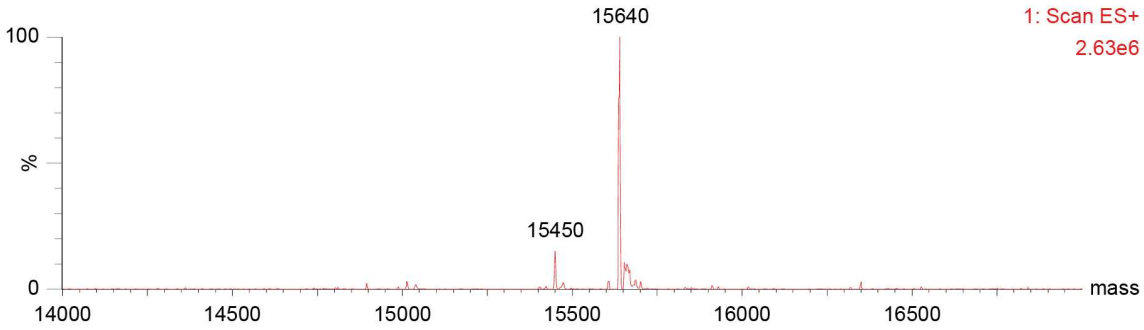


Figure 86. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-maleimide-DBCO protein (25 µM) with Ellman's reagent (250 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3K4C-maleimide-DBCO (calculated mass = 15641 Da; observed mass = 15640 Da).

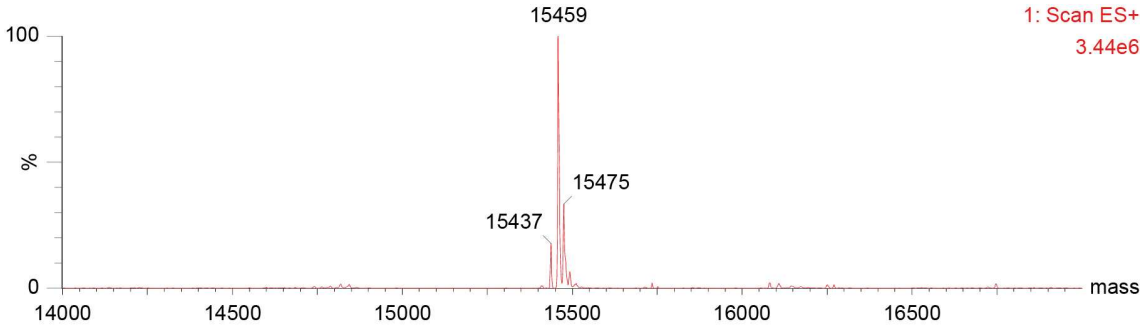


Figure 87. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-hydrolyzed-maleimide-dummy protein (25 µM) with Ellman's reagent (250 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C at 400 rpm. Major product identified: Protein H3K4C-hydrolyzed-maleimide-dummy (calculated mass = 15464 Da; observed mass = 15459 Da).

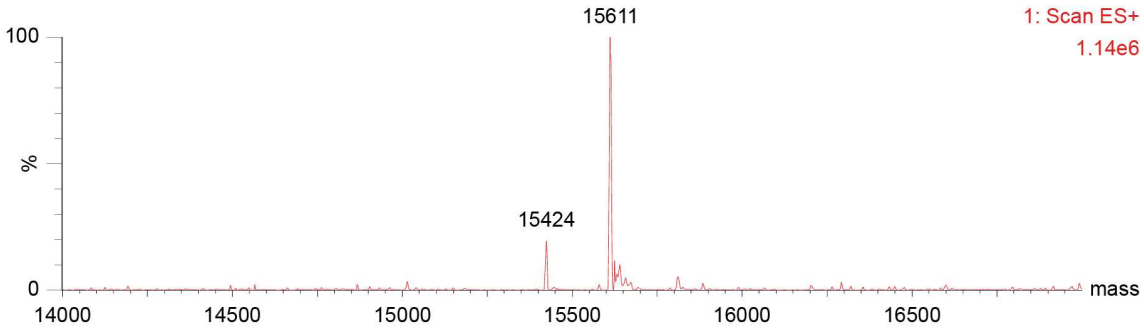


Figure 88. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C-maleimide-DBCO protein (25 µM) with Ellman's reagent (250 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3R52C-maleimide-DBCO (calculated mass = 15613 Da; observed mass = 15611 Da).

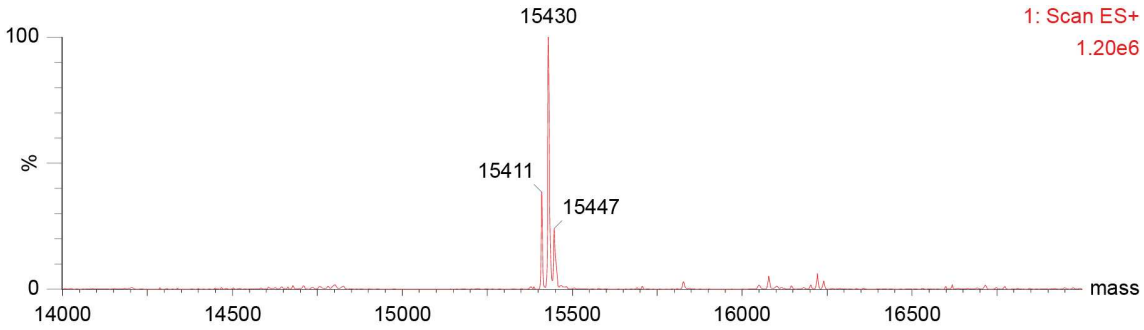


Figure 89. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C-hydrolyzed-maleimide-dummy protein (25 µM) with Ellman's reagent (250 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 ºC at 400 rpm. Major product identified: Protein H3R52C-hydrolyzed-maleimide-dummy (calculated mass = 15436 Da; observed mass = 15430 Da).

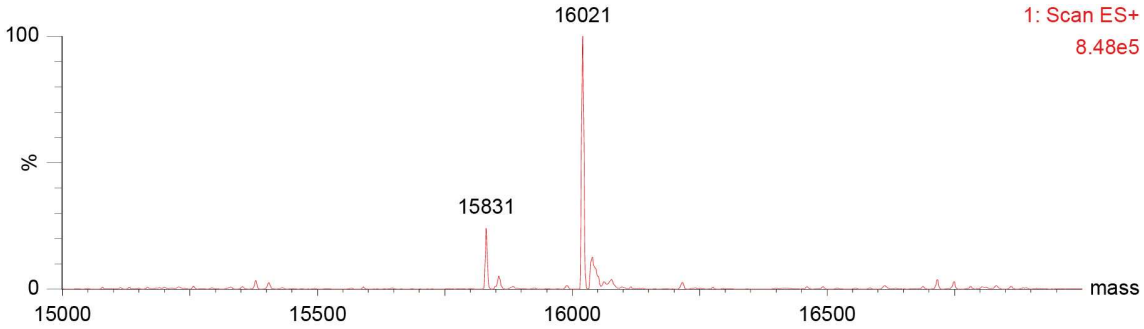


Figure 90. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4CR52C-maleimide-DBCO protein (25 µM) with Ellman's reagent (250 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 ºC and 400 rpm. Major product identified: Protein H3K4CR52C-maleimide-DBCO (calculated mass = 16016 Da; observed mass = 16021 Da).

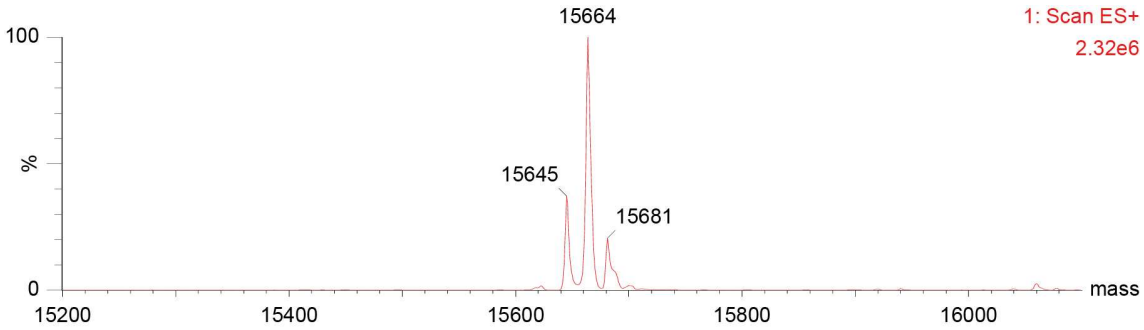


Figure 91. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4CR52C-hydrolyzed-maleimide-dummy protein (25 µM) with Ellman's reagent (250 µM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 ºC at 400 rpm. Major product identified: Protein H3K4CR52C-hydrolyzed-maleimide-dummy (calculated mass = 15661 Da; observed mass = 15664 Da).

3.4.3 Control reactions with the ubiquitin UbK63C protein

To demonstrate that acetylation occurring after the SPAAC reaction is dependent on the existence of a proximal lysine near the chemically modified cysteine, the same set of reactions was attempted in the globular UbK63C protein. This ubiquitin mutant contains six lysine residues, none of which is proximal to its single cysteine, thus making it an ideal control protein to test the mentioned hypothesis that SPAAC-dependent acetylation requires spatial proximity between the maleimide-DBCO-modified cysteine and the target lysine (**Figure 92**).

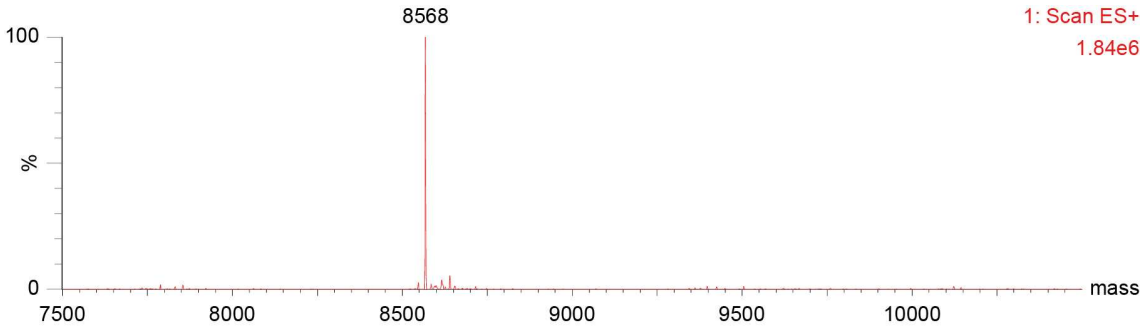


Figure 92. Deconvoluted ESI-MS spectrum of the UbK63C protein (50 μM) in ammonium acetate buffer (20 mM, pH 8.0). Major product identified: Protein UbK63C (calculated mass = 8567 Da; observed mass = 8568 Da).

Initially, similar experimental conditions to the ones used at the beginning of the H3K4C conjugation experiments were employed for the UbK63C protein. However, these were not sufficient to promote full conversion to the corresponding UbK63C-maleimide-DBCO product. To achieve complete conversion, UbK63C was first treated with TCEP (4 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C) to increase the nucleophilicity of its cysteine residue. Addition of a large excess of maleimide-DBCO (374 equiv.) to the reduced protein afforded the expected UbK63C-maleimide-DBCO product (**Figure 93**).

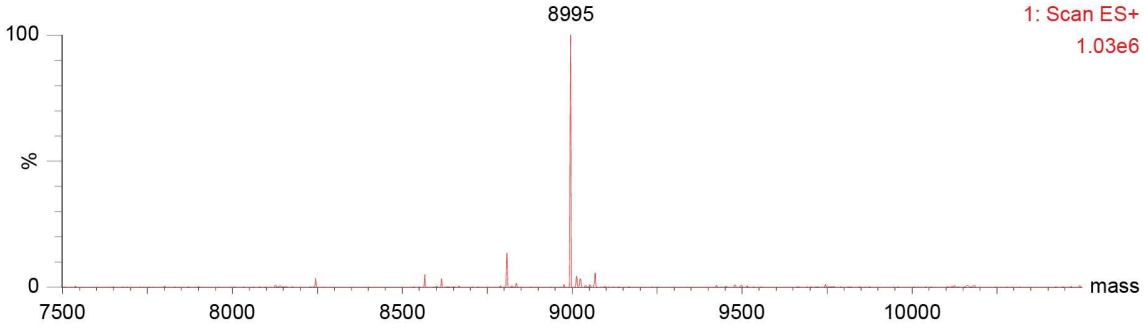


Figure 93. Deconvoluted ESI-MS spectrum obtained following incubation of the UbK63C protein (50 μM) with TCEP (200 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 °C and 400 rpm, with subsequent addition of a large excess of maleimide-DBCO (18.7 mM). Major product identified: Protein UbK63C-maleimide-DBCO (calculated mass = 8994 Da; observed mass = 8995 Da).

The SPAAC reaction was then attempted by treating the modified UbK63C-maleimide-DBCO protein with compound **4** (1 equiv., $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C) and the expected unacetylated UbK63C SPAAC-containing product with the *gem*-dithioacetate group still left to react, henceforth referred to as unAcUbK63C**, was obtained (**Figure 94**).

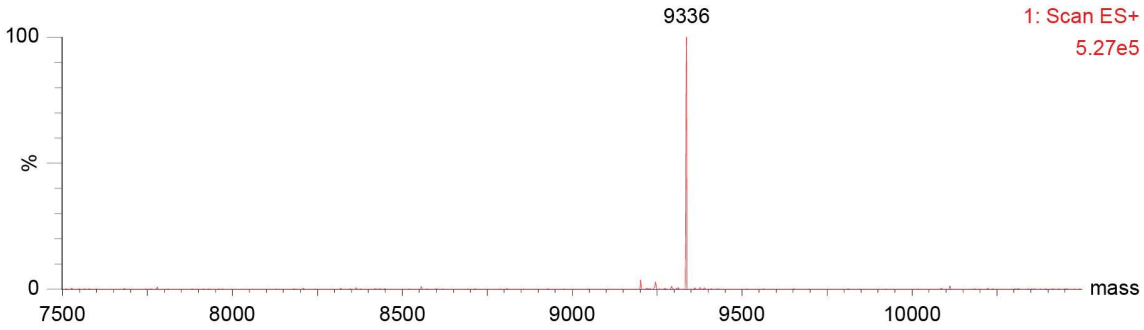


Figure 94. Deconvoluted ESI-MS spectrum obtained following incubation of the UbK63C-maleimide-DBCO protein (25 μM) with compound **4** (25 μM) in ammonium acetate buffer (20 mM, pH 8.0) for 30 min at 25 ºC and 400 rpm. Major product identified: Protein unAcUbK63C** (calculated mass = 9333 Da; observed mass = 9336 Da).

3.5 Generation of unacetylated histones containing the SPAAC by-product at the cysteine

To serve as control proteins in subsequent studies, namely those presented in subsections 3.6 and 3.7, unacetylated H3K4C, H3R52C and H3K4CR52C histones bearing the SPAAC by-product that remains attached at the cysteine were also prepared. Thus, H3K4C, H3R52C and H3K4CR52C proteins were reduced and conjugated to maleimide-DBCO according to the optimized protocols already described. Then, the SPAAC reaction was performed by incubating the modified proteins with compound **2** using the same reaction parameters reported previously for compound **4**. Unsurprisingly, full conversion to the corresponding H3K4C** (**Figure 95**), H3R52C** (**Figure 96**) and H3K4C**R52C** (**Figure 97**) proteins containing the SPAAC by-product was observed.

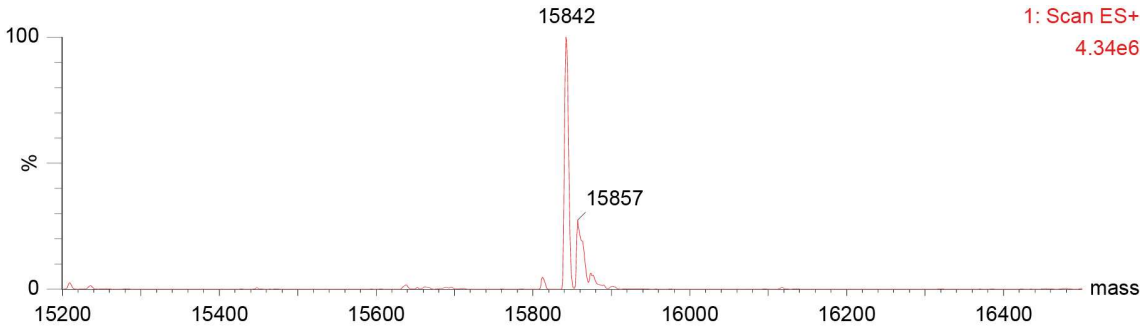


Figure 95. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4C-maleimide-DBCO protein (25 μM) with compound **2** (100 μM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 ºC and 400 rpm. Major product identified: Protein H3K4C** (calculated mass = 15847 Da; observed mass = 15842 Da).

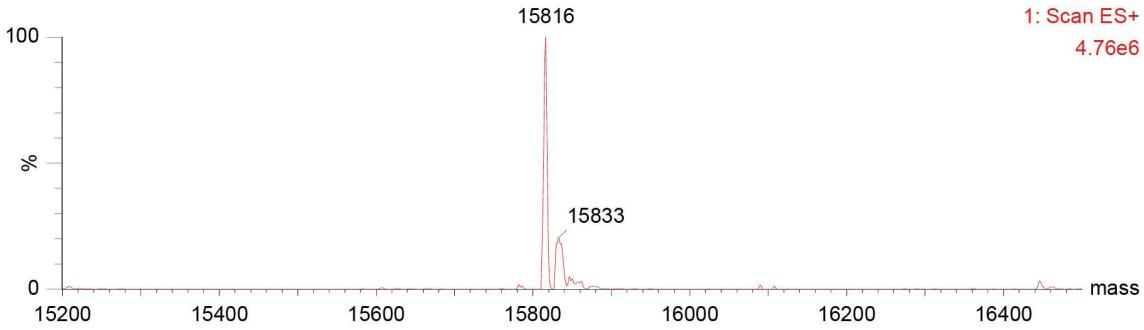


Figure 96. Deconvoluted ESI-MS spectrum obtained following incubation of the H3R52C-maleimide-DBCO protein (25 µM) with compound **2** (50 µM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3R52C** (calculated mass = 15819 Da; observed mass = 15816 Da).

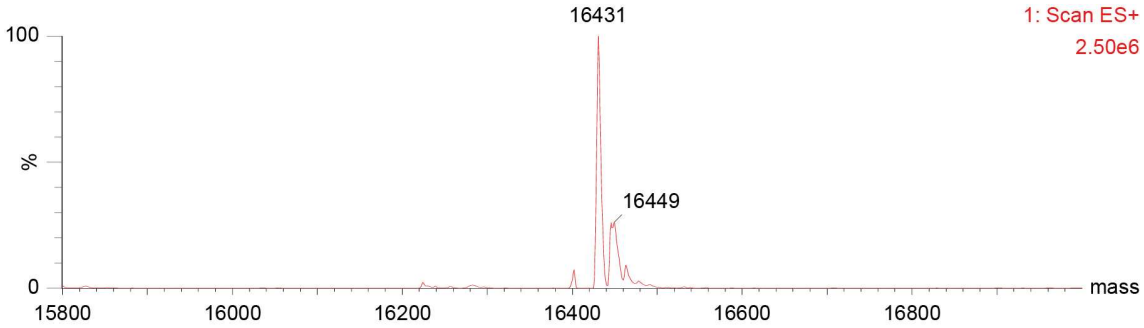


Figure 97. Deconvoluted ESI-MS spectrum obtained following incubation of the H3K4CR52C-maleimide-DBCO protein (25 µM) with compound **2** (100 µM) in ammonium acetate buffer (20 mM, pH 4.9) for 30 min at 25 °C and 400 rpm. Major product identified: Protein H3K4C**R52C** (calculated mass = 16426 Da; observed mass = 16431 Da).

3.6 Evaluation of secondary structure by circular dichroism spectroscopy

CD spectroscopy is an established and worthwhile technique for the study of protein structure. This optical spectroscopic method takes advantage of the differential absorption of left- and right-circularly polarized light in the UV region by chromophores to create representative signatures of distinct structural features in proteins. Relevant chromophores in proteins include the peptide bond (absorption below 240 nm), aromatic amino acid side chains (absorption between 260 and 320 nm) and disulfide bonds (weak broad absorption centered around 260 nm). Particularly, electronic transitions in the far UV region (~ 240 to 170 nm) produce characteristic CD spectra that reflects the secondary structure content of a protein.[286,305,306]

3.6.1 Secondary structure evaluation of the H3K4C protein following cysteine conjugation and SPAAC reactions

CD experiments were performed to establish whether the H3K4C protein maintains its overall secondary structure following cysteine conjugation and subsequent SPAAC reaction. The CD profiles of the H3K4C, H3K4C** and H3K4C**K9Ac proteins were found to be quite similar, indicating that the overall secondary structure of this single-cysteine histone mutant is maintained throughout the conjugation and SPAAC reactions (**Figure 98**).

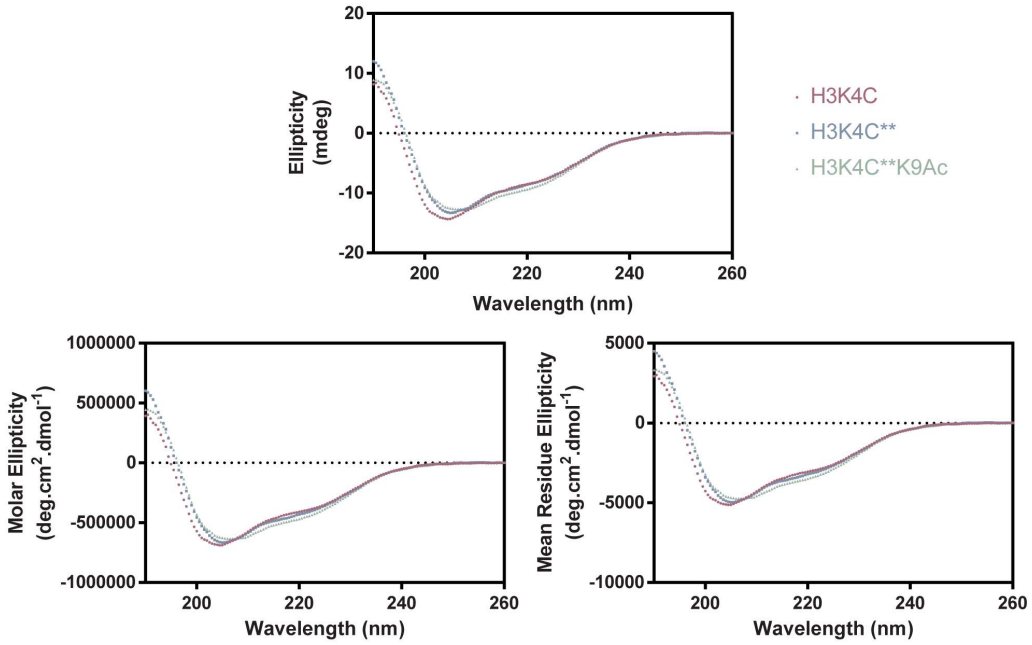


Figure 98. CD profiles of the H3K4C, H3K4C** and H3K4C**K9Ac proteins (20 μM) in sodium phosphate buffer (20 mM, pH 7.4) at 25 °C represented as ellipticity, molar ellipticity and mean residue ellipticity.

3.6.2 Secondary structure evaluation of the H3K4C, H3R52C and H3K4CR52C proteins

CD spectroscopy was also performed to evaluate whether differences exist in the secondary structure of the various H3 mutants generated in this thesis. Appreciable differences in the CD profiles were not observed between the H3K4C, H3R52C and H3K4CR52C proteins, suggesting that all three mutants have similar secondary structures irrespective of the number and location of cysteine residues at positions 4 and/or 52 (**Figure 99**).

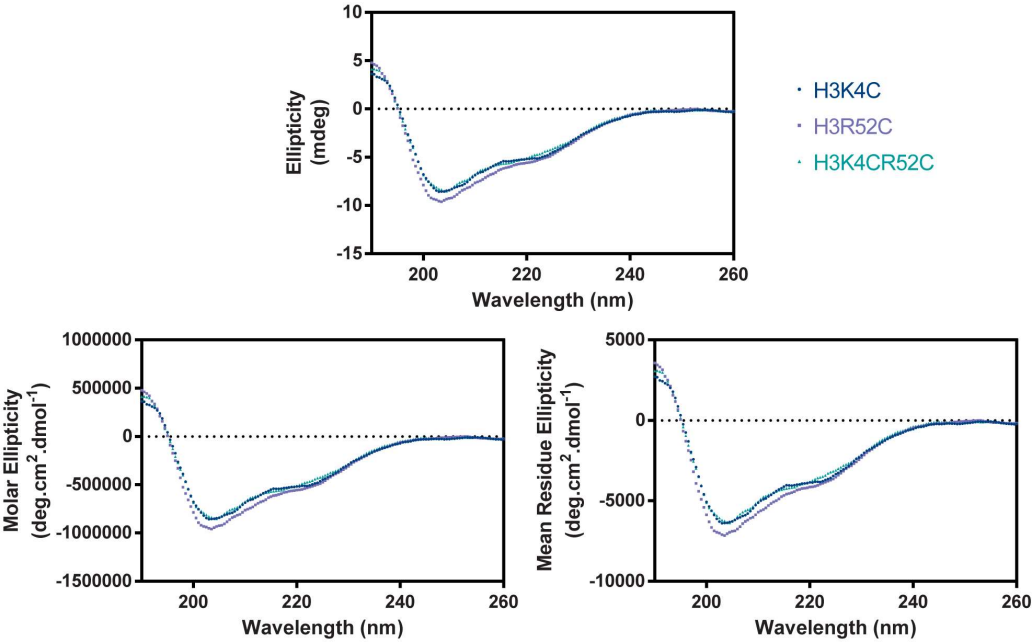


Figure 99. CD profiles of the H3K4C, H3R52C and H3K4CR52C proteins (10 μM) in sodium phosphate buffer (20 mM, pH 7.4) at 20 °C represented as ellipticity, molar ellipticity and mean residue ellipticity.

3.7 Evaluation of acetylated histones as targets of biologically relevant partners

The final step in the project consisted of evaluating whether the chemically acetylated H3K4C**K9Ac, H3R52C**K56Ac and H3K4C**K9AcR52C**K56Ac proteins can be recognized as potential targets of biologically relevant partners, namely antibodies and enzymes. Specifically, Western blot and ELISA experiments were first performed to assess the ability of these chemically acetylated histones to be recognized by antibodies raised against the natural ε-acetyl-lysine at positions 9 and 56 of histone H3. Concomitantly, Coomassie staining experiments were also performed to evaluate the purity of the histone proteins by SDS-PAGE and match the resulting bands to those obtained by Western blot. Subsequently, deacetylation assays with enzymes from the sirtuin family of deacylases were performed and their outcome evaluated by Western blot and ELISA experiments.

3.7.1 Western blot analysis of K9 and K56 acetylation

With respect to Western blot experiments to evaluate whether the H3K4C**K9Ac protein can be recognized by its cognate antibody, a clear band with a molecular weight of approximately 15 kDa corresponding to the acetylated H3 histone at position K9 was observed in the lane loaded with H3K4C**K9Ac. However, no bands were visible in the control lanes containing the H3K4C and H3K4C** proteins (**Figure 100a**). The membrane was then re-probed with the anti-H3 antibody, which served as a control to confirm that the absence of the K9 acetylation signal in the lanes loaded with the H3K4C and H3K4C** proteins was not related to any technical issue (such as variations in protein loading between wells, uneven membrane transfer or antibody probing). Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were observed in the lanes loaded with H3K4C, H3K4C** and H3K4C**K9Ac (**Figure 100b**). Thus, the chemically acetylated H3K4C**K9Ac protein is recognized by antibodies raised against the natural ε-acetyl-lysine at position 9.

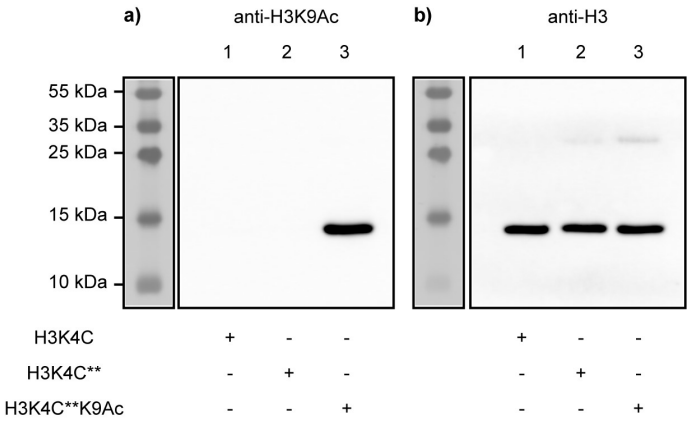


Figure 100. Western blot analysis to confirm K9 acetylation in the H3K4C**K9Ac protein. 300 ng each of H3K4C (lane1), H3K4C** (lane 2) and H3K4C**K9Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K9) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

In addition to showing that the H3K4C, H3K4C** and H3K4C**K9Ac proteins used in the Western blot mentioned above were extremely pure, the corresponding Coomassie staining also revealed a match between the bands resulting from both experiments (**Figure 101**).

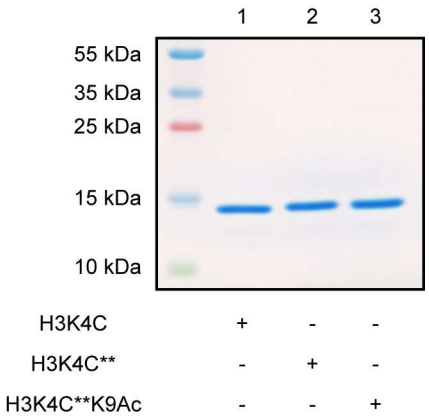


Figure 101. Coomassie staining of the H3K4C, H3K4C** and H3K4C**K9Ac proteins. 2 µg each of H3K4C (lane 1), H3K4C** (lane 2) and H3K4C**K9Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, the gel was stained with Coomassie. The protein ladder and protein bands were visualized by colorimetric detection.

Concerning the Western blot experiments to evaluate whether the H3R52C**K56Ac protein can be recognized by its cognate antibody, a clear band with a molecular weight of approximately 15 kDa corresponding to the acetylated H3 histone at position K56 was observed in the lane loaded with H3R52C**K56Ac. However, no bands were visible in the control lanes containing the H3R52C and H3R52C** proteins (**Figure 102a**). The membrane was then re-probed with the anti-H3 antibody, which served as a control to confirm that the absence of the K56 acetylation signal in the lanes loaded with the H3R52C and H3R52C** proteins was not related to any technical issue (such as variations in protein loading between wells, uneven membrane transfer or antibody probing). Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were observed in the lanes loaded with H3R52C, H3R52C** and H3R52C**K56Ac (**Figure 102b**). Thus, the chemically acetylated H3R52C**K56Ac protein is specifically recognized by antibodies raised against the natural ε-acetyl-lysine at position 56.

The Coomassie staining revealed that the H3R52C, H3R52C** and H3R52C**K56Ac proteins used in the corresponding Western blot were extremely pure. In addition, a match between the bands resulting from both experiments was observed (**Figure 103**).

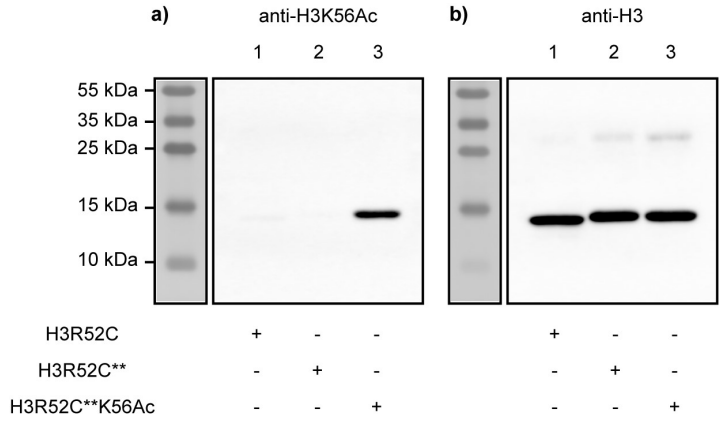


Figure 102. Western blot analysis to confirm K56 acetylation in the H3R52C**K56Ac protein. 2 µg each of H3R52C (lane 1), H3R52C** (lane 2) and H3R52C**K56Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K56) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

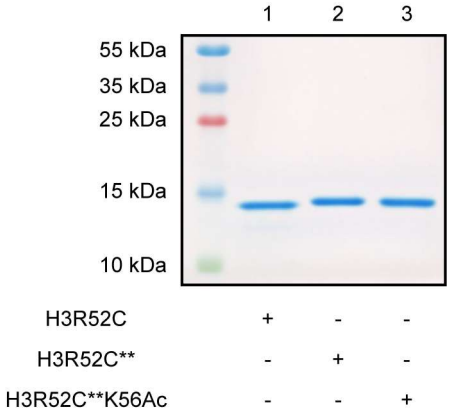


Figure 103. Coomassie staining of the H3R52C, H3R52C** and H3R52C**K56Ac proteins. 2 µg each of H3R52C (lane 1), H3R52C** (lane 2) and H3R52C**K56Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, the gel was stained with Coomassie. The protein ladder and protein bands were visualized by colorimetric detection.

Having established that antibodies raised against the natural ε-acetyl-lysine at positions 9 and 56 of H3 can recognize the H3K4C**K9Ac and H3R52C**K56Ac proteins, respectively, two additional Western blot experiments were devised. In the first, the H3R52C**K56Ac protein was probed with the anti-H3K9Ac antibody to demonstrate that the K9 acetylation signal is only detected for the chemically acetylated H3K4C**K9Ac (**Figure 104a**). The membrane was then re-probed with the anti-H3 antibody, which served as a control to confirm that the absence of K9 acetylation signal in the lane loaded with the H3R52C**K56Ac protein was not related to any technical issue (such as variations in protein loading between wells, uneven membrane transfer or antibody probing). Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were observed in the lanes loaded with H3K4C**K9Ac and H3R52C**K56Ac (**Figure 104b**).

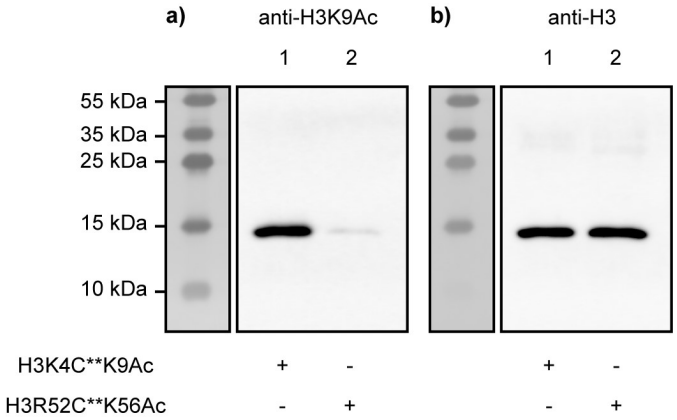


Figure 104. Western blot analysis to confirm the site-specificity of K9 acetylation in the H3K4C**K9Ac protein. 300 ng each of H3K4C**K9Ac (lane 1) and H3R52C**K56Ac was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K9) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

In the second Western blot experiment, the anti-H3K9Ac and anti-H3K56Ac antibodies were used to probe the H3K4C**K9AcR52C**K56Ac protein and confirm that the latter is dually acetylated at lysine residues 9 and 56. When probing separately with the anti-H3K9Ac and anti-H3K56Ac antibodies, two clear bands with a molecular weight of approximately 15 kDa corresponding to the acetylated H3 histone at positions K9 and K56 were observed exclusively in the lanes loaded with H3K4C**K9AcR52C**K56Ac, while no bands were visible in the control lanes containing the H3K4CR52C and H3K4C**R52C** proteins (**Figure 105a** and **Figure 106a**). The membranes were then re-probed with the anti-H3 antibody, which served as a control to confirm that the absence of K9 and K56 acetylation signals in the lanes loaded with the H3K4CR52C and H3K4C**R52C** proteins was not related to any technical issue (such as variations in protein loading between wells, uneven membrane transfer or antibody probing). Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were observed in the lanes loaded with H3K4CR52C, H3K4C**R52C** and H3K4C**K9AcR52C**K56Ac (**Figure 105b** and **Figure 106b**).

The Coomassie staining showed that the H3K4CR52C, H3K4C**R52C** and H3K4C**K9AcR52C**K56Ac proteins used in the corresponding Western blot were extremely pure, while simultaneously revealing a match between the bands resulting from both experiments (**Figure 107**).

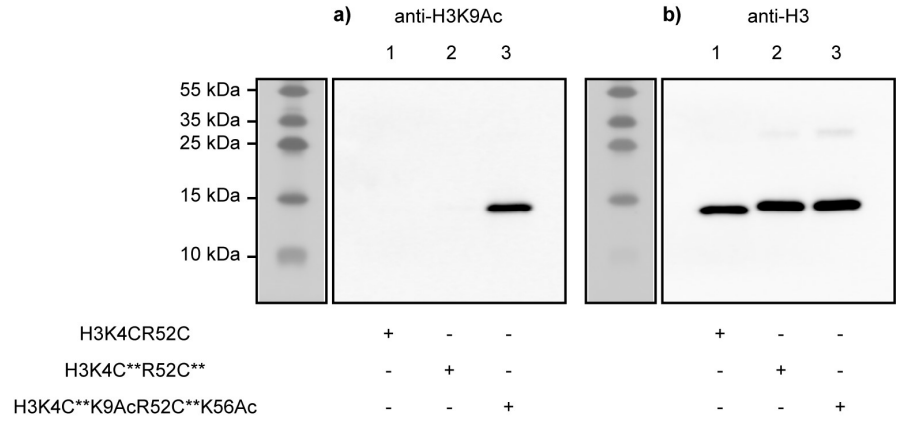


Figure 105. Western blot analysis to confirm K9 acetylation in the H3K4C**K9AcR52C**K56Ac protein. 300 ng each of H3K4CR52C (lane 1), H3K4C**R52C** (lane 2) and H3K4C**K9AcR52C**K56Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K9) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

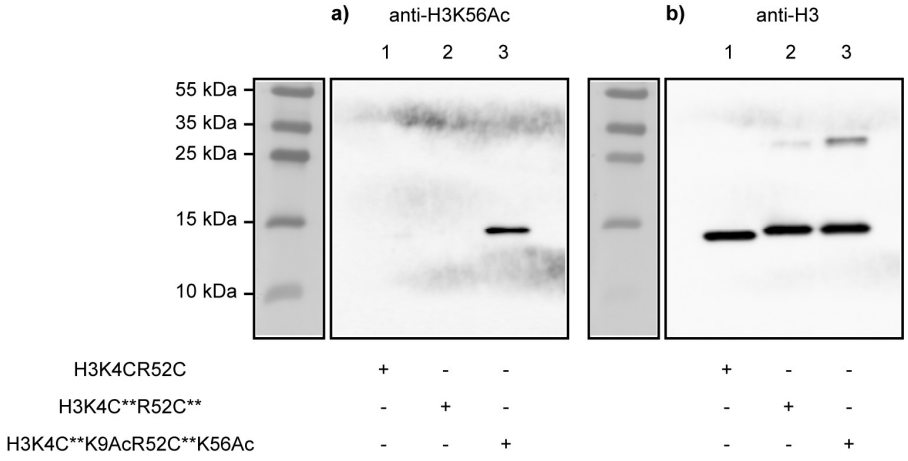


Figure 106. Western blot analysis to confirm K56 acetylation in the H3K4C**K9AcR52C**K56Ac protein. 2 µg each of H3K4CR52C (lane 1), H3K4C**R52C** (lane 2) and H3K4C**K9AcR52C**K56Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K56) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

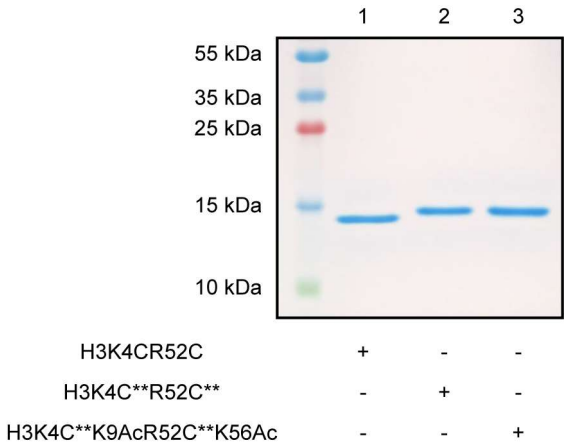


Figure 107. Coomassie staining of the H3K4CR52C, H3K4C**R52C** and H3K4C**K9AcR52C**K56Ac proteins. 2 µg each of H3K4CR52C (lane 1), H3K4C**R52C** (lane 2) and H3K4C**K9AcR52C**K56Ac (lane 3) was loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, the gel was stained with Coomassie. The protein ladder and protein bands were visualized by colorimetric detection.

To summarize, Western blot experiments showed that antibodies raised against the natural ε-acetyl-lysine at positions 9 and 56 of histone H3 recognize and bind to the H3K4C**K9Ac and H3R52C**K56Ac proteins, respectively, and that the presence of the SPAAC by-product at the nearby cysteine is not detrimental to this process. Importantly, Western blot using the anti-H3K9Ac antibody did not show a signal for the H3R52C**K56Ac protein. In the case of the H3K4C**K9AcR52C**K56Ac protein, a blot signal was obtained for both acetylated lysine residues at positions 9 and 56. Together, this data suggests that the cysteine-assisted click-chemistry strategy allows for the site-specific installation of one or two acetyl groups at pre-defined lysine residues of histone H3.

3.7.2 Development of an ELISA protocol for the detection and quantitation of H3K9 acetylation

In addition to Western blot experiments, an ELISA was also performed to detect K9 acetylation in the H3K4C**K9Ac protein. This assay ultimately served to quantitate the amount of K9 acetylation present in the modified protein following incubation with two sirtuin deacetylases.

In agreement with the Western blot results described previously, ELISA experiments showed that antibodies raised against the natural ε-acetyl-lysine at position 9 of histone H3 recognize and bind to the H3K4C**K9Ac protein and that the presence of the SPAAC by-product at the nearby cysteine is not detrimental to this process (**Figure 108**). The developed ELISA assay can specifically detect H3K9 acetylation since the colorimetric signal was only significantly observed for H3K4C**K9Ac and not for the H3K4C or H3K4C** proteins. While a basal signal of K9 acetylation was also observed for high concentrations of H3K4C and H3K4C** proteins, its value is of the same magnitude as the one registered for the lowest amounts of H3K4C**K9Ac. Furthermore, a strong H3 colorimetric signal was observed in all three proteins, indicating that the absence of acetyl-lysine detection at position 9 in the H3K4C and H3K4C** proteins is not related to assay or technical issues (such as variations in protein loading or antibody probing between wells).

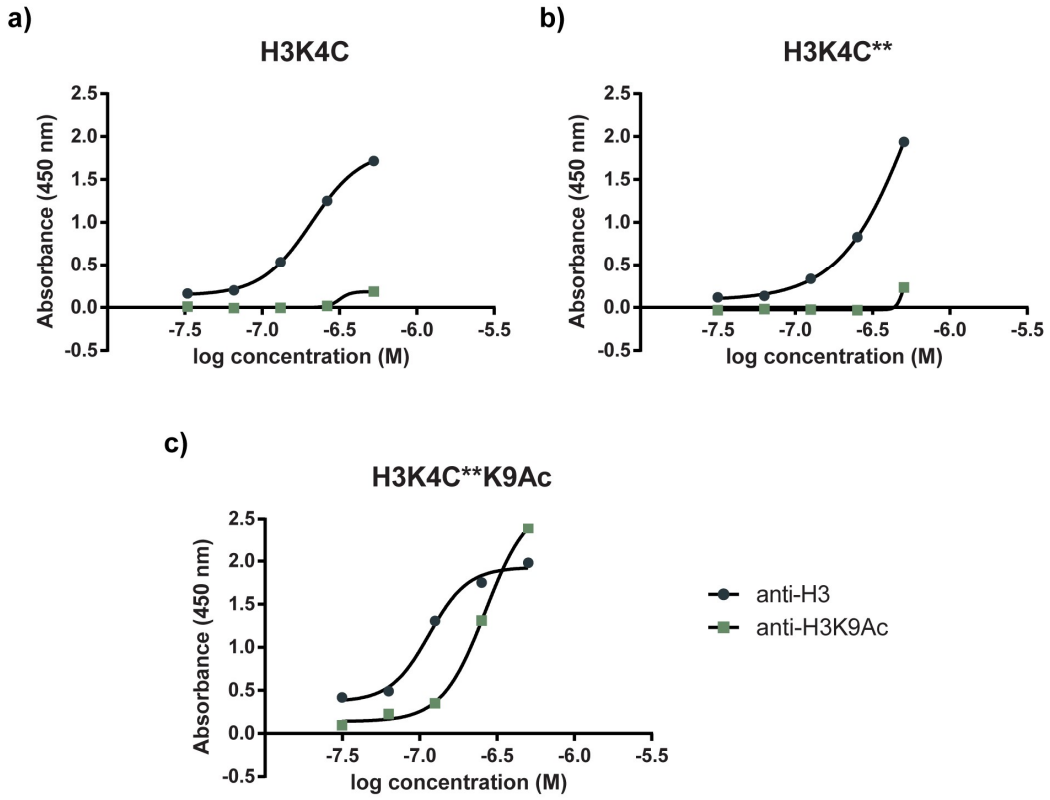


Figure 108. Development of an ELISA protocol for the detection and quantitation of K9 acetylation in the H3K4C**K9Ac protein. Standard curves obtained following probing of the **a)** H3K4C, **b)** H3K4C** and **c)** H3K4C**K9Ac proteins with anti-H3 and anti-H3K9Ac antibodies.

3.7.3 Deacetylation assays using sirtuin enzymes

To establish that H3K4C**K9Ac and H3R52C**K56Ac can be recognized by histone-modifying enzymes, these chemically acetylated proteins were used as substrates in NAD^+-dependent deacetylation assays with Sirt3. Furthermore, the H3K4C**K9Ac protein was also used as a substrate in the NAD^+-dependent deacetylation assay with Sirt6. The outcome of the H3K4C**K9Ac and H3R52C**K56Ac deacetylation assays was evaluated by Western blot experiments. In addition, the result of the H3K4C**K9Ac deacetylation assay was also evaluated by ELISA.

Deacetylation assays with Sirt3

With respect to the deacetylation assay of H3K4C**K9Ac with Sirt3 evaluated by Western blot, a clear band with a molecular weight of approximately 15 kDa corresponding to the acetylated H3 histone at position K9 was observed in the lane loaded with the control sample (**Figure 109a**). However, no band was visible in the lane loaded with the reaction sample, indicating that deacetylation was complete following incubation with Sirt3 for 2 h at 37 ºC. Thus, the chemically acetylated H3K4C**K9Ac protein is recognized by Sirt3 and the presence of the SPAAC by-product at the cysteine does not interfere to a significant extent with this process. The membrane was then re-probed with the anti-H3 antibody, which served as a control to confirm that absence of the K9 acetylation signal in the reaction sample is not related to any technical issue (such as variations in protein loading between wells, uneven membrane transfer or antibody

probing) and due exclusively to the deacetylase activity of Sirt3. Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were observed in the lanes loaded with the control and reaction samples (**Figure 109b**).

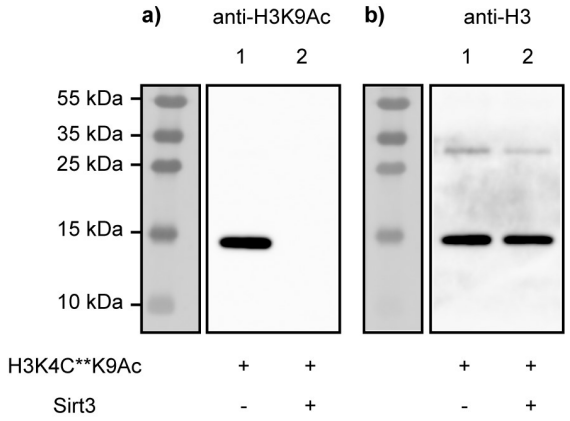


Figure 109. Western blot analysis showing complete K9 deacetylation of the H3K4C**K9Ac protein with Sirt3. Following incubation of H3K4C**K9Ac for 2 h at 37 ºC and 400 rpm in the absence or presence of Sirt3, 300 ng of protein from control (lane 1) and reaction (lane 2) samples, respectively, were loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K9) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

An ELISA was also performed to corroborate the results from the Western blot experiments showing K9 deacetylation of H3K4C**K9Ac following incubation with Sirt3. As observed in **Figure 110b**, incubation with Sirt3 for 2 h at 37 ºC resulted in complete deacetylation of the H3K4C**K9Ac protein. Moreover, the interpolated value of K9 acetylation obtained for the control sample is very similar to the concentration of the H3K4C**K9Ac protein loaded into the ELISA plate wells (**Figure 110c**). To confirm that the absence of K9 acetylation signal in the reaction sample is not related to any technical issue (such as variations in protein loading or antibody probing between wells) and due only to the deacetylase activity of Sirt3, the H3 concentration in control and reaction samples was interpolated from the standard curve of H3K4C**K9Ac probed with the anti-H3 antibody. As observed in **Figure 110d**, while the interpolated H3 concentration values are less than half of those that would be expected, these are quite similar between control and reaction samples. By normalizing the individual values to the average H3 signal of the control, it is observed that both samples generate similar colorimetric signals following H3 probing, which indicates that similar protein quantities were loaded into the wells. Therefore, together with the Western blot experiments, these results show that the remarkable differences in K9 acetylation levels between control and reaction samples following Sirt3 incubation can only be explained by enzymatic deacetylation.

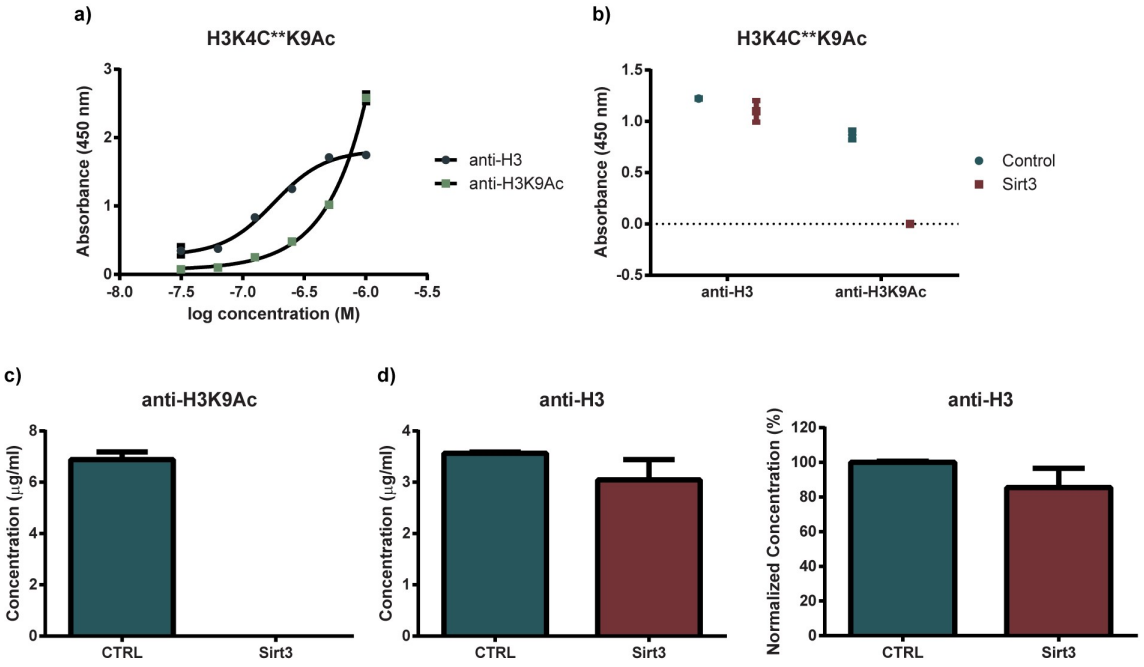


Figure 110. ELISA results showing complete K9 deacetylation of the H3K4C**K9Ac protein with Sirt3. H3K4C**K9Ac was incubated for 2 h at 37 ºC and 400 rpm in the absence or presence of Sirt3; **a)** standard curves obtained for the H3K4C**K9Ac protein following probing with the anti-H3 and anti-H3K9Ac antibodies, **b)** absorbance and **c)** interpolated H3K9Ac concentration values (in µg/mL) in control and Sirt3-treated samples, and **d)** interpolated H3 concentration values (in µg/mL) in control and Sirt3-treated samples and respective normalization (in percentage) to the average H3 signal of the control.

Concerning the deacetylation assay of H3R52C**K56Ac with Sirt3 evaluated by Western blot, a strong and clear band with a molecular weight of approximately 15 kDa corresponding to the acetylated H3 histone at position K56 was observed in the lane loaded with the control sample (**Figure 111a**). While the corresponding band in the lane loaded with the reaction sample was still visible, a very significant reduction in its intensity was seen. Therefore, although deacetylation was not complete following incubation with Sirt3 for 4h at 37ºC, these results indicate that Sirt3 also recognizes the chemically acetylated H3R52C**K56Ac protein, even in the presence of the SPAAC by-product at the cysteine. The membrane was then re-probed with the anti-H3 antibody, which served as a control to confirm that reduction of the K56 acetylation signal in the reaction sample is not related to any technical issue (such as variations in protein loading between wells, uneven membrane transfer or antibody probing) and due only to the deacetylase activity of Sirt3. Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were observed in the lanes loaded with the control and reaction samples (**Figure 111b**).

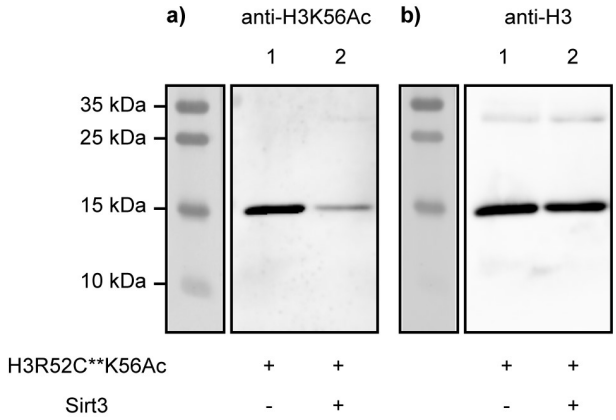


Figure 111. Western blot analysis showing significant K56 deacetylation of the H3R52C**K56Ac protein with Sirt3. Following incubation of H3R52C**K56Ac for 4 h at 37 ºC and 400 rpm in the absence or presence of Sirt3, 2 µg of protein from control (lane 1) and reaction (lane 2) samples, respectively, were loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K56) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

Deacetylation assays with Sirt6

Regarding the deacetylation assay of H3K4C**K9Ac with Sirt6 evaluated by Western blot, clear bands with equivalent densitometry values and a molecular weight of approximately 15 kDa corresponding to the acetylated H3 histone at position K9 were observed in the lanes loaded with control and reaction samples (**Figure 112a**). The membrane was then re-probed with the anti-H3 antibody, which served as a control to confirm that similar protein amounts were loaded in the two wells. Clear bands with equivalent densitometry values and a molecular weight of roughly 15 kDa corresponding to histone H3 were seen in the lanes loaded with the control and reaction samples (**Figure 112b**).

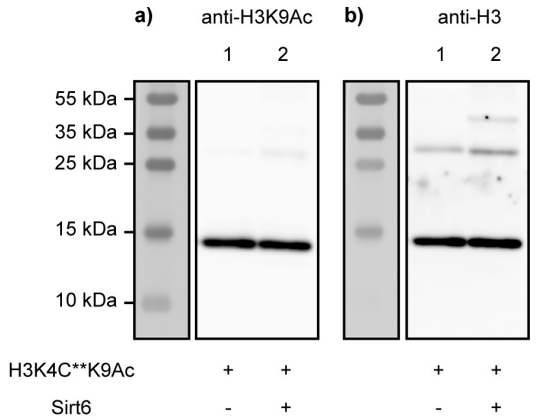


Figure 112. Western blot analysis showing no K9 deacetylation of the H3K4C**K9Ac protein with Sirt6. Following incubation of H3K4C**K9Ac for 4 h at 37 ºC and 400 rpm in the absence or presence of Sirt6, 300 ng of protein from control (lane 1) and reaction (lane 2) samples, respectively, were loaded into the wells of a 15% SDS-PAGE gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane, which was probed with **a)** an anti-acetyl-histone H3 (K9) antibody, and **b)** an anti-histone H3 antibody following acidic stripping. The protein ladder was visualized by colorimetric detection and the protein bands were visualized by chemiluminescence.

An ELISA was also performed to corroborate the results from the Western blot experiments showing no K9 deacetylation of H3K4C**K9Ac following incubation with Sirt6. As observed in

Figure 113b and **Figure 113c**, incubation with Sirt6 for 4 h at 37 ºC did not promote deacetylation of the H3K4C**K9Ac protein. The interpolated values of K9 acetylation obtained for the control and reaction samples are very similar to the concentration of the H3K4C**K9Ac protein loaded into the ELISA plate wells. To confirm that the identical K9 acetylation signal obtained in the control and reaction samples is not related to any technical issue (such as variations in protein loading or antibody probing between wells), the H3 concentration in control and reaction samples was interpolated from the standard curve of H3K4C**K9Ac probed with the anti-H3 antibody. As observed in **Figure 113d**, the interpolated H3 concentration values are less than half of those that would be expected and are approximately 30% lower in the reaction sample with respect to the control, indicating that less Sirt6-treated protein was loaded into the wells. Therefore, together with the Western blot experiments, these results show that the H3K4C**K9Ac protein is not recognized as a substrate by the Sirt6 enzyme.

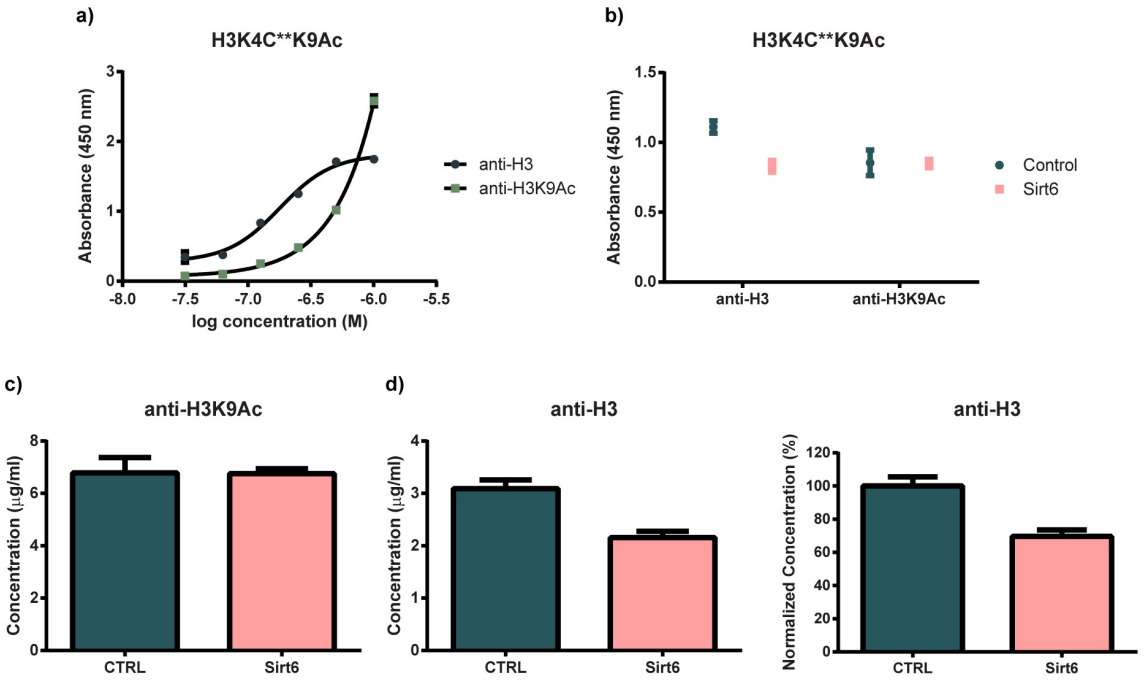


Figure 113. ELISA results showing no K9 deacetylation of the H3K4C**K9Ac protein with Sirt6. H3K4C**K9Ac was incubated for 4 h at 37 ºC and 400 rpm in the absence or presence of Sirt6; **a)** standard curves obtained for the H3K4C**K9Ac protein following probing with the anti-H3 and anti-H3K9Ac antibodies, **b)** absorbance and **c)** interpolated H3K9Ac concentration values (in µg/mL) in control and Sirt6-treated samples, and **d)** interpolated H3 concentration values (in µg/mL) in control and Sirt6-treated samples and respective normalization (in percentage) to the average H3 signal of the control.

4. Discussion

The work presented here was built upon previous research which demonstrated that the sequential combination of cysteine conjugation with click chemistry could be used to acetylate the K9 residue of N-terminal tail peptides of histone H3 with site-specificity. Specifically, by placing a cysteine residue in proximity to the lysine of interest, the peptides could subsequently be modified with the sulfhydryl-reactive maleimide-DBCO. Due to the presence of a SPAAC clickable handle, this heterobifunctional crosslinker served to connect the peptide to a lysine modifying reagent. In the presence of the latter, which contains a click counterpart (azide) for the SPAAC reaction and a stable acetyl donor (*gem*-dithioacetate) with a fixed spacer, it was then demonstrated that mono-acetylation occurred at the nearest lysine at position 9.

The aim of this thesis consisted of extending this concept to the full-length protein, thereby contributing to the development of a novel strategy for the site-specific acetylation of proximal lysine residues in histones. Initially, this strategy was tested to acetylate the recombinant H3K4C at the more external K9 residue due to three main factors, namely the biological significance of this epigenetic mark, the absence of other lysine residues in the primary sequence at equal or similar distance from the cysteine, and the accessibility of K9 conferred by being in the N-terminal tail of the histone. H3K9 acetylation proceeded without considerable technical difficulty or optimization effort, which attests to the value of choosing the K9 residue as a first target of the developed strategy. MS/MS experiments later confirmed the site-specificity of H3K9 acetylation.

Having successfully applied this strategy to generate the corresponding mono-acetylated protein at K9, its broader applicability to acetylate other positions within the H3 sequence was evaluated. To this end, the internal K56 residue of the recombinant H3R52C histone was chosen as the second target of the directed acetylated strategy. Here, site-specific acetylation was achieved only following extensive optimization efforts, primarily due to the excessive use of TCEP and maleimide-DBCO in prior steps. Following the successful generation of the corresponding mono-acetylated protein at K56, the possibility of applying this strategy to acetylate histone H3 simultaneously at K9 and K56 was tested, for which the outcome was favorable as well.

The chemically acetylated proteins were subsequently evaluated as potential targets of biologically relevant partners, namely antibodies and enzymes. In this context, Western blot and ELISA experiments revealed that antibodies raised against the natural ε-acetyl-lysine at positions 9 and 56 of H3 bind to the H3K4C**K9Ac and H3R52C**K56Ac proteins, respectively, as well as to H3K4C**K9AcR52C**K56Ac. Such experiments showed that the presence of the SPAAC by-product at the nearby cysteine residue is not detrimental to this recognition process, and further corroborated the site-specificity of the strategy. The outcome of enzymatic assays showed that the acetyl group at K9 of the chemically modified H3K4C**K9Ac protein is fully removed by Sirt3, as evaluated by Western blot and ELISA experiments. In addition, almost complete deacetylation of the H3R52C**K56Ac protein was observed following incubation with Sirt3, which demonstrates that this sirtuin can recognize and remove the chemically installed acetylation at K56. The presence of the SPAAC by-product also did not appear to interfere with this enzymatic process,

which enables the use of such chemically acetylated proteins in the study of histone-modifying enzymes.

Previous results obtained by MS with the H3K4C**-(1-15)-K9Ac peptide had established that the chemically installed acetylation is susceptible to Sirt6-promoted deacetylation. However, the corresponding experiment of the H3K4C**K9Ac protein with Sirt6 did not yield the expected outcome since deacetylation was not observed as evaluated by Western blot and ELISA. This might be due to several factors at play, namely the poor enzymatic recognition by Sirt6 of the chemically acetylated protein at K9, low specific activity of this sirtuin, or even loss of its biological activity due to prolonged or improper storage and manipulation. Another plausible explanation for the lack of Sirt6-mediated deacetylation in the H3K4C**K9Ac protein pertains to the fact that the *in vitro* deacetylase activity of this enzyme on acetylated H3K9 peptides is intrinsically low and appears to be nucleosome-dependent.[307] Therefore, to investigate the ability of Sirt6 to bind to the chemically acetylated protein, a more suitable experiment would be to incubate this sirtuin with reconstituted nucleosomes containing the H3K4C**K9Ac protein.

While not achieved in this work, removal of the SPAAC by-product is presumably possible using several known procedures, namely desulfurization, palladium(II)-promoted cleavage of the thiosuccinimide linkage and thiol exchange via retro-Michael addition. Desulfurization to generate an alanine residue from the modified cysteine was attempted using the H3K4C**-(1-15)-K9Ac peptide. However, this strategy ultimately proved to be ineffective, resulting in the formation of the expected product with an extra acetylation. This unfavorable outcome can most likely be surpassed by including a purification step between the SPAAC and desulfurization reactions to remove the excess acetyl donor. In this context, dialysis was attempted but purification using this method was not adequate, since most of the acetylated peptide was not recovered during the process. The similar molecular weight values between the acetylated peptide and the dialysis membrane cut-off suggest that the previous result can be explained by the peptide interacting with or crossing the membrane.

Finally, while not explored in the context of this work, the cysteine-assisted click-chemistry strategy described here should be applicable not only to histones, but to other model proteins in which acetylation also plays an important biological role. Additionally, this strategy can also be extended to incorporate acylations other acetylation.

5. Conclusions and Future Perspectives

5.1 Conclusions

Histones represent a family of small and highly conserved proteins found in the nuclei of eukaryotic cells. These basic proteins organize the negatively charged DNA into the repeating cored structures, termed nucleosomes, that form chromatin. In addition to packaging DNA into chromatin, histones are also responsible for regulating its accessibility to replication, transcription and repair machineries. To accomplish this, histones are extensively and dynamically decorated on specific residues with various PTMs, particularly at their flexible N-terminal tails. These chemical modifications can influence chromatin structure and function directly, by influencing the strength of their interactions with DNA, or indirectly by recruiting chromatin-associated proteins.

First discovered more than 50 years ago, acetylation is one of the most common PTMs. In histones, this modification mostly occurs at the side chain amine of lysine residues, where the transfer of an acetyl group neutralizes their positive charge. As a result, acetylation decreases the interaction of histones with the negatively charged DNA, thereby transforming the condensed chromatin into a more relaxed structure that facilitates the accessibility of transcription factors to DNA and is associated with greater levels of gene expression. The acetylation of lysine 9 in histone H3 is a particularly relevant PTM, since it is frequently found at the promoter regions of actively transcribed genes and can recruit transcription factor TFIID. In addition, acetylation of lysine 56 in histone H3 is critical for genomic stability by promoting efficient chromatin assembly following DNA replication and repair.

Although many acetylation marks have been characterized in histones, a biochemical understanding of how these modifications combine to regulate chromatin function is essential. However, access to homogeneous samples of histones containing a single acetyl group in the quantities required for these studies is difficult to achieve from biological sources or enzymatic methods. Thus, the ability to introduce this modification in a site-specific manner is essential to evaluate its functional role. In this context, several strategies have been reported to achieve site-specific acetylation of histones, namely chemical or post-expression mutagenesis, genetic code expansion, protein ligation, and the use of affinity ligands tethered to acetyl transfer catalysts.

This thesis describes the development of a novel strategy that couples cysteine conjugation with click chemistry for the oriented, site-specific acetylation of pre-defined lysine residues in the histone H3 protein. The strategy outlined here comprises two sequential steps: (1) a Michael addition reaction between the unique cysteine of the protein and the maleimide moiety of the heterobifunctional reagent, followed by (2) a SPAAC reaction between the DBCO clickable handle of the latter and an azide bearing an acetyl donor. Here, the SPAAC locks the acetyl donor in proximity to the lysine of interest. Spontaneous acetylation is then triggered by the nucleophilic attack of the side chain amine of the nearest lysine on the carbonyl of the acetyl donor.

By using this strategy, the more external K9 residue and the more internal K56 residue of recombinant H3 proteins were acetylated with site-specificity. Single-cysteine H3K4C and H3R52C mutants were mono-acetylated, whereas the double-cysteine H3K4CR52C mutant was

di-acetylated, without unspecific transfer of the acetyl group to other, more distant nucleophilic lysine residues. Without resorting to complex biochemical techniques, site-specific acetylation was achieved through a combination of distinct conventional and well-established methods, which include recombinant protein expression in *E. coli*, protein purification, cysteine-based chemical modification using a maleimide-based reagent and click chemistry. The resulting acetylated histone proteins are recognized by their cognate antibodies and can interact with relevant biological partners, namely Sirt3, even in the presence of the SPAAC by-product at the nearby cysteine.

In this way, the cysteine-assisted click-chemistry strategy outlined here for the proximity-driven, site-specific acetylation of histones overcomes many of the disadvantages presented by current methods. These include the mutation of the amino acids of interest to generate acetyl-lysine mimics, which may lead to sub-optimal interactions with their natural binding partners in the case of chemical or post-expression mutagenesis, the lower yield of modified histones with genetic code expansion, the more challenging nature of protein ligation, and the dependency on the existence of corresponding interacting ligands that can promote acetylation of proximal lysine residues in the protein sequence, when using affinity ligands tethered to acetyl transfer catalysts.

5.2 Future Perspectives

The work described in this thesis serves as a proof-of-concept to demonstrate that site-specific acetylation of histones can be achieved by sequentially coupling cysteine conjugation with click chemistry. While applied to the acetylation of the K9 and K56 sites of H3, the strategy presented here can in principle be used to introduce the same modification at additional lysine residues, either in the same or other histone proteins. Since this strategy is not histone-specific, it can also be easily adapted to acetylate proteins in which this PTM also plays an important functional role. Furthermore, due to its intrinsic nature involving a S-to-N acyl transfer reaction, this strategy can in theory be extended to incorporate acylations other than acetylation, such as propionylation, butyrylation or crotonylation, whose biological role remains poorly understood. In this context, the *gem*-dithiol reagent used for the SPAAC reaction can be synthesized to include the corresponding propionyl, butyryl or crotonyl groups. Thus, in addition to acetylated histones, the work presented here opens new avenues to prepare site-specifically acylated proteins.

The acetylated histones generated in the context of this dissertation still contained the SPAAC by-product at the engineered cysteine. Based on previously published work, its removal should be attainable, particularly by palladium(II)-promoted cleavage of the thiosuccinimide linkage.[308] In addition, structurally similar reagents to maleimide that enable stable yet reversible thiol-specific conjugation exist, as is the case of 5-methylene pyrrolone (5MP).[309] The latter could be functionalized with an appropriate spacer and DBCO-amine to create a reagent with an identical role to maleimide-DBCO. The use of 5MP-DBCO rather than maleimide-DBCO would thus enable the traceless regeneration of the acetylated histone via a retro-Michael reaction or thiol-exchange. In any case, whether by cleavage of the thiosuccinimide linkage or reversible 5MP conjugation, the nucleophilicity of the resulting free cysteine could be blocked through

alkylation or conversion to a non-nucleophilic alanine residue using known desulfurization procedures.

While the acetylated histones prepared in this work could be recognized by biologically relevant partners, these proteins act as modulators of chromatin structure and function when assembled in nucleosomes. The assembly of the acetylated histones was not explored in the context of this work, but it is expected that this should proceed without difficulty. Nucleosome reconstitution from recombinant histones using different DNA sequences has been described in several publications by various authors. In the specific case presented here, two possibilities are envisioned. In the first, the histone mutant modified with the maleimide-DBCO clickable handle would be assembled with its partner, the H4 protein, to form a heterotetramer that can be used to reconstitute nucleosomes in conjunction with DNA and H2A-H2B dimers. These nucleosomes would subsequently be tested as substrates for the site-specific acetylation of histone H3 following the SPAAC reaction with the acetyl donor. Since the K9 residue is located at the flexible N-terminal tail, it is unlikely that other neighboring nucleophiles could interfere with acetylation at this lysine. However, this could prove more troublesome for the K56 residue, due to its internal location in the H3 sequence and within the nucleosome itself. In this case, the second possibility would consist of preparing the acetylated H3 protein and then assembling it with the remaining histone and DNA partners to form nucleosomes, thereby bypassing the potential occurrence of unspecific cross-reactions with other nucleophiles.

With the assembled nucleosome core particles in hand, it would be possible to perform several biological studies with the aim of investigating the role of acetylation events in histones. For instance, experiments could be designed to investigate the effects of acetylated sites in nucleosome dynamics and whether these could be attributed to DNA unwrapping or repositioning. The possible influence of these acetylated lysine residues on the activity of ATP-dependent chromatin remodeling complexes could also be explored. Another interesting study would consist of performing mononucleosome immunoprecipitation assays in various cell lines, namely those derived from cancerous tissues, to identify possible novel binding partners by proteomics.[310] The above experiments could also be performed by using nucleosomes containing histones bearing more than one acetylated lysine or even multiple acylation types, given that this strategy can be expanded to accommodate acyl groups other than acetyl. Such studies could therefore help elucidate the combinatorial effects of distinct PTMs on nucleosome structure, dynamics and function.

6. Supplementary Material

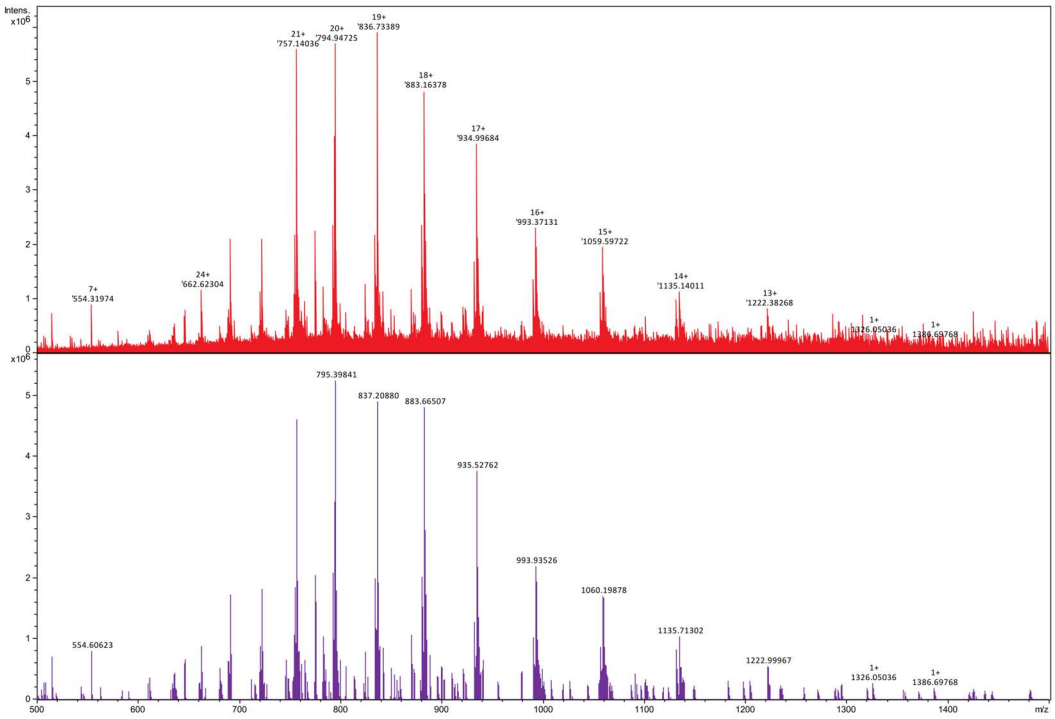


Figure 114. Broadband mass spectra showing charge state distribution of the H3K4C**K9Ac protein. The measured and calculated charge state distributions are shown in red and purple, respectively.

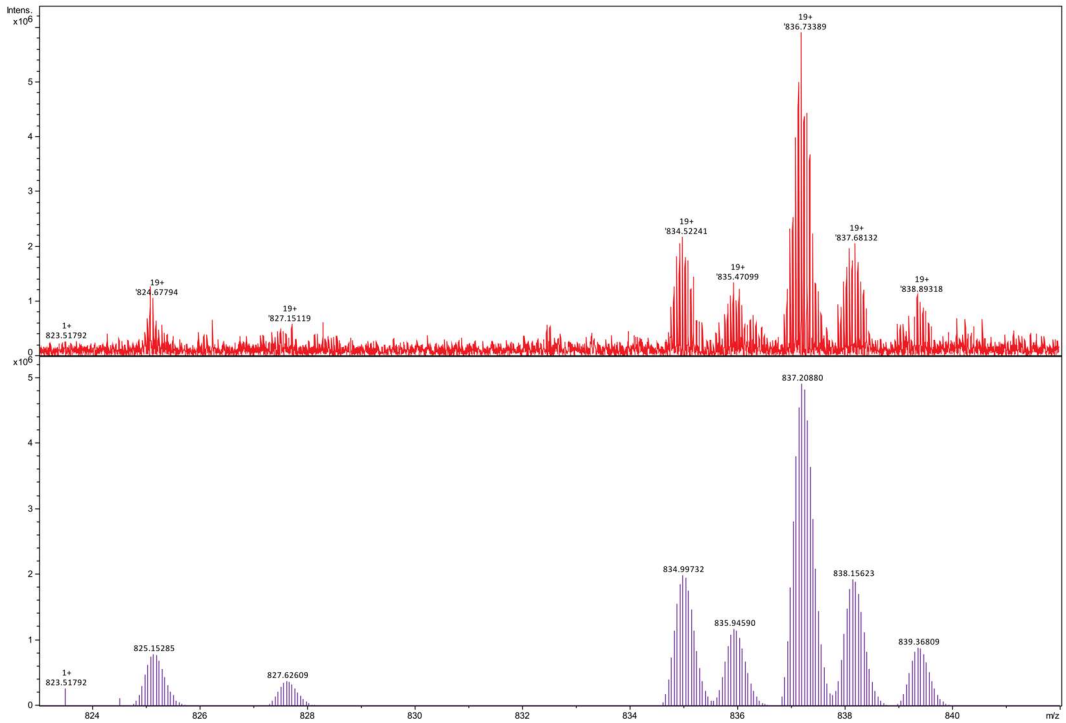


Figure 115. 19^{+} charge state of the H34K4C**K9Ac protein, showing isotopic distribution of carbon. The measured and calculated isotopic distributions of carbon are shown in red and purple, respectively. Seven major forms are present, corresponding to the molecular masses (single charged ions) indicated in **Table 11**. Monoisotopic mass of the neutral modified protein is 15877.665 Da according to the sequence and including the modification at C4 and one acetylation.

Table 11. Molecular masses and corresponding abundances of the major forms present in the injected H3K4C**K9Ac sample.

Mass	Abundance	Relative Abundance	Standard Deviation
15879.81181	24282774	100.00	0.017447
15837.80705	8987362	37.01	0.016178
15896.82382	8051163	33.16	0.023637
15650.73088	5443948	22.42	0.022362
15897.82785	5284286	21.76	0.024171
15917.81366	4674036	19.25	0.016074
15854.80362	2228403	9.18	0.019087

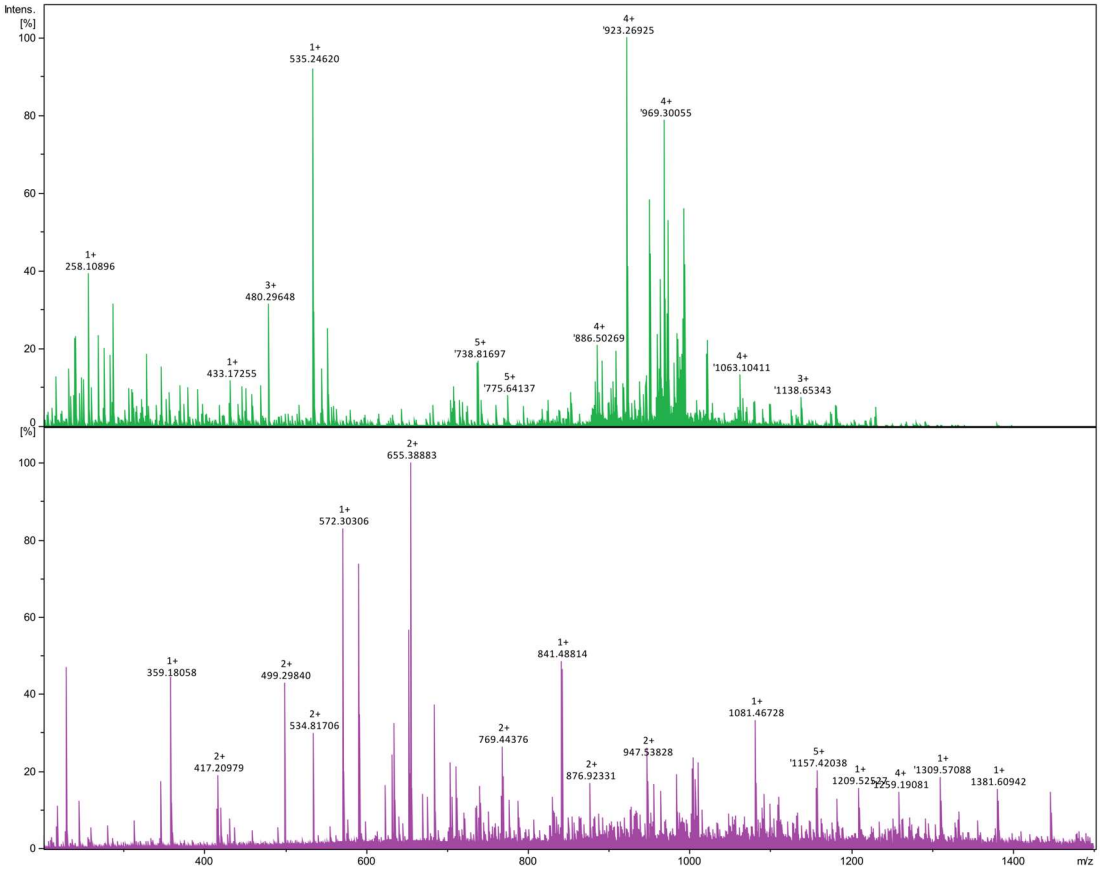


Figure 116. MS/MS spectra of the most abundant 19^+ charged ion corresponding to the H3K4C**K9Ac protein. CID (green) and ECD (purple) were used. Monoisotopic masses (single charged) were determined and combined before matching to the modified sequence.

7. Facsimile of the Published Paper

GDCh Communications Angewandte International Edition Chemie www.angewandte.org

Check for updates

Protein Modifications Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2022**, *61*, e202208543
International Edition: doi.org/10.1002/anie.202208543
German Edition: doi.org/10.1002/ange.202208543

Cysteine-Assisted Click-Chemistry for Proximity-Driven, Site-Specific Acetylation of Histones

*Cláudia F. Afonso, Marta C. Marques, João P. M. António, Carlos Cordeiro, Pedro M. P. Gois, Pedro M. S. D. Cal, and Gonçalo J. L. Bernardes**

Abstract: Post-translational modifications of histones are essential in the regulation of chromatin structure and function. Among these modifications, lysine acetylation is one of the most established. Earlier studies relied on the use of chromatin containing heterogeneous mixtures of histones acetylated at multiple sites. Differentiating the individual contribution of single acetylation events towards chromatin regulation is thus of great relevance. However, it is difficult to access homogeneous samples of histones, with a single acetylation, in sufficient quantities for such studies. By engineering histone H3 with a cysteine in proximity of the lysine of interest, we demonstrate that conjugation with maleimide-DBCO followed by a strain-promoted alkyne-azide cycloaddition reaction results in the acetylation of a single lysine in a controlled, site-specific manner. The chemical precision offered by our click-to-acetylate approach will facilitate access to and the study of acetylated histones.

Histones are proteins that organize the negatively charged DNA into the repeating cored structures, termed nucleosomes, that form chromatin.[1] The nucleosome structure is stabilized by electrostatic interactions, however histone PTMs, particularly at their flexible *N*-terminal tail, constitute an essential epigenetic mechanism that regulates the accessibility of DNA to replication, transcription and repair machineries. Examples of these PTMs include acetylation, methylation and phosphorylation and the reversible nature of these modifications allows modulation of chromatin structure by directly altering histone interactions with DNA or by recruiting chromatin-associated proteins.[2]

Acetylation is a common PTM that occurs at the ε-amine side chain of lysine (Lys or K) residues in histones.[3] The transfer of an acetyl group neutralizes the positive charge of lysine residues, which decreases the interaction of histones with DNA. As a result, the condensed chromatin is transformed into a more relaxed structure that facilitates accessibility of transcription factors to DNA and is associated with greater levels of gene expression. The acetylation of lysine 9 in histone H3 (H3K9Ac) is a particularly relevant PTM, because it is commonly found at the promoter regions of active genes and can induce transcription by promoting transcription factor TFIID binding.[4]

Although many acetylation marks have been characterized in histones, a biochemical understanding of how these modifications combine to regulate chromatin function is essential. However, access to homogeneous samples of histones containing a single acetyl group in the quantities required for these studies is difficult to achieve from biological sources or enzymatic methods.[5] Several approaches have been developed to achieve site-specific acetylation of histones, namely protein ligation, genetic code expansion, cysteine-selective modification and the use of affinity ligands tethered to acyl transfer catalysts.[6] Our technology seeks to overcome many of the disadvantages presented by these methods, namely the technically challenging nature of protein ligation methods;[7] the lower yield of modified histones with genetic code expansion;[8] mutagenesis of the amino acids of interest to generate acetyllysine mimics, which may lead to sub-optimal interactions with their natural binding partners in the case of cysteine conjugation methods;[9] and the dependency on the existence of corresponding interacting ligands that can promote acetylation of particular lysines in the protein sequence.[10]

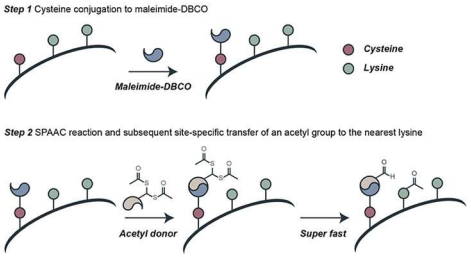
Herein, we present a novel strategy that couples cysteine conjugation with click-chemistry for site-specific acetylation of a single proximate lysine residue on histones. We introduce an artificial "acetyltransferase-like" histone that is chemically and spatially controlled by the addition of a click handle counterpart (Scheme 1).

[*] C. F. Afonso, Dr. M. C. Marques, Dr. P. M. S. D. Cal, Prof. G. J. L. Bernardes
Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa
Avenida Professor Egas Moniz, 1649-028 Lisboa (Portugal)
E-mail: gb453@cam.ac.uk

Dr. J. P. M. António, Prof. P. M. P. Gois
Research Institute for Medicines (iMed.ULisboa), Faculdade de Farmácia, Universidade de Lisboa
Av. Prof. Gama Pinto, 1649-003 Lisboa (Portugal)

Dr. C. Cordeiro
Laboratório de FT-ICR e Espectrometria de Massa Estrutural, Faculdade de Ciências Universidade de Lisboa
Campo Grande, 1749-016 Lisboa (Portugal)

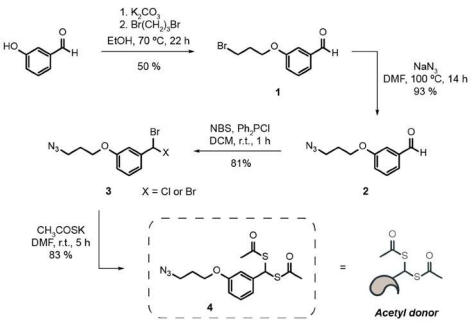
Prof. G. J. L. Bernardes
Yusuf Hamied Department of Chemistry, University of Cambridge
Lensfield Road, CB2 1EW, Cambridge (UK)



Scheme 1. Overview of the site-specific acetylation strategy presented in this work.

Considering the biological significance that acetylation of lysine 9 in histone H3 has, we decided to use it as a target to test our proximity-driven acetylation strategy. For this purpose, the initial sequence of 15 amino acids of histone H3 was synthesized, in which single cysteine residues replaced both surrounding lysines of K9 to generate two different peptides—peptides K4C and K14C (Figures S1, S2)—which enables the installation of the clickable handles. Two chemical reagents were sought at this stage, the first a hetero-crosslinker with a cysteine-modifying handle (maleimide) and a strain-promoted alkyne-azide cycloaddition (SPAAC) clickable handle (cyclooctyne) that can connect the peptide to the second, a lysine modifying reagent. The lysine modifying reagent has a click counterpart for the SPAAC reaction (azide) and a stable acetyl donor (*gem*-dithioacetate) with a fixed spacer. The first reagent is commercially available (maleimide-DBCO) and the second was synthesized (Scheme 2).

We started with an alkylation reaction between 3-hydroxybenzaldehyde and a dibromoalkyl reagent, followed by a substitution of the remaining bromine by sodium azide. From the azide-aldehyde bearing compound, we performed a bromination reaction of the aldehyde function, which was



Scheme 2. Synthesis of the acetyl donor that has an azide SPAAC counterpart and a *gem*-dithioacetate. DMF = dimethyl formamide. NBS = *N*-bromosuccinimide.

subsequently reacted with potassium thioacetate to afford **4** with an overall yield of 31 %. Smaller-sized dibromoalkyl reagents were also trialled but did not afford the intended product. To assess if the synthesized chemical handle was stable for the intended purpose, that is a SPAAC with a cysteine derivative bearing a click handle, a doubly protected S-propargyl-cysteine derivative was used as a model. This reaction confirmed that the *gem*-dithioacetate is stable under the conjugation reaction conditions, which demonstrates the viability of such a molecule. As reported,[11] *gem*-dithioacetate **4** by itself acetylates free cysteines in peptides under the same conditions (Figures S3, S4).

Next, we moved on to test our proposed strategy in H3 tail peptides. Maleimide-DBCO was used to perform the initial cysteine bioconjugation method, which installed a clickable handle within five residues of the target K9 for acetylation. After optimization, peptides K4C and K14C reacted under mild conditions with a slight excess of maleimide-DBCO (2 equiv, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 2 h at 25 °C, Figures S5, S6). With these modified peptides, the directed acetylation of K9 was attempted. Both peptides reacted with **4** to afford the click product from the SPAAC and acetylation of the nearest nucleophilic residue, in this case K9. The reaction conditions were optimized to give the expected product with complete conversion (8 equiv of **4**, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 1 h at 25 °C; Figures 1 and S7, S8). To ensure their stability, the acetylated peptides were evaluated after 1 month and no discernible degradation was seen (Figures S9, S10). To confirm that acetylation occurs in the nearest nucleophilic residue K9, we used high-resolution FT-ICR mass spectrometry and MS/MS fragmentation experiments to map the resulting peptides. For both peptides acetylation always occurred in the lysine closest to the modified cysteine (Figure 1 and Tables S1, S2). As a control, mono-acetylated peptides were reacted with iodoacetamide to confirm that reaction with the maleimide had taken place exclusively at the cysteine. As expected, no changes were noted (Figures S11–S14). All the above-mentioned experiments were performed with peptides K4C and K14C. Furthermore, by using a model peptide containing several reactive amino acids together with a cysteine residue, but without lysines, we demonstrated that under the same reaction conditions acetylation did not occur and that the resulting SPAAC product was stable to hydrolysis (Figures S15–S17).

To establish that the SPAAC reaction is essential for orienting the acetylation reaction towards the nearby K9 residue, we conjugated both peptides with a maleimide-dummy under the same reaction conditions as those described above and added either **4** by itself or the product resulting from the SPAAC between this compound and maleimide-DBCO. As expected, acetylation was not detected in either of the K4C and K14C peptides (Figures 1 and S18–S21).

Given that the conjugation reactions and their respective controls in peptides supported our claims, we performed two additional assays to confer a greater biological significance to our strategy. First, we evaluated the possibility of removing the side product of the SPAAC reaction still

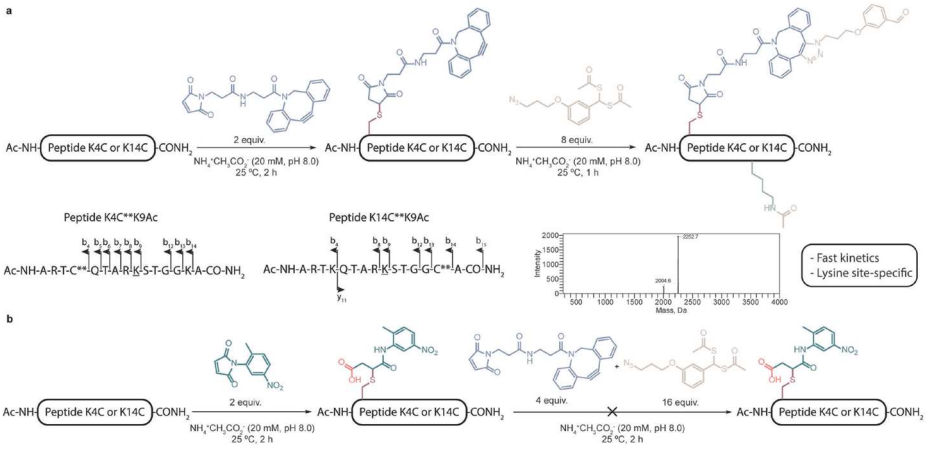


Figure 1. Strategy for acetylating histone H3 peptides. a) Sequential reaction of histone H3K4C and K14C peptides with maleimide-DBCO and **4** results in site-specific acetylation of the nearest lysine residue as confirmed by MS/MS analysis. b) Sequential reaction of histone H3K4C and K14C peptides with a maleimide reagent without an alkyne (maleimide-dummy) and the click product formed between maleimide-DBCO and **4** did not promote acetylation.

attached to the cysteine residue by using thiol-exchange via a retro-Michael addition. β-mercaptoethanol was used as the thiol compound and although somewhat successful, irreversible hydrolysis of the maleimide occurred during the process, which resulted in a significant amount (20–30 %) of the side product unaltered. In addition, hydrolysis of the acetylated K9 following the retro-Michael addition was also observed (20 %), which affords the unmodified K4C and K14C peptides (Figures S22–S25). Together, these experiments demonstrate the feasibility of the cysteine-assisted click-chemistry for lysine acetylation as well as proof-of-concept for its potential for proximity-driven residue-specific acetylation on peptides.

Next, we assessed the ability of the acetylated peptides to be recognized as substrates by one of the NAD^+-dependent histone deacetylase enzymes. Under conditions suitable for deacetylation to occur, modified K4C and K14C peptides were incubated with Sirt6, which can remove the acetyl group of specific lysine residues of histone H3, namely K9 and K56.[12] Although deacetylation was incomplete, a significant difference between the results of both peptides was observed for the first time. For the modified K4C peptide, conversion towards the deacetylated product was ≈40 % (Figure S26), whereas for the modified K14C peptide it was <20 % (Figure S27). These results serve as preliminary data to demonstrate the ability of the modified peptides to be recognized by histone deacetylases, even in the presence of the SPAAC side product at the nearby cysteine.

The same set of bioconjugation reactions as those previously described for the K4C and K14C peptides were performed for the full-length protein, which was recombinantly expressed with a single cysteine mutant H3K4C (C110A) (Figure S28).[9a,b] Treatment of H3K4C with an excess of maleimide-DBCO resulted in complete conversion of the protein (2.5 equiv, $NH_4^+CH_3CO_2^-$ 20 mM, pH 7.0, 1 h at 25 °C; Figures 2 and S34). No reaction with Ellman's reagent confirmed that conjugation occurred at the free cysteine of H3K4C (10 equiv, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C; Figure S62). After purification, site-specific directed acetylation was attempted by treating the modified protein with **4**. Incubating the H3K4C-maleimide-DBCO protein with a 4-equiv excess of the reagent yielded the respective mono-acetylated product, H3K4C**K9Ac ($NH_4^+CH_3CO_2^-$ 20 mM, 30 min at 25 °C; Figures 2, S37 and S43,S44). Analysis of the circular dichroism (CD) profiles of the three proteins (H3K4C, H3K4C** and H3K4C**K9Ac) showed that the overall secondary structure of H3K4C did not change following cysteine conjugation and subsequent SPAAC reaction (Figure S71). The site-specificity of our approach was confirmed by high-resolution FT-ICR mass spectrometry and MS/MS fragmentation experiments which showed acetylation only at residue K9 (Figure S75).

We then questioned if this strategy could be generalized to other lysine residues for which the biological role of acetylation is also important or unknown. To test this possibility, we chose the more internal residue K56 of histone H3 as the second target of our directed acetylation strategy. H3K56 acetylation (H3K56Ac) is a conserved epigenetic mark found in yeast and higher eukaryotes that plays a critical role in genomic stability by promoting efficient nucleosome assembly following DNA replication.[13] Consequently, the single cysteine mutant H3R52C (C110A) was generated, recombinantly expressed and purified (Figure S29). The first conjugation reaction of H3R52C with

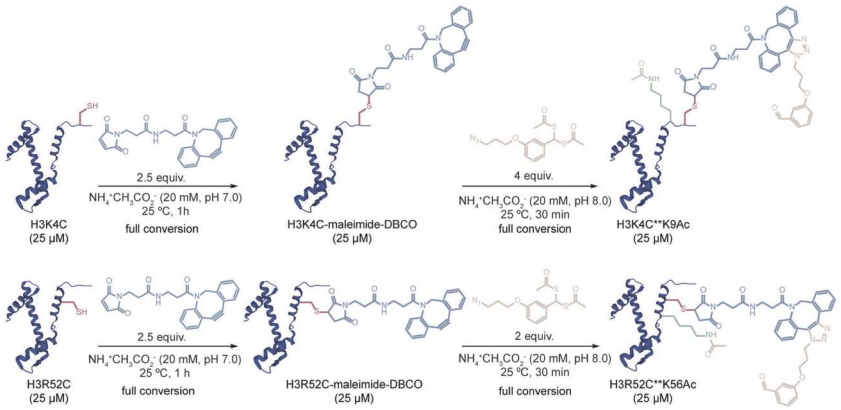


Figure 2. Strategy for acetylating histone H3 proteins. The sequential reaction of histones H3K4C and H3R52C with maleimide-DBCO and **4** results in the mono-acetylation of these proteins.

maleimide-DBCO was attempted using the same conditions as those described previously for the H3K4C protein, with complete conversion observed with 2.5 equiv of the reagent (Figures 2 and S35). The modified protein did not react with Ellman's reagent, which confirmed that conjugation of maleimide-DBCO occurred at the free cysteine of H3R52C (10 equiv, $NH_4^+CH_3CO_2^-$ 20 mM, pH 8.0, 30 min at 25 °C, Figure S64). Following size-exclusion purification of the modified protein, treatment with 2 equiv of **4** yielded the corresponding mono-acetylated product, H3R52C**K56Ac ($NH_4^+CH_3CO_2^-$ 20 mM, 30 min at 25 °C; Figures 2, S38 and S45). Thus, attesting greater value to the strategy, not only easily accessible residues could be modified but also more hindered and constrained residues within the protein sequence. Furthermore, to establish that the SPAAC reaction is essential to direct the acetyl group towards the nearest lysine and thus for acetylation to occur in proteins, H3K4C and H3R52C were conjugated to the same maleimide-dummy as the one described previously for the corresponding peptide experiments (Figures S50, S51). Following purification of the modified proteins, these were treated with **4** under the same conditions as those used for the SPAAC reactions. Similarly to K4C and K14C peptides, acetylation was not detected in either of the two proteins (Figures S53, S54, S56, S57).

We then checked whether the chemically acetylated H3K4C**K9Ac and H3R52C**K56Ac proteins are recognized as antigens by their cognate antibodies. Western blot and ELISA experiments showed that antibodies raised against the natural ε-acetyl-lysine at positions 9 and 56 of H3 bind to H3K4C**K9Ac and H3R52C**K56Ac, respectively, and that the presence of the SPAAC product at the nearby cysteine residue is not detrimental to this process (Figures 3, S76–S79 and S84). Importantly, western blot using the anti-H3K9Ac antibody did not show a signal for the H3R52C**K56Ac protein (Figure S80). When a dual site-specific H3K4C**K9AcR52C**K56Ac was produced (Figures S39 and S46), a blot signal was obtained for both acetylated lysines at positions 9 and 56 (Figures S81–S83). Together this data suggests that our cysteine-assisted click-chemistry approach allows for site-specific installation of one or two acetyl groups at pre-determined lysines in H3.

Finally, we wondered whether our acetylated H3K4C**K9Ac and H3R52C**K56Ac proteins can be used in the study of histone-modifying enzymes. Both proteins were incubated with Sirt3, one of the NAD^+-dependent histone deacetylase enzymes that exhibits a high preference for removing acetyl groups of many lysine residues of histone H3.[14] After incubating H3K4C**K9Ac with Sirt3 for 2 h at 37 °C, full removal of the acetyl group was observed as indicated by western blot and ELISA experiments (Figures 3 and S85, S86). Significant deacetylation of H3R52C**K56Ac was also seen following incubation with Sirt3 for 4 h at 37 °C (Figures 3 and S87). Collectively, this data shows that our method is useful to create functional H3 proteins with acetyl group at precise sites.

In summary, we describe an approach which combines cysteine conjugation with click chemistry for the oriented site-specific acetylation of pre-defined lysine residues. Our strategy comprises two sequential steps: a Michael addition reaction between the unique cysteine of the protein and the maleimide moiety of the heterocrosslinker reagent, followed by a SPAAC reaction between the DBCO clickable handle of the latter and an azide bearing an acetyl donor. Here, the SPAAC locks the acetyl donor in proximity to the lysine of interest. Spontaneous acetylation is then triggered by nucleophilic attack of the side-chain amine of the nearest lysine on the carbonyl of the acetyl donor. By using this approach, we modified recombinant histones at the more external K9 residue or at the more internal K56 residue of histone H3. H3K4C and H3R52C mutants were mono-acetylated (and H3K4CR52C di-acetylated), without unspe-

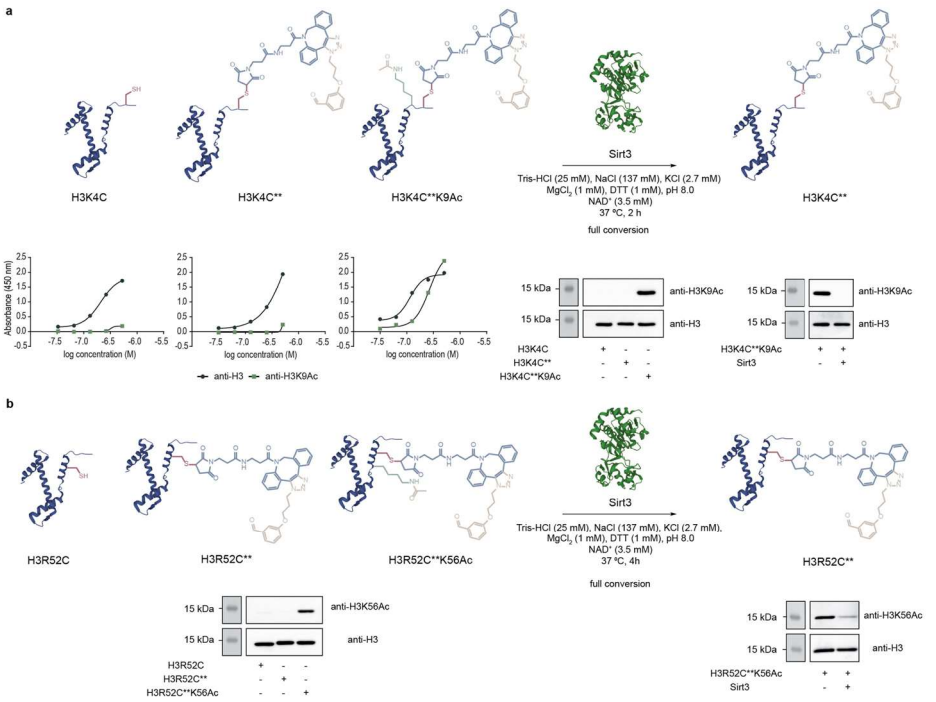


Figure 3. Western Blot and ELISA experiments confirmed the site-specificity of our approach for acetylating a) position K9 and b) position K56 of histone H3. Deacetylation experiments with Sirt3 demonstrated that the chemically installed acetyl group in these proteins can be removed even in the presence of the SPAAC product at the nearby cysteine. The Western Blot images displayed here are duplicates of the full-size images reported in the Supporting Information.

cific transfer of the acetyl group to other, more distant nucleophilic lysines nor by resorting to any complex biological or purification techniques. These acetylated histones interact with their relevant biological partners, even in the presence of the SPAAC product at the nearby cysteine. Removal of this by-product may be achieved using known procedures such as Pd^{II}-promoted cleavage of the thiosuccinimide linkage[15] or thiol exchange via retro-Michael addition.[16] The nucleophilicity of the resulting free cysteine could be blocked through alkylation or conversion to a non-nucleophilic alanine using known desulfurization procedures.[17] The strategy reported here can be used to probe the role of acetylation events at precise sites, which may help decipher their contribution towards a given biological function.

Acknowledgements

We thank Prof. Ernest Laue for providing the histone H3 plasmid. We are grateful for the support from the Portuguese Mass Spectrometry Network (LISBOA-01-0145-FEDER-022125) and the Project EU FT-ICR MS, funded by the Europe Union's Horizon 2020 research and innovation program under grant agreement nr.731077. Funding from FCT Portugal (PhD scholarship PD/BD/135512/2018 to C.F.A, Postdoctoral Fellowship SFRH/BPD/118731/2016 to M.C.M and PTDC/QUI-OUT/3989/2021 to J.P.M.A, and FCT Stimulus to P.M.S.D.C. CEECIND/04518/2017 and G.J.L.B. CEECIND/00453/2018) is also acknowledged. The authors thank Dr Vikki Cantrill for her help with the editing of this manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the Supporting Information of this article.

Keywords: Click Chemistry · Cysteine Bioconjugation · Histones · Lysine Acetylation · Maleimide

[1] Y. L. Deribe, T. Pawson, I. Dikic, *Nat. Struct. Mol. Biol.* **2010**, *17*, 666–672.
[2] S. Venkatesh, J. L. Workman, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 178–189.
[3] A. Drazic, L. M. Myklebust, R. Ree, T. Arnesen, *Biochim. Biophys. Acta Proteins Proteomics* **2016**, *1864*, 1372–1401.
[4] a) T. Agalioti, G. Chen, D. Thanos, *Cell* **2002**, *111*, 381–392; b) M. Vermeulen, K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann, H. T. Timmers, *Cell* **2007**, *131*, 58–69; c) Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, K. Zhao, *Nat. Genet.* **2008**, *40*, 897–903.
[5] A. Dhall, C. Chatterjee, *ACS Chem. Biol.* **2011**, *6*, 987–999.
[6] C. A. Musselman, T. G. Kutateladze, *ACS Chem. Biol.* **2019**, *14*, 579–586.
[7] a) J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier, J. J. Ottesen, *J. Mol. Biol.* **2011**, *408*, 187–204; b) Y. K. Qi, H. S. Ai, Y. M. Li, B. Yan, *Front. Chem.* **2018**, *6*, 19.
[8] H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, J. W. Chin, *Mol. Cell* **2009**, *36*, 153–163.
[9] a) R. Huang, M. A. Holbert, M. K. Tarrant, S. Curtet, D. R. Colquhoun, B. M. Dancy, B. C. Dancy, Y. Hwang, Y. Tang, K. Meeth, R. Marmorstein, R. N. Cole, S. Khochbin, P. A. Cole, *J. Am. Chem. Soc.* **2010**, *132*, 9986–9987; b) F. Li, A. Allahverdi, R. Yang, G. B. Lua, X. Zhang, Y. Cao, N. Korolev, L. Nordenskiold, C. F. Liu, *Angew. Chem. Int. Ed.* **2011**, *50*, 9611–9614; *Angew. Chem.* **2011**, *123*, 9785–9788; c) J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed.* **2012**, *51*, 1835–1839; *Angew. Chem.* **2012**, *124*, 1871–1875; d) Y. Jing, Z. Liu, G. Tian, X. Bao, T. Ishibashi, X. D. Li, *Cell Chem. Biol.* **2018**, *25*, 166–174.e7; e) S. Bhat, Y. Hwang, M. D. Gibson, M. T. Morgan, S. D. Taverna, Y. Zhao, C. Wolberger, M. G. Poirier, P. A. Cole, *J. Am. Chem. Soc.* **2018**, *140*, 9478–9485.
[10] a) Y. Koshi, E. Nakata, M. Miyagawa, S. Tsukiji, T. Ogawa, I. Hamachi, *J. Am. Chem. Soc.* **2008**, *130*, 245–251; b) Y. Amamoto, Y. Aoi, N. Nagashima, H. Suto, D. Yoshidome, Y. Arimura, A. Osakabe, D. Kato, H. Kurumizaka, S. A. Kawashima, K. Yamatsugu, M. Kanai, *J. Am. Chem. Soc.* **2017**, *139*, 7568–7576.
[11] Y. Zhao, J. Kang, C. M. Park, P. E. Bagdon, B. Peng, M. Xian, *Org. Lett.* **2014**, *16*, 4536–4539.
[12] a) E. Michishita, R. A. McCord, E. Berber, M. Kioi, H. Padilla-Nash, M. Damian, P. Cheung, R. Kusumoto, T. L. Kawahara, J. C. Barrett, H. Y. Chang, V. A. Bohr, T. Ried, O. Gozani, K. F. Chua, *Nature* **2008**, *452*, 492–496; b) E. Michishita, R. A. McCord, L. D. Boxer, M. F. Barber, T. Hong, O. Gozani, K. F. Chua, *Cell Cycle* **2009**, *8*, 2664–2666.
[13] a) C. Das, M. S. Lucia, K. C. Hansen, J. K. Tyler, *Nature* **2009**, *459*, 113–117; b) S. Topal, P. Vasseur, M. Radman-Livaja, C. L. Peterson, *Nat. Commun.* **2019**, *10*, 4372.
[14] K. Tanabe, J. Liu, D. Kato, H. Kurumizaka, K. Yamatsugu, M. Kanai, S. A. Kawashima, *Sci. Rep.* **2018**, *8*, 2656.
[15] G. B. Vamisetti, G. Satish, P. Sulkshane, G. Mann, M. H. Glickman, A. Brik, *J. Am. Chem. Soc.* **2020**, *142*, 19558–19569.
[16] Y. Zhang, X. Zhou, Y. Xie, M. M. Greenberg, Z. Xi, C. Zhou, *J. Am. Chem. Soc.* **2017**, *139*, 6146–6151.
[17] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

Manuscript received: June 10, 2022
Accepted manuscript online: September 20, 2022
Version of record online: October 18, 2022

8. Bibliography

1. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; W.H. Freeman & Company, 2008.
2. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; et al. How many human proteoforms are there? *Nat. Chem. Biol.* **2018**, *14* (3), 206-214. DOI: 10.1038/nchembio.2576
3. Bradshaw, R. A.; Medzihradszky, K. F.; Chalkley, R. J. Protein PTMs: post-translational modifications or pesky trouble makers? *J. Mass Spectrom.* **2010**, *45* (10), 1095-1097. DOI: 10.1002/jms.1786
4. Varland, S.; Osberg, C.; Arnesen, T. N-terminal modifications of cellular proteins: the enzymes involved, their substrate specificities and biological effects. *Proteomics* **2015**, *15* (14), 2385-2401. DOI: 10.1002/pmic.201400619
5. Chuh, K. N.; Batt, A. R.; Pratt, M. R. Chemical methods for encoding and decoding of posttranslational modifications. *Cell Chem. Biol.* **2016**, *23* (1), 86-107. DOI: 10.1016/j.chembiol.2015.11.006
6. Tsikas, D. Post-translational modifications (PTM): analytical approaches, signaling, physiology and pathophysiology-part I. *Amino Acids* **2021**, *53* (4), 485-487. DOI: 10.1007/s00726-021-02984-y
7. Spoel, S. H. Orchestrating the proteome with post-translational modifications. *J. Exp. Bot.* **2018**, *69* (19), 4499-4503. DOI: 10.1093/jxb/ery295
8. Pan, S.; Chen, R. Pathological implication of protein post-translational modifications in cancer. *Mol. Asp. Med.* **2022**, *86*, 101097. DOI: 10.1016/j.mam.2022.101097
9. Zhong, Q.; Xiao, X.; Qiu, Y.; Xu, Z.; Chen, C.; Chong, B.; Zhao, X.; Hai, S.; Li, S.; An, Z.; et al. Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications. *MedComm* **2023**, *4* (3), e261. DOI: 10.1002/mco2.261
10. Wang, Y. C.; Peterson, S. E.; Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **2014**, *24* (2), 143-160. DOI: 10.1038/cr.2013.151
11. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-242. DOI: 10.1093/nar/28.1.235
12. Wood, C. M.; Sodngam, S.; Nicholson, J. M.; Lambert, S. J.; Reynolds, C. D.; Baldwin, J. P. Crystal structure of the native histone octamer to 2.1 angstrom resolution, crystalised in the presence of S-nitrosoglutathione. **2005**. DOI: 10.2210/pdb2aro/pdb
13. Wood, C. M.; Sodngam, S.; Nicholson, J. M.; Lambert, S. J.; Reynolds, C. D.; Baldwin, J. P. The oxidised histone octamer does not form a H3 disulphide bond. *Biochim. Biophys. Acta* **2006**, *1764* (8), 1356-1362. DOI: 10.1016/j.bbapap.2006.06.014
14. Chen, L.; Kashina, A. Post-translational modifications of the protein termini. *Front. Cell Dev. Biol.* **2021**, *9*, 719590. DOI: 10.3389/fcell.2021.719590
15. Marino, G.; Eckhard, U.; Overall, C. M. Protein termini and their modifications revealed by positional proteomics. *ACS Chem. Biol.* **2015**, *10* (8), 1754-1764. DOI: 10.1021/acschembio.5b00189
16. Schilling, O.; Barré, O.; Huesgen, P. F.; Overall, C. M. Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat. Methods* **2010**, *7*, 508-511. DOI: 10.1038/nmeth.1467
17. Muller, M. M. Post-translational modifications of protein backbones: unique functions, mechanisms, and challenges. *Biochemistry* **2018**, *57* (2), 177-185. DOI: 10.1021/acs.biochem.7b00861
18. Suskiewicz, M. J. The logic of protein post-translational modifications (PTMs): Chemistry, mechanisms and evolution of protein regulation through covalent attachments. *Bioessays* **2024**, *46* (3), 2300178. DOI: 10.1002/bies.202300178
19. Li, Z.; Li, S.; Luo, M.; Jhong, J. H.; Li, W.; Yao, L.; Pang, Y.; Wang, Z.; Wang, R.; Ma, R.; et al. dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications. *Nucleic Acids Res.* **2022**, *50* (D1), D471–D479. DOI: 10.1093/nar/gkab1017
20. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271-280. DOI: 10.3892/ijmm.2017.3036

21. Chen, M. J.; Dixon, J. E.; Manning, G. Genomics and evolution of protein phosphatases. *Sci. Signal.* **2017**, *10* (474), eaag1796. DOI: 10.1126/scisignal.aag1796
22. Yu, L.; Xu, L.; Xu, M.; Wan, B.; Yu, L.; Huang, Q. Role of Mg^{2+} ions in protein kinase phosphorylation: insights from molecular dynamics simulations of ATP-kinase complexes. *Mol. Simul.* **2011**, *37* (14), 1143-1150. DOI: 10.1080/08927022.2011.561430
23. Wang, Z.; Cole, P. A. Chapter One - Catalytic mechanisms and regulation of protein kinases. In *Protein Kinase Inhibitors in Research and Medicine*, Shokat, K. M. Ed.; Methods in Enzymology, Vol. 548; Academic Press, 2014; pp 1-21.
24. Shi, Y. Serine/threonine phosphatases: mechanism through structure. *Cell* **2009**, *139* (3), 468-484. DOI: 10.1016/j.cell.2009.10.006
25. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127* (3), 635-648. DOI: 10.1016/j.cell.2006.09.026
26. Seok, S. H. Structural insights into protein regulation by phosphorylation and substrate recognition of protein kinases/phosphatases. *Life* **2021**, *11* (9), 957. DOI: 10.3390/life11090957
27. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114* (13), 6589-6631. DOI: 10.1021/cr400525m
28. Endicott, J. A.; Noble, M. E.; Johnson, L. N. The structural basis for control of eukaryotic protein kinases. *Annu. Rev. Biochem.* **2012**, *81*, 587-613. DOI: 10.1146/annurev-biochem-052410-090317
29. Hunter, T. Why nature chose phosphate to modify proteins. *Phil. Trans. R. Soc. B* **2012**, *367* (1602), 2513-2516. DOI: 10.1098/rstb.2012.0013
30. Johnson, L. N. The regulation of protein phosphorylation. *Biochem. Soc. Trans.* **2009**, *37* (4), 627-641. DOI: 10.1042/BST0370627
31. Duronio, R. J.; Xiong, Y. Signaling pathways that control cell proliferation. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (3), a008904. DOI: 10.1101/cshperspect.a008904
32. Ward, P. S.; Thompson, C. B. Signaling in control of cell growth and metabolism. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (7), a006783. DOI: 10.1101/cshperspect.a006783
33. Yue, J.; López, J. M. Understanding MAPK signaling pathways in apoptosis. *Int. J. Mol. Sci.* **2020**, *21* (7), 2346. DOI: 10.3390/ijms21072346
34. Davis, C.; Spaller, B. L.; Matouschek, A. Mechanisms of substrate recognition by the 26S proteasome. *Curr. Opin. Struct. Biol.* **2021**, *67*, 161-169. DOI: 10.1016/j.sbi.2020.10.010
35. Pickart, C. M.; Eddins, M. J. Ubiquitin: structures, functions, mechanisms. *Biochim. Biophys. Acta* **2004**, *1695* (1-3), 55-72. DOI: 10.1016/j.bbamcr.2004.09.019
36. Etlinger, J. D.; Goldberg, A. L. A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74* (1), 54-58. DOI: 10.1073/pnas.74.1.54
37. Ciechanover, A.; Heller, H.; Elias, S.; Haas, A. L.; Hershko, A. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc Natl Acad Sci U S A* **1980**, *77* (3), 1365-1368. DOI: 10.1073/pnas.77.3.1365
38. Hershko, A.; Ciechanover, A.; Heller, H.; Haas, A. L.; Rose, I. A. Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77* (4), 1783-1786. DOI: 10.1073/pnas.77.4.1783
39. Ciechanover, A.; Elias, S.; Heller, H.; Hershko, A. "Covalent affinity" purification of ubiquitin-activating enzyme. *J. Biol. Chem.* **1982**, *257* (5), 2537-2542. DOI: 10.1016/S0021-9258(18)34957-3
40. Hershko, A.; Heller, H.; Elias, S.; Ciechanover, A. Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* **1983**, *258* (13), 8206-8214. DOI: 10.1016/S0021-9258(20)82050-X
41. Varshavsky, A. The early history of the ubiquitin field. *Protein Sci.* **2006**, *15* (3), 647-654. DOI: 10.1110/ps.052012306
42. Damgaard, R. B. The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities. *Cell Death Differ.* **2021**, *28* (2), 423-426. DOI: 10.1038/s41418-020-00703-w
43. Mackinnon, E.; Stone, S. L. The ubiquitin proteasome system and nutrient stress response. *Front. Plant Sci.* **2022**, *13*, 867419. DOI: 10.3389/fpls.2022.867419

44. Snyder, N. A.; Silva, G. M. Deubiquitinating enzymes (DUBs): regulation, homeostasis, and oxidative stress response. *J. Biol. Chem.* **2021**, *297* (3), 101077. DOI: 10.1016/j.jbc.2021.101077
45. Breitschopf, K.; Bengal, E.; Ziv, T.; Admon, A.; Ciechanover, A. A novel site for ubiquitination: the N-terminal residue, and not internal lysines of MyoD, is essential for conjugation and degradation of the protein. *EMBO J.* **1998**, *17* (20), 5964-5973. DOI: 10.1093/emboj/17.20.5964
46. Ben-Saadon, R.; Fajerman, I.; Ziv, T.; Hellman, U.; Schwartz, A. L.; Ciechanover, A. The tumor suppressor protein $p16^{INK4a}$ and the human papillomavirus oncoprotein-58 E7 are naturally occurring lysine-less proteins that are degraded by the ubiquitin system: direct evidence for ubiquitination at the N-terminal residue. *J. Biol. Chem.* **2004**, *279* (40), 41414-41421. DOI: 10.1074/jbc.M407201200
47. Pietrocola, F.; Galluzzi, L.; Bravo-San Pedro, J. M.; Madeo, F.; Kroemer, G. Acetyl coenzyme A: a central metabolite and second messenger. *Cell Metab.* **2015**, *21* (6), 805-821. DOI: 10.1016/j.cmet.2015.05.014
48. Drazic, A.; Myklebust, L. M.; Ree, R.; Arnesen, T. The world of protein acetylation. *Biochim. Biophys. Acta* **2016**, *1864* (10), 1372-1401. DOI: 10.1016/j.bbapap.2016.06.007
49. Arnesen, T.; Van Damme, P.; Polevoda, B.; Helsens, K.; Evjenth, R.; Colaert, N.; Varhaug, J. E.; Vandekerckhove, J.; Lillehaug, J. R.; Sherman, F.; et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (20), 8157-8162. DOI: 10.1073/pnas.0901931106
50. Lee, K. E.; Heo, J. E.; Kim, J. M.; Hwang, C. S. N-terminal acetylation-targeted N-end rule proteolytic system: the Ac/N-end rule pathway. *Mol. Cells* **2016**, *39* (3), 169-178. DOI: 10.14348/molcells.2016.2329
51. Aksnes, H.; Ree, R.; Arnesen, T. Co-translational, post-translational, and non-catalytic roles of N-terminal acetyltransferases. *Mol. Cell* **2019**, *73* (6), 1097-1114. DOI: 10.1016/j.molcel.2019.02.007
52. Hershko, A.; Heller, H.; Eytan, E.; Kaklij, G.; Rose, I. A. Role of the α-amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81* (22), 7021-7025. DOI: 10.1073/pnas.81.22.7021
53. Hwang, C. S.; Shemorry, A.; Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **2010**, *327* (5968), 973-977. DOI: 10.1126/science.1183147
54. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **2011**, *20* (8), 1298-1345. DOI: 10.1002/pro.666
55. Shemorry, A.; Hwang, C. S.; Varshavsky, A. Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* **2013**, *50* (4), 540-551. DOI: 10.1016/j.molcel.2013.03.018
56. Polevoda, B.; Sherman, F. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J. Mol. Biol.* **2003**, *325* (4), 595-622. DOI: 10.1016/s0022-2836(02)01269-x
57. Goetze, S.; Qeli, E.; Mosimann, C.; Staes, A.; Gerrits, B.; Roschitzki, B.; Mohanty, S.; Niederer, E. M.; Laczko, E.; Timmerman, E.; et al. Identification and functional characterization of N-terminally acetylated proteins in *Drosophila melanogaster*. *PLoS Biol.* **2009**, *7* (11), e1000236. DOI: 10.1371/journal.pbio.1000236
58. Phillips, D. M. The presence of acetyl groups of histones. *Biochem. J.* **1963**, *87* (2), 258-263. DOI: 10.1042/bj0870258
59. Allfrey, V. G.; Mirsky, A. E. Structural modifications of histones and their possible role in the regulation of RNA synthesis. *Science* **1964**, *144* (3618), 559-569. DOI: 10.1126/science.144.3618.559
60. Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *51* (5), 786-794. DOI: 10.1073/pnas.51.5.786
61. Allis, C. D.; Berger, S. L.; Cote, J.; Dent, S.; Jenuwien, T.; Kouzarides, T.; Pillus, L.; Reinberg, D.; Shi, Y.; Shiekhattar, R.; et al. New nomenclature for chromatin-modifying enzymes. *Cell* **2007**, *131* (4), 633-636. DOI: 10.1016/j.cell.2007.10.039
62. Kouzarides, T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* **2000**, *19* (6), 1176-1179. DOI: 10.1093/emboj/19.6.1176

63. Yang, X. J.; Grégoire, S. Metabolism, cytoskeleton and cellular signalling in the grip of protein N$^{\epsilon}$- and O-acetylation. *EMBO Rep.* **2007**, *8* (6), 556-562. DOI: 10.1038/sj.embor.7400977
64. Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; et al. Regulation of cellular metabolism by protein lysine acetylation. *Science* **2010**, *327* (5968), 1000-1004. DOI: 10.1126/science.1179689
65. Gu, W.; Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **1997**, *90* (4), 595-606. DOI: 10.1016/s0092-8674(00)80521-8
66. Sullivan, K. D.; Galbraith, M. D.; Andrysik, Z.; Espinosa, J. M. Mechanisms of transcriptional regulation by p53. *Cell Death Differ.* **2018**, *25* (1), 133-143. DOI: 10.1038/cdd.2017.174
67. Schemies, J.; Uciechowska, U.; Sippl, W.; Jung, M. NAD^{+}-dependent histone deacetylases (sirtuins) as novel therapeutic targets. *Med. Res. Rev.* **2010**, *30* (6), 861-889. DOI: 10.1002/med.20178
68. Dang, F.; Wei, W. Targeting the acetylation signaling pathway in cancer therapy. *Semin. Cancer Biol.* **2022**, *85*, 209-218. DOI: 10.1016/j.semcancer.2021.03.001
69. Shukla, S.; Tekwani, B. L. Histone deacetylases inhibitors in neurodegenerative diseases, neuroprotection and neuronal differentiation. *Front. Pharmacol.* **2020**, *11*, 537. DOI: 10.3389/fphar.2020.00537
70. Ree, R.; Varland, S.; Arnesen, T. Spotlight on protein N-terminal acetylation. *Exp. Mol. Med.* **2018**, *50* (7), 1-13. DOI: 10.1038/s12276-018-0116-z
71. Wagner, G. R.; Payne, R. M. Widespread and enzyme-independent N$^{\epsilon}$-acetylation and N$^{\epsilon}$-succinylation of proteins in the chemical conditions of the mitochondrial matrix. *J. Biol. Chem.* **2013**, *288* (40), 29036-29045. DOI: 10.1074/jbc.M113.486753
72. James, A. M.; Hoogewijs, K.; Logan, A.; Hall, A. R.; Ding, S.; Fearnley, I. M.; Murphy, M. P. Non-enzymatic N-acetylation of lysine residues by acetylCoA often occurs via a proximal S-acetylated thiol intermediate sensitive to Glyoxalase II. *Cell Rep.* **2017**, *18* (9), 2105-2112. DOI: 10.1016/j.celrep.2017.02.018
73. Clarke, S. Protein methylation. *Curr. Opin. Cell Biol.* **1993**, *5* (6), 977-983. DOI: 10.1016/0955-0674(93)90080-a
74. Schapira, M. Chemical inhibition of protein methyltransferases. *Cell Chem. Biol.* **2016**, *23* (9), 1067-1076. DOI: 10.1016/j.chembiol.2016.07.014
75. Zhang, J.; Zheng, Y. G. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chem. Biol.* **2016**, *11* (3), 583–597. DOI: 10.1021/acschembio.5b00812
76. Ambler, R. P.; Rees, M. W. ε-N-methyl-lysine in bacterial flagellar protein. *Nature* **1959**, *184*, 56-57. DOI: 10.1038/184056b0
77. Murray, K. The occurrence of ε-N-methyl lysine in histones. *Biochemistry* **1964**, *3*, 10-15. DOI: 10.1021/bi00889a003
78. Falnes, P. Ø.; Jakobsson, M. E.; Davydova, E.; Ho, A.; Małecki, J. Protein lysine methylation by seven-β-strand methyltransferases. *Biochem. J.* **2016**, *473* (14), 1995-2009. DOI: 10.1042/BCJ20160117
79. Husmann, D.; Gozani, O. Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **2019**, *26* (10), 880-889. DOI: 10.1038/s41594-019-0298-7
80. Maiques-Diaz, A.; Somervaille, T. C. LSD1: biologic roles and therapeutic targeting. *Epigenomics* **2016**, *8* (8), 1103-1116. DOI: 10.2217/epi-2016-0009
81. Hamamoto, R.; Saloura, V.; Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer* **2015**, *15* (2), 110-124. DOI: 10.1038/nrc3884
82. Zhang, J.; Jing, L.; Li, M.; He, L.; Guo, Z. Regulation of histone arginine methylation/demethylation by methylase and demethylase. *Mol. Med. Rep.* **2019**, *19* (5), 3963-3971. DOI: 10.3892/mmr.2019.10111
83. Bachand, F. Protein arginine methyltransferases: from unicellular eukaryotes to humans. *Eukaryot. Cell* **2007**, *6* (6), 889-898. DOI: 10.1128/EC.00099-07
84. Jain, K.; Warmack, R. A.; Debler, E. W.; Hadjikyriacou, A.; Stavropoulos, P.; Clarke, S. G. Protein arginine methyltransferase product specificity is mediated by distinct active-site architectures. *J. Biol. Chem.* **2016**, *291* (35), 18299-18308. DOI: 10.1074/jbc.M116.740399

85. Chang, B.; Chen, Y.; Zhao, Y.; Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science* **2007**, *318* (5849), 444-447. DOI: 10.1126/science.1145801
86. Walport, L. J.; Hopkinson, R. J.; Chowdhury, R.; Schiller, R.; Ge, W.; Kawamura, A.; Schofield, C. J. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **2016**, *7*, 11974. DOI: 10.1038/ncomms11974
87. Beaver, J. E.; Waters, M. L. Molecular recognition of Lys and Arg methylation. *ACS Chem. Biol.* **2016**, *11* (3), 643-653. DOI: 10.1021/acschembio.5b00996
88. Ruthenburg, A. J.; Allis, C. D.; Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **2007**, *25* (1), 15-30. DOI: 10.1016/j.molcel.2006.12.014
89. Padeken, J.; Methot, S. P.; Gasser, S. M. Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nature Reviews Molecular Cell Biology* **2022**, *23* (9), 623-640. DOI: 10.1038/s41580-022-00483-w
90. Migliori, V.; Müller, J.; Phalke, S.; Low, D.; Bezzi, M.; Mok, W. C.; Sahu, S. K.; Gunaratne, J.; Capasso, P.; Bassi, C.; et al. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat. Struct. Mol. Biol.* **2012**, *19* (2), 136-144. DOI: 10.1038/nsmb.2209
91. Kirmizis, A.; Santos-Rosa, H.; Penkett, C. J.; Singer, M. A.; Green, R. D.; Kouzarides, T. Distinct transcriptional outputs associated with mono- and dimethylated histone H3 arginine 2. *Nat. Struct. Mol. Biol.* **2009**, *16* (4), 449-451. DOI: 10.1038/nsmb.1569
92. Stock, A.; Clarke, S.; Clarke, C.; Stock, J. N-terminal methylation of proteins: structure, function and specificity. *FEBS Lett.* **1987**, *220* (1), 8-14. DOI: 10.1016/0014-5793(87)80866-9
93. Tooley, C. E.; Petkowski, J. J.; Muratore-Schroeder, T. L.; Balsbaugh, J. L.; Shabanowitz, J.; Sabat, M.; Minor, W.; Hunt, D. F.; Macara, I. G. NRMT is an α-N-methyltransferase that methylates RCC1 and retinoblastoma protein. *Nature* **2010**, *466* (7310), 1125-1128. DOI: 10.1038/nature09343
94. Webb, K. J.; Lipson, R. S.; Al-Hadid, Q.; Whitelegge, J. P.; Clarke, S. G. Identification of protein N-terminal methyltransferases in yeast and humans. *Biochemistry* **2010**, *49* (25), 5225-5235. DOI: 10.1021/bi100428x
95. Chen, P.; Huang, R.; Hazbun, T. R. Unlocking the mysteries of alpha-N-terminal methylation and its diverse regulatory functions. *J. Biol. Chem.* **2023**, *299* (7), 104843. DOI: 10.1016/j.jbc.2023.104843
96. Huang, R. Chemical biology of protein N-terminal methyltransferases. *ChemBioChem* **2019**, *20* (8), 976-984. DOI: 10.1002/cbic.201800615
97. Brown, T.; Nguyen, T.; Zhou, B.; Zheng, Y. G. Chemical probes and methods for the study of protein arginine methylation. *RSC Chem. Biol.* **2023**, *4* (9), 647-669. DOI: 10.1039/d3cb00018d
98. Chen, T.; Muratore, T. L.; Schaner-Tooley, C. E.; Shabanowitz, J.; Hunt, D. F.; Macara, I. G. N-terminal α-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* **2007**, *9* (5), 596-603. DOI: 10.1038/ncb1572
99. Petkowski, J. J.; Bonsignore, L. A.; Tooley, J. G.; Wilkey, D. W.; Merchant, M. L.; Macara, I. G.; Schaner Tooley, C. E. NRMT2 is an N-terminal monomethylase that primes for its homologue NRMT1. *Biochem. J.* **2013**, *456* (3), 453-462. DOI: 10.1042/BJ20131163
100. Petkowski, J. J.; Schaner Tooley, C. E.; Anderson, L. C.; Shumilin, I. A.; Balsbaugh, J. L.; Shabanowitz, J.; Hunt, D. F.; Minor, W.; Macara, I. G. Substrate specificity of mammalian N-terminal α-amino methyltransferase NRMT. *Biochemistry* **2012**, *51* (30), 5942–5950. DOI: 10.1021/bi300278f
101. Giglione, C.; Boularot, A.; Meinnel, T. Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **2004**, *61* (12), 1455-1474. DOI: 10.1007/s00018-004-3466-8
102. Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T. The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics* **2006**, *5* (12), 2336-2349. DOI: 10.1074/mcp.M600225-MCP200
103. Venkatesh, S.; Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (3), 178-189. DOI: 10.1038/nrm3941
104. McGinty, R. K.; Tan, S. Nucleosome structure and function. *Chem. Rev.* **2015**, *115* (6), 2255-2273. DOI: 10.1021/cr500373h
105. Arents, G.; Burlingame, R. W.; Wang, B. C.; Love, W. E.; Moudrianakis, E. N. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a

left-handed superhelix. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88* (22), 10148-10152. DOI: 10.1073/pnas.88.22.10148
106. Corujo, D.; Buschbeck, M. Post-translational modifications of H2A histone variants and their role in cancer. *Cancers* **2018**, *10* (3), 59. DOI: 10.3390/cancers10030059
107. Luger, K.; Dechassa, M. L.; Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (7), 436-447. DOI: 10.1038/nrm3382
108. Polach, K. J.; Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **1995**, *254* (2), 130-149. DOI: 10.1006/jmbi.1995.0606
109. Cutter, A. R.; Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **2015**, *589* (20 Pt A), 2914-2922. DOI: 10.1016/j.febslet.2015.05.016
110. Lai, W. K. M.; Pugh, B. F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (9), 548-562. DOI: 10.1038/nrm.2017.47
111. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thåström, A.; Field, Y.; Moore, I. K.; Wang, J.-P. Z.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442* (7104), 772-778. DOI: 10.1038/nature04979
112. Struhl, K.; Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **2013**, *20* (3), 267-273. DOI: 10.1038/nsmb.2506
113. Nelson, H. C.; Finch, J. T.; Luisi, B. F.; Klug, A. The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature* **1987**, *330* (6145), 221-226. DOI: 10.1038/330221a0
114. Suter, B.; Schnappauf, G.; Thoma, F. Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.* **2000**, *28* (21), 4083-4089. DOI: 10.1093/nar/28.21.4083
115. Segal, E.; Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 65-71. DOI: 10.1016/j.sbi.2009.01.004
116. Tessarz, P.; Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (11), 703-708. DOI: 10.1038/nrm3890
117. Masumoto, H.; Hawke, D.; Kobayashi, R.; Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **2005**, *436* (7048), 294-298. DOI: 10.1038/nature03714
118. Li, Q.; Zhou, H.; Wurtele, H.; Davies, B.; Horazdovsky, B.; Verreault, A.; Zhang, Z. Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* **2008**, *134* (2), 244-255. DOI: 10.1016/j.cell.2008.06.018
119. Williams, S. K.; Truong, D.; Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (26), 9000-9005. DOI: 10.1073/pnas.0800057105
120. Kurat, C. F.; Recht, J.; Radovani, E.; Durbic, T.; Andrews, B.; Fillingham, J. Regulation of histone gene transcription in yeast. *Cell. Mol. Life Sci.* **2014**, *71* (4), 599-613. DOI: 10.1007/s00018-013-1443-9
121. Talbert, P. B.; Henikoff, S. Histone variants on the move: substrates for chromatin dynamics. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (2), 115-126. DOI: 10.1038/nrm.2016.148
122. Lee, K.-M.; Narlikar, G. Assembly of nucleosomal templates by salt dialysis. *Curr. Protoc. Mol. Biol.* **2001**, *54* (1). DOI: 10.1002/0471142727.mb2106s54
123. Luger, K.; Rechsteiner, T. J.; Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. In *Chromatin*, Wassarman, P. M., Wolffe, A. P. Eds.; Methods in Enzymology, Vol. 304; Academic Press, 1999; pp 3-19.
124. Laskey, R. A.; Honda, B. M.; Mills, A. D.; Finch, J. T. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **1978**, *275* (5679), 416-420. DOI: 10.1038/275416a0
125. Gurard-Levin, Z. A.; Quivy, J. P.; Almouzni, G. Histone chaperones: assisting histone traffic and nucleosome dynamics. *Annu. Rev. Biochem.* **2014**, *83*, 487-517. DOI: 10.1146/annurev-biochem-060713-035536
126. Park, Y.-J.; Luger, K. The structure of nucleosome assembly protein 1. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (5), 1248-1253. DOI: 10.1073/pnas.0508002103
127. Bernardes, N. E.; Fung, H. Y. J.; Li, Y.; Chen, Z.; Chook, Y. M. Structure of IMPORTIN-4 bound to the H3-H4-ASF1 histone-histone chaperone complex. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (38), e2207177119. DOI: 10.1073/pnas.2207177119

128. Cook, A. J. L.; Gurard-Levin, Z. A.; Vassias, I.; Almouzni, G. A specific function for the histone chaperone NASP to fine-tune a reservoir of soluble H3-H4 in the histone supply chain. *Mol. Cell* **2011**, *44* (6), 918-927. DOI: 10.1016/j.molcel.2011.11.021
129. Han, J.; Zhou, H.; Li, Z.; Xu, R. M.; Zhang, Z. Acetylation of lysine 56 of histone H3 catalyzed by RTT109 and regulated by ASF1 is required for replisome integrity. *J. Biol. Chem.* **2007**, *282* (39), 28587-28596. DOI: 10.1074/jbc.M702496200
130. Sauer, P. V.; Gu, Y.; Liu, W. H.; Mattiroli, F.; Panne, D.; Luger, K.; Churchill, M. E. Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Res.* **2018**, *46* (19), 9907-9917. DOI: 10.1093/nar/gky823
131. Mazina, M. Y.; Vorobyeva, N. E. Chromatin modifiers in transcriptional regulation: new findings and prospects. *Acta Naturae* **2021**, *13* (1), 16-30. DOI: 10.32607/actanaturae.11101
132. Tang, L.; Nogales, E.; Ciferri, C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* **2010**, *102* (2-3), 122-128. DOI: 10.1016/j.pbiomolbio.2010.05.001
133. Clapier, C. R.; Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **2009**, *78*, 273-304. DOI: 10.1146/annurev.biochem.77.062706.153223
134. Zhang, W.; Feng, J.; Li, Q. The replisome guides nucleosome assembly during DNA replication. *Cell Biosci.* **2020**, *10*, 37. DOI: 10.1186/s13578-020-00398-z
135. Jian, Y.; Shim, W.-B.; Ma, Z. Multiple functions of SWI/SNF chromatin remodeling complex in plant-pathogen interactions. *Stress Biol.* **2021**, *1* (1), 18. DOI: 10.1007/s44154-021-00019-w
136. Patel, A. B.; Moore, C. M.; Greber, B. J.; Luo, J.; Zukin, S. A.; Ranish, J.; Nogales, E. Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. *eLife* **2019**, *8*, e54449. DOI: 10.7554/eLife.54449
137. Clapier, C. R.; Kasten, M. M.; Parnell, T. J.; Viswanathan, R.; Szerlong, H.; Sirinakis, G.; Zhang, Y.; Cairns, B. R. Regulation of DNA translocation efficiency within the chromatin remodeler RSC/Sth1 potentiates nucleosome sliding and ejection. *Mol. Cell* **2016**, *62* (3), 453-461. DOI: 10.1016/j.molcel.2016.03.032
138. Alver, B. H.; Kim, K. H.; Lu, P.; Wang, X.; Manchester, H. E.; Wang, W.; Haswell, J. R.; Park, P. J.; Roberts, C. W. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* **2017**, *8*, 14648. DOI: 10.1038/ncomms14648
139. Mittal, P.; Roberts, C. W. M. The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat. Rev. Clin. Oncol.* **2020**, *17* (7), 435-448. DOI: 10.1038/s41571-020-0357-3
140. Gangaraju, V. K.; Bartholomew, B. Mechanisms of ATP dependent chromatin remodeling. *Mutat. Res.* **2007**, *618* (1-2), 3-17. DOI: 10.1016/j.mrfmmm.2006.08.015
141. Mills, A. A. The chromodomain helicase DNA-binding chromatin remodelers: family traits that protect from and promote cancer. *Cold Spring Harb. Perspect. Med.* **2017**, *7* (4), a026450. DOI: 10.1101/cshperspect.a026450
142. Murawska, M.; Brehm, A. CHD chromatin remodelers and the transcription cycle. *Transcription* **2011**, *2* (6), 244-253. DOI: 10.4161/trns.2.6.17840
143. Manning, B. J.; Yusufzai, T. The ATP-dependent chromatin remodeling enzymes CHD6, CHD7, and CHD8 exhibit distinct nucleosome binding and remodeling activities. *J. Biol. Chem.* **2017**, *292* (28), 11927-11936. DOI: 10.1074/jbc.M117.779470
144. Zhang, W.; Aubert, A.; Gomez de Segura, J. M.; Karuppasamy, M.; Basu, S.; Murthy, A. S.; Diamante, A.; Drury, T. A.; Balmer, J.; Cramard, J.; et al. The nucleosome remodeling and deacetylase complex NuRD is built from preformed catalytically active sub-modules. *J. Mol. Biol.* **2016**, *428* (14), 2931-2942. DOI: 10.1016/j.jmb.2016.04.025
145. Shimada, K.; Oma, Y.; Schleker, T.; Kugou, K.; Ohta, K.; Harata, M.; Gasser, S. M. Ino80 chromatin remodeling complex promotes recovery of stalled replication forks. *Curr. Biol.* **2008**, *18* (8), 566-575. DOI: 10.1016/j.cub.2008.03.049
146. Poli, J.; Gerhold, C. B.; Tosi, A.; Hustedt, N.; Seeber, A.; Sack, R.; Herzog, F.; Pasero, P.; Shimada, K.; Hopfner, K. P.; et al. Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication conflicts through RNAPII removal during replication stress. *Genes Dev.* **2016**, *30* (3), 337-354. DOI: 10.1101/gad.273813.115
147. Clapier, C. R.; Iwasa, J.; Cairns, B. R.; Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (7), 407-422. DOI: 10.1038/nrm.2017.26

148. Poli, J.; Gasser, S. M.; Papamichos-Chronakis, M. The INO80 remodeller in transcription, replication and repair. *Phil. Trans. R. Soc. B* **2017**, *372* (1731), 20160290. DOI: 10.1098/rstb.2016.0290
149. Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **2012**, *11* (5), 384-400. DOI: 10.1038/nrd3674
150. Trerotola, M.; Relli, V.; Simeone, P.; Alberti, S. Epigenetic inheritance and the missing heritability. *Hum. Genomics.* **2015**, *9* (1), 17. DOI: 10.1186/s40246-015-0041-3
151. Turner, B. M. Decoding the nucleosome. *Cell* **1993**, *75* (1), 5-8. DOI: 10.1016/s0092-8674(05)80078-9
152. Strahl, B. D.; Allis, C. D. The language of covalent histone modifications. *Nature* **2000**, *403* (6765), 41-45. DOI: 10.1038/47412
153. Govin, J.; Caron, C.; Lestrat, C.; Rousseaux, S.; Khochbin, S. The role of histones in chromatin remodelling during mammalian spermiogenesis. *Eur. J. Biochem.* **2004**, *271* (17), 3459-3469. DOI: 10.1111/j.1432-1033.2004.04266.x
154. Agalioti, T.; Chen, G.; Thanos, D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **2002**, *111* (3), 381-392. DOI: 10.1016/s0092-8674(02)01077-2
155. Liang, G.; Lin, J. C. Y.; Wei, V.; Yoo, C.; Cheng, J. C.; Nguyen, C. T.; Weisenberger, D. J.; Egger, G.; Takai, D.; Gonzales, F. A.; et al. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (19), 7357-7362. DOI: 10.1073/pnas.0401866101
156. Wang, Z.; Zang, C.; Rosenfeld, J. A.; Schones, D. E.; Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Peng, W.; Zhang, M. Q.; et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **2008**, *40* (7), 897-903. DOI: 10.1038/ng.154
157. Karmodiya, K.; Krebs, A. R.; Oulad-Abdelghani, M.; Kimura, H.; Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **2012**, *13*, 424. DOI: 10.1186/1471-2164-13-424
158. Jacobson, R. H.; Ladurner, A. G.; King, D. S.; Tjian, R. Structure and function of a human $TAF_{II}250$ double bromodomain module. *Science* **2000**, *288* (5470), 1422-1425. DOI: 10.1126/science.288.5470.1422
159. Gates, L. A.; Shi, J.; Rohira, A. D.; Feng, Q.; Zhu, B.; Bedford, M. T.; Sagum, C. A.; Jung, S. Y.; Qin, J.; Tsai, M.-J.; et al. Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.* **2017**, *292* (35), 14456-14472. DOI: 10.1074/jbc.M117.802074
160. Winston, F.; Allis, C. D. The bromodomain: a chromatin-targeting module? *Nat. Struct. Mol. Biol.* **1999**, *6* (7), 601-604. DOI: 10.1038/10640
161. Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M.-M. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **1999**, *399* (6735), 491-496. DOI: 10.1038/20974
162. Josling, G. A.; Selvarajah, S. A.; Petter, M.; Duffy, M. F. The role of bromodomain proteins in regulating gene expression. *Genes* **2012**, *3* (2), 320-343. DOI: 10.3390/genes3020320
163. Trojer, P. Targeting BET bromodomains in cancer. *Annu. Rev. Cancer Biol.* **2022**, *6* (1), 313-336. DOI: 10.1146/annurev-cancerbio-070120-103531
164. Zaware, N.; Zhou, M.-M. Bromodomain biology and drug discovery. *Nat. Struct. Mol. Biol.* **2019**, *26* (10), 870-879. DOI: 10.1038/s41594-019-0309-8
165. Filippakopoulos, P.; Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.* **2014**, *13* (5), 337-356. DOI: 10.1038/nrd4286
166. Zeng, L.; Zhang, Q.; Li, S.; Plotnikov, A. N.; Walsh, M. J.; Zhou, M.-M. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **2010**, *466* (7303), 258-262. DOI: 10.1038/nature09139
167. Ali, M.; Yan, K.; Lalonde, M.-E.; Degerny, C.; Rothbart, S. B.; Strahl, B. D.; Côté, J.; Yang, X.-J.; Kutateladze, T. G. Tandem PHD fingers of MORF/MOZ acetyltransferases display selectivity for acetylated histone H3 and are required for the association with chromatin. *J. Mol. Biol.* **2012**, *424* (5), 328-338. DOI: 10.1016/j.jmb.2012.10.004
168. Qiu, Y.; Liu, L.; Zhao, C.; Han, C.; Li, F.; Zhang, J.; Wang, Y.; Li, G.; Mei, Y.; Wu, M.; et al. Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev.* **2012**, *26* (12), 1376-1391. DOI: 10.1101/gad.188359.112

169. Su, D.; Hu, Q.; Li, Q.; Thompson, J. R.; Cui, G.; Fazly, A.; Davies, B. A.; Botuyan, M. V.; Zhang, Z.; Mer, G. Structural basis for recognition of H3K56-acetylated histone H3-H4 by the chaperone Rtt106. *Nature* **2012**, *483* (7387), 104-107. DOI: 10.1038/nature10861
170. Zhao, D.; Li, Y.; Xiong, X.; Chen, Z.; Li, H. YEATS domain - a histone acylation reader in health and disease. *J. Mol. Biol.* **2017**, *429* (13), 1994-2002. DOI: 10.1016/j.jmb.2017.03.010
171. Yap, K. L.; Zhou, M.-M. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry* **2011**, *50* (12), 1966-1980. DOI: 10.1021/bi101885m
172. Maity, S. K.; Jbara, M.; Brik, A. Chemical and semisynthesis of modified histones. *J. Pept. Sci.* **2016**, *22* (5), 252-259. DOI: 10.1002/psc.2848
173. Barber, K. W.; Rinehart, J. The ABCs of PTMs. *Nat. Chem. Biol.* **2018**, *14* (3), 188-192. DOI: 10.1038/nchembio.2572
174. Graves, H. K.; Wang, P.; Lagarde, M.; Chen, Z.; Tyler, J. K. Mutations that prevent or mimic persistent post-translational modifications of the histone H3 globular domain cause lethality and growth defects in *Drosophila*. *Epigenetics Chromatin* **2016**, *9*, 9. DOI: 10.1186/s13072-016-0059-3
175. Kamieniarz, K.; Schneider, R. Tools to tackle protein acetylation. *Chem. Biol.* **2009**, *16* (10), 1027-1029. DOI: 10.1016/j.chembiol.2009.10.002
176. Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S.-G. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One* **2012**, *7* (7), e40410. DOI: 10.1371/journal.pone.0040410
177. Weirich, S.; Jeltsch, A. Limited choice of natural amino acids as mimetics restricts design of protein lysine methylation studies. *Nat. Commun.* **2023**, *14* (1), 4097. DOI: 10.1038/s41467-023-39777-8
178. Fujimoto, H.; Higuchi, M.; Koike, M.; Ode, H.; Pinak, M.; Bunta, J. K.; Nemoto, T.; Sakudoh, T.; Honda, N.; Maekawa, H.; et al. A possible overestimation of the effect of acetylation on lysine residues in KQ mutant analysis. *J. Comput. Chem.* **2012**, *33* (3), 239-246. DOI: 10.1002/jcc.21956
179. David, Y.; Muir, T. W. Emerging chemistry strategies for engineering native chromatin. *J. Am. Chem. Soc.* **2017**, *139* (27), 9090-9096. DOI: 10.1021/jacs.7b03430
180. Bachman, J. Chapter Nineteen - Site-Directed Mutagenesis. In *Laboratory Methods in Enzymology: DNA*, Lorsch, J. Ed.; Methods in Enzymology, Vol. 529; Academic Press, 2013; pp 241-248.
181. Chalker, J. M.; Davis, B. G. Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **2010**, *14* (6), 781-789. DOI: 10.1016/j.cbpa.2010.10.007
182. Zioudrou, C.; Wilchek, M.; Patchornik, A. Conversion of the L-serine residue to an L-cysteine residue in peptides*. *Biochemistry* **2002**, *4* (9), 1811-1822. DOI: 10.1021/bi00885a018
183. Polgar, L.; Bender, M. L. A new enzyme containing a synthetically formed active site. Thiol-subtilisin. *J. Am. Chem. Soc.* **1966**, *88* (13), 3153-3154. DOI: 10.1021/ja00965a060
184. Neet, K. E.; Koshland, D. E., Jr. The conversion of serine at the active site of subtilisin to cysteine: a "chemical mutation". *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *56* (5), 1606-1611. DOI: 10.1073/pnas.56.5.1606
185. Sornay, C.; Vaur, V.; Wagner, A.; Chaubet, G. An overview of chemo- and site-selectivity aspects in the chemical conjugation of proteins. *R. Soc. Open Sci.* **2022**, *9* (1), 211563. DOI: 10.1098/rsos.211563
186. Hintzen, J. C. J.; Mecinović, J. Synthetic cysteine-based tools for probing protein posttranslational modifications. *Tetrahedron Lett.* **2023**, *124*, 154602. DOI: 10.1016/j.tetlet.2023.154602
187. Qi, Y.-K.; Ai, H.-S.; Li, Y.-M.; Yan, B. Total chemical synthesis of modified histones. *Front. Chem.* **2018**, *6*, 19. DOI: 10.3389/fchem.2018.00019
188. Dadová, J.; Galan, S. R.; Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **2018**, *46*, 71-81. DOI: 10.1016/j.cbpa.2018.05.022
189. Huang, R.; Holbert, M. A.; Tarrant, M. K.; Curtet, S.; Colquhoun, D. R.; Dancy, B. M.; Dancy, B. C.; Hwang, Y.; Tang, Y.; Meeth, K.; et al. Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J. Am. Chem. Soc.* **2010**, *132* (29), 9986-9987. DOI: 10.1021/ja103954u

190. Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C.-F. A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed.* **2011**, *50* (41), 9611-9614. DOI: 10.1002/anie.201103754
191. Griffiths, R. C.; Smith, F. R.; Long, J. E.; Scott, D.; Williams, H. E. L.; Oldham, N. J.; Layfield, R.; Mitchell, N. J. Site-selective installation of N^{ϵ}-modified sidechains into peptide and protein scaffolds via visible-light-mediated desulfurative C-C bond formation. *Angew. Chem. Int. Ed.* **2022**, *61* (2), e202110223. DOI: 10.1002/anie.202110223
192. Fu, X.-P.; Yuan, Y.; Jha, A.; Levin, N.; Giltrap, A. M.; Ren, J.; Mamalis, D.; Mohammed, S.; Davis, B. G. Stereoretentive post-translational protein editing. *ACS Cent. Sci.* **2023**, *9* (3), 405-416. DOI: 10.1021/acscentsci.2c00991
193. Hewings, D. S.; Wang, M.; Philpott, M.; Fedorov, O.; Uttarkar, S.; Filippakopoulos, P.; Picaud, S.; Vuppusetty, C.; Marsden, B.; Knapp, S.; et al. 3,5-Dimethylisoxazoles act as acetyl-lysine-mimetic bromodomain ligands. *J. Med. Chem.* **2011**, *54* (19), 6761-6770. DOI: 10.1021/jm200640v
194. Sekirnik (née Measures), A. R.; Hewings, D. S.; Theodoulou, N. H.; Jursins, L.; Lewendon, K. R.; Jennings, L. E.; Rooney, T. P. C.; Heightman, T. D.; Conway, S. J. Isoxazole-derived amino acids are bromodomain-binding acetyl-lysine mimics: incorporation into histone H4 peptides and histone H3. *Angew. Chem. Int. Ed.* **2016**, *55* (29), 8353-8357. DOI: 10.1002/anie.201602908
195. Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev.* **2017**, *117* (8), 5457-5520. DOI: 10.1021/acs.chemrev.6b00591
196. Wang, Z.; Lyons, B.; Truscott, R. J. W.; Schey, K. L. Human protein aging: modification and crosslinking through dehydroalanine and dehydrobutyrine intermediates. *Aging Cell* **2014**, *13* (2), 226-234. DOI: 10.1111/acel.12164
197. Humphrey, J. M.; Chamberlin, A. R. Chemical synthesis of natural product peptides: coupling methods for the incorporation of noncoded amino acids into peptides. *Chem. Rev.* **1997**, *97* (6), 2243-2266. DOI: 10.1021/cr950005s
198. Strumeyer, D. H.; White, W. N.; Koshland, D. E., Jr. Role of serine in chymotrypsin action. Conversion of the active serine to dehydroalanine. *Proc. Natl. Acad. Sci. U.S.A.* **1963**, *50* (5), 931-935. DOI: 10.1073/pnas.50.5.931
199. Weiner, H.; White, W. N.; Hoare, D. G.; Koshland Jr., D. E. The formation of anhydrochymotrypsin by removing the elements of water from the serine at the active site. *J. Am. Chem. Soc.* **1966**, *88* (16), 3851-3859. DOI: 10.1021/ja00968a033
200. Ekici, Ö. D.; Paetzel, M.; Dalbey, R. E. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci.* **2008**, *17* (12), 2023-2037. DOI: 10.1110/ps.035436.108
201. Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Söll, D.; Lee, H.-Y.; Park, H.-S. A chemical biology route to site-specific authentic protein modifications. *Science* **2016**, *354* (6312), 623-626. DOI: 10.1126/science.aah4428
202. Zhao, Z.; Shimon, D.; Metanis, N. Chemoselective copper-mediated modification of selenocysteines in peptides and proteins. *J. Am. Chem. Soc.* **2021**, *143* (32), 12817-12824. DOI: 10.1021/jacs.1c06101
203. Mobli, M.; Morgenstern, D.; King, G. F.; Alewood, P. F.; Muttenthaler, M. Site-specific pK_a determination of selenocysteine residues in selenovasopressin by using ^{77}Se NMR spectroscopy. *Angew. Chem. Int. Ed.* **2011**, *50* (50), 11952-11955. DOI: 10.1002/anie.201104169
204. Reich, H. J.; Hondal, R. J. Why Nature chose selenium. *ACS Chem. Biol.* **2016**, *11* (4), 821-841. DOI: 10.1021/acschembio.6b00031
205. Metanis, N.; Hilvert, D. Natural and synthetic selenoproteins. *Curr. Opin. Chem. Biol.* **2014**, *22*, 27-34. DOI: 10.1016/j.cbpa.2014.09.010
206. Liu, J.; Chen, Q.; Rozovsky, S. Utilizing selenocysteine for expressed protein ligation and bioconjugations. *J. Am. Chem. Soc.* **2017**, *139* (9), 3430-3437. DOI: 10.1021/jacs.6b10991
207. Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed.* **2008**, *47* (34), 6399-6401. DOI: 10.1002/anie.200802336
208. Wang, Z. U.; Wang, Y.-S.; Pai, P.-J.; Russell, W. K.; Russell, D. H.; Liu, W. R. A facile method to synthesize histones with posttranslational modification mimics. *Biochemistry* **2012**, *51* (26), 5232-5234. DOI: 10.1021/bi300535a

209. Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2011**, *2* (9), 1666-1676. DOI: 10.1039/c1sc00185j
210. Dadová, J.; Wu, K.-J.; Isenegger, P. G.; Errey, J. C.; Bernardes, G. J. L.; Chalker, J. M.; Raich, L.; Rovira, C.; Davis, B. G. Precise probing of residue roles by post-translational β,γ-C,N aza-Michael mutagenesis in enzyme active sites. *ACS Cent. Sci.* **2017**, *3* (11), 1168-1173. DOI: 10.1021/acscentsci.7b00341
211. 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. *Angew. Chem. Int. Ed.* **2012**, *51* (8), 1835-1839. DOI: 10.1002/anie.201106432
212. Josephson, B.; Fehl, C.; Isenegger, P. G.; Nadal, S.; Wright, T. H.; Poh, A. W. J.; Bower, B. J.; Giltrap, A. M.; Chen, L.; Batchelor-McAuley, C.; et al. Light-driven post-translational installation of reactive protein side chains. *Nature* **2020**, *585* (7826), 530-537. DOI: 10.1038/s41586-020-2733-7
213. Gomez, M. A. R.; Ibba, M. Aminoacyl-tRNA synthetases. *RNA* **2020**, *26* (8), 910-936. DOI: 10.1261/rna.071720.119
214. Rodnina, M. V.; Beringer, M.; Wintermeyer, W. How ribosomes make peptide bonds. *Trends Biochem. Sci.* **2007**, *32* (1), 20-26. DOI: 10.1016/j.tibs.2006.11.007
215. Beringer, M.; Rodnina, M. V. The ribosomal peptidyl transferase. *Mol. Cell* **2007**, *26* (3), 311-321. DOI: 10.1016/j.molcel.2007.03.015
216. Leisle, L.; Valiyaveetil, F.; Mehl, R. A.; Ahern, C. A. Incorporation of non-canonical amino acids. In *Novel Chemical Tools to Study Ion Channel Biology*, Ahern, C. A., Pless, S. Eds.; Advances in Experimental Medicine and Biology, Vol. 869; Springer, 2015; pp 119-151.
217. Chaudhuri, B. N.; Yeates, T. O. A computational method to predict genetically encoded rare amino acids in proteins. *Genome Biol.* **2005**, *6* (9), R79. DOI: 10.1186/gb-2005-6-9-r79
218. Mariotti, M.; Salinas, G.; Gabaldón, T.; Gladyshev, V. N. Utilization of selenocysteine in early-branching fungal phyla. *Nat Microbiol.* **2019**, *4* (5), 759-765. DOI: 10.1038/s41564-018-0354-9
219. Fu, X.; Crnković, A.; Sevostyanova, A.; Söll, D. Designing seryl-tRNA synthetase for improved serylation of selenocysteine tRNAs. *FEBS Lett.* **2018**, *592* (22), 3759-3768. DOI: 10.1002/1873-3468.13271
220. Fischer, N.; Neumann, P.; Bock, L. V.; Maracci, C.; Wang, Z.; Paleskava, A.; Konevega, A. L.; Schröder, G. F.; Grubmüller, H.; Ficner, R.; et al. The pathway to GTPase activation of elongation factor SelB on the ribosome. *Nature* **2016**, *540* (7631), 80-85. DOI: 10.1038/nature20560
221. Yuan, J.; O'Donoghue, P.; Ambrogelly, A.; Gundllapalli, S.; Sherrer, R. L.; Palioura, S.; Simonović, M.; Söll, D. Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett.* **2010**, *584* (2), 342-349. DOI: 10.1016/j.febslet.2009.11.005
222. Ho, J. M. L.; Miller, C. A.; Smith, K. A.; Mattia, J. R.; Bennett, M. R. Improved pyrrolysine biosynthesis through phage assisted non-continuous directed evolution of the complete pathway. *Nat. Commun.* **2021**, *12* (1), 3914. DOI: 10.1038/s41467-021-24183-9
223. Gaston, M. A.; Zhang, L.; Green-Church, K. B.; Krzycki, J. A. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **2011**, *471* (7340), 647-650. DOI: 10.1038/nature09918
224. Tharp, J. M.; Ehnbom, A.; Liu, W. R. $tRNA^{Pyl}$: structure, function, and applications. *RNA Biol.* **2018**, *15* (4-5), 441-452. DOI: 10.1080/15476286.2017.1356561
225. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **2017**, *550* (7674), 53-60. DOI: 10.1038/nature24031
226. Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **2014**, *1844* (6), 1059-1070. DOI: 10.1016/j.bbapap.2014.03.002
227. Crnković, A.; Suzuki, T.; Söll, D.; Reynolds, N. M. Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. *Croat. Chem. Acta* **2016**, *89* (2), 163-174. DOI: 10.5562/cca2825

228. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* **2014**, *12* (1), 35-48. DOI: 10.1038/nrmicro3155
229. Conibear, A. C. Deciphering protein post-translational modifications using chemical biology tools. *Nat. Rev. Chem.* **2020**, *4* (12), 674-695. DOI: 10.1038/s41570-020-00223-8
230. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically encoding N^{ε}-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008**, *4* (4), 232-234. DOI: 10.1038/nchembio.73
231. Neumann, H.; Hancock, S. M.; Buning, R.; Routh, A.; Chapman, L.; Somers, J.; Owen-Hughes, T.; van Noort, J.; Rhodes, D.; Chin, J. W. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* **2009**, *36* (1), 153-163. DOI: 10.1016/j.molcel.2009.07.027
232. Algar, W. R.; Hildebrandt, N.; Vogel, S. S.; Medintz, I. L. FRET as a biomolecular research tool - understanding its potential while avoiding pitfalls. *Nat. Methods* **2019**, *16* (9), 815-829. DOI: 10.1038/s41592-019-0530-8
233. Xu, F.; Zhang, K.; Grunstein, M. Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell* **2005**, *121* (3), 375-385. DOI: 10.1016/j.cell.2005.03.011
234. Hellman, L. M.; Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2007**, *2* (8), 1849-1861. DOI: 10.1038/nprot.2007.249
235. Bruno, M.; Flaus, A.; Stockdale, C.; Rencurel, C.; Ferreira, H.; Owen-Hughes, T. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol. Cell* **2003**, *12* (6), 1599-1606. DOI: 10.1016/s1097-2765(03)00499-4
236. Wang, Z. A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y.-J.; Sharma, V.; Lin, H.; Dai, S. Y.; Liu, W. R. A versatile approach for site-specific lysine acylation in proteins. *Angew. Chem. Int. Ed.* **2017**, *56* (6), 1643-1647. DOI: 10.1002/anie.201611415
237. Shandell, M. A.; Tan, Z.; Cornish, V. W. Genetic code expansion: a brief history and perspective. *Biochemistry* **2021**, *60* (46), 3455-3469. DOI: 10.1021/acs.biochem.1c00286
238. González, S. S.; Ad, O.; Shah, B.; Zhang, Z.; Zhang, X.; Chatterjee, A.; Schepartz, A. Genetic code expansion in the engineered organism Vmax X2: high yield and exceptional fidelity. *ACS Cent. Sci.* **2021**, *7* (9), 1500-1507. DOI: 10.1021/acscentsci.1c00499
239. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (20), 7566-7571. DOI: 10.1073/pnas.0401517101
240. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464* (7287), 441-444. DOI: 10.1038/nature08817
241. Zhang, Y.; Ptacin, J. L.; Fischer, E. C.; Aerni, H. R.; Caffaro, C. E.; Jose, K. S.; Feldman, A. W.; Turner, C. R.; Romesberg, F. E. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **2017**, *551* (7682), 644-647. DOI: 10.1038/nature24659
242. Fischer, E. C.; Hashimoto, K.; Zhang, Y.; Feldman, A. W.; Dien, V. T.; Karadeema, R. J.; Adhikary, R.; Ledbetter, M. P.; Krishnamurthy, R.; Romesberg, F. E. New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat. Chem. Biol.* **2020**, *16* (5), 570-576. DOI: 10.1038/s41589-020-0507-z
243. Fredens, J.; Wang, K.; de la Torre, D.; Funke, L. F. H.; Robertson, W. E.; Christova, Y.; Chia, T.; Schmied, W. H.; Dunkelmann, D. L.; Beránek, V.; et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **2019**, *569* (7757), 514-518. DOI: 10.1038/s41586-019-1192-5
244. Petropoulos, A. D.; McDonald, M. E.; Green, R.; Zaher, H. S. Distinct roles for release factor 1 and release factor 2 in translational quality control. *J. Biol. Chem.* **2014**, *289* (25), 17589-17596. DOI: 10.1074/jbc.M114.564989
245. Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; et al. Genomically recoded organisms expand biological functions. *Science* **2013**, *342* (6156), 357-360. DOI: 10.1126/science.1241459
246. Kuru, E.; Määttälä, R.-M.; Noguera, K.; Stork, D. A.; Narasimhan, K.; Rittichier, J.; Wiegand, D.; Church, G. M. Release factor inhibiting antimicrobial peptides improve nonstandard amino acid incorporation in wild-type bacterial cells. *ACS Chem. Biol.* **2020**, *15* (7), 1852-1861. DOI: 10.1021/acschembio.0c00055

247. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776-779. DOI: 10.1126/science.7973629
248. Yan, L. Z.; Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **2001**, *123* (4), 526-533. DOI: 10.1021/ja003265m
249. Siman, P.; Karthikeyan, S. V.; Brik, A. Native chemical ligation at glutamine. *Org. Lett.* **2012**, *14* (6), 1520-1523. DOI: 10.1021/ol300254y
250. Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J. Chemoselective peptide ligation-desulfurization at aspartate. *Angew. Chem. Int. Ed.* **2013**, *52* (37), 9723-9727. DOI: 10.1002/anie.201304793
251. Crich, D.; Banerjee, A. Native chemical ligation at phenylalanine. *J. Am. Chem. Soc.* **2007**, *129* (33), 10064-10065. DOI: 10.1021/ja072804l
252. Haase, C.; Rohde, H.; Seitz, O. Native chemical ligation at valine. *Angew. Chem. Int. Ed.* **2008**, *47* (36), 6807-6810. DOI: 10.1002/anie.200801590
253. Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Dual native chemical ligation at lysine. *J. Am. Chem. Soc.* **2009**, *131* (38), 13592-13593. DOI: 10.1021/ja905491p
254. Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J. A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. *Tetrahedron* **2010**, *66* (13), 2277-2283. DOI: 10.1016/j.tet.2010.01.067
255. Harpaz, Z.; Siman, P.; Kumar, K. S. A.; Brik, A. Protein synthesis assisted by native chemical ligation at leucine. *ChemBioChem* **2010**, *11* (9), 1232-1235. DOI: 10.1002/cbic.201000168
256. Malins, L. R.; Cergol, K. M.; Payne, R. J. Peptide ligation-desulfurization chemistry at arginine. *ChemBioChem* **2013**, *14* (5), 559-563. DOI: 10.1002/cbic.201300049
257. Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149-2154. DOI: 10.1021/ja00897a025
258. Jensen, K. J. Solid-phase peptide synthesis: an introduction. In *Peptide Synthesis and Applications*, Jensen, K. J., Tofteng Shelton, P., Pedersen, S. L. Eds.; Methods in Molecular Biology, Vol. 1047; Humana Press, 2013; pp 1-21.
259. Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. Methods and protocols of modern solid phase peptide synthesis. *Mol. Biotechnol.* **2006**, *33* (3), 239-254. DOI: 10.1385/MB:33:3:239
260. Shelton, P. T.; Jensen, K. J. Linkers, resins, and general procedures for solid-phase peptide synthesis. In *Peptide Synthesis and Applications*, Jensen, K. J., Tofteng Shelton, P., Pedersen, S. L. Eds.; Methods in Molecular Biology, Humana Press, 2013; pp 23-41.
261. Nadal, S.; Raj, R.; Mohammed, S.; Davis, B. G. Synthetic post-translational modification of histones. *Curr. Opin. Chem. Biol.* **2018**, *45*, 35-47. DOI: 10.1016/j.cbpa.2018.02.004
262. Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (12), 6705-6710. DOI: 10.1073/pnas.95.12.6705
263. Eryilmaz, E.; Shah, N. H.; Muir, T. W.; Cowburn, D. Structural and dynamical features of inteins and implications on protein splicing. *J. Biol. Chem.* **2014**, *289* (21), 14506-14511. DOI: 10.1074/jbc.R113.540302
264. Shah, N. H.; Muir, T. W. Inteins: Nature's gift to protein chemists. *Chem. Sci.* **2014**, *5* (1), 446-461. DOI: 10.1039/C3SC52951G
265. Flavell, R. R.; Muir, T. W. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **2009**, *42* (1), 107-116. DOI: 10.1021/ar800129c
266. He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.; Meng, F.; Kelleher, N.; McCafferty, D. G. Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (21), 12033-12038. DOI: 10.1073/pnas.2035256100
267. Loyola, A.; LeRoy, G.; Wang, Y.-H.; Reinberg, D. Reconstitution of recombinant chromatin establishes a requirement for histone-tail modifications during chromatin assembly and transcription. *Genes Dev.* **2001**, *15* (21), 2837-2851. DOI: 10.1101/gad.937401
268. Poulard, C.; Noureddine, L. M.; Pruvost, L.; Le Romancer, M. Structure, activity, and function of the protein lysine methyltransferase G9a. *Life* **2021**, *11* (10), 1082. DOI: 10.3390/life11101082

269. Ferreira, H.; Flaus, A.; Owen-Hughes, T. Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. *J. Mol. Biol.* **2007**, *374* (3), 563-579. DOI: 10.1016/j.jmb.2007.09.059
270. Manohar, M.; Mooney, A. M.; North, J. A.; Nakkula, R. J.; Picking, J. W.; Edon, A.; Fishel, R.; Poirier, M. G.; Ottesen, J. J. Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. *J. Biol. Chem.* **2009**, *284* (35), 23312-23321. DOI: 10.1074/jbc.M109.003202
271. 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. *J. Mol. Biol.* **2011**, *408* (2), 187-204. DOI: 10.1016/j.jmb.2011.01.003
272. Li, G.; Widom, J. Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.* **2004**, *11* (8), 763-769. DOI: 10.1038/nsmb801
273. Cheal, S. M.; Ng, M.; Barrios, B.; Miao, Z.; Kalani, A. K.; Meares, C. F. Mapping protein-protein interactions by localized oxidation: consequences of the reach of hydroxyl radical. *Biochemistry* **2009**, *48* (21), 4577-4586. DOI: 10.1021/bi900273j
274. Koshi, Y.; Nakata, E.; Miyagawa, M.; Tsukiji, S.; Ogawa, T.; Hamachi, I. Target-specific chemical acylation of lectins by ligand-tethered DMAP catalysts. *J. Am. Chem. Soc.* **2008**, *130* (1), 245-251. DOI: 10.1021/ja075684q
275. Amamoto, Y.; Aoi, Y.; Nagashima, N.; Suto, H.; Yoshidome, D.; Arimura, Y.; Osakabe, A.; Kato, D.; Kurumizaka, H.; Kawashima, S. A.; et al. Synthetic posttranslational modifications: chemical catalyst-driven regioselective histone acylation of native chromatin. *J. Am. Chem. Soc.* **2017**, *139* (22), 7568-7576. DOI: 10.1021/jacs.7b02138
276. Barbera, A. J.; Chodaparambil, J. V.; Kelley-Clarke, B.; Joukov, V.; Walter, J. C.; Luger, K.; Kaye, K. M. The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **2006**, *311* (5762), 856-861. DOI: 10.1126/science.1120541
277. Fujiwara, Y.; Yamanashi, Y.; Fujimura, A.; Sato, Y.; Kujirai, T.; Kurumizaka, H.; Kimura, H.; Yamatsugu, K.; Kawashima, S. A.; Kanai, M. Live-cell epigenome manipulation by synthetic histone acetylation catalyst system. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (4), e2019554118. DOI: 10.1073/pnas.2019554118
278. Hamajima, W.; Fujimura, A.; Fujiwara, Y.; Yamatsugu, K.; Kawashima, S. A.; Kanai, M. Site-selective synthetic acylation of a target protein in living cells promoted by a chemical catalyst/donor system. *ACS Chem. Biol.* **2019**, *14* (6), 1102-1109. DOI: 10.1021/acschembio.9b00102
279. Afonso, C. F.; Marques, M. C.; António, J. P. M.; Cordeiro, C.; Gois, P. M. P.; Cal, P. M. S. D.; Bernardes, G. J. L. Cysteine-assisted click-chemistry for proximity-driven, site-specific acetylation of histones. *Angew. Chem. Int. Ed.* **2022**, *61* (46), e202208543. DOI: 10.1002/anie.202208543
280. Guo, S.; Ma, L.; Zhao, J.; Küçüköz, B.; Karatay, A.; Hayvali, M.; Yaglioglu, H. G.; Elmali, A. BODIPY triads triplet photosensitizers enhanced with intramolecular resonance energy transfer (RET): broadband visible light absorption and application in photooxidation. *Chem. Sci.* **2014**, *5* (2), 489-500. DOI: 10.1039/C3SC52323C
281. Aghapour, G.; Afzali, A. Facile conversion of aldehydes and ketones to gem-dichlorides using chlorodiphenylphosphine/N-chlorosuccinimide as a new and neutral system. *Synth. Commun.* **2008**, *38* (22), 4023-4035. DOI: 10.1080/00397910802271271
282. Zhao, Y.; Kang, J.; Park, C.-M.; Bagdon, P. E.; Peng, B.; Xian, M. Thiol-activated gem-dithiols: a new class of controllable hydrogen sulfide donors. *Org. Lett.* **2014**, *16* (17), 4536-4539. DOI: 10.1021/ol502088m
283. Klinker, H.; Haas, C.; Harrer, N.; Becker, P. B.; Mueller-Planitz, F. Rapid purification of recombinant histones. *PLoS One* **2014**, *9* (8), e104029. DOI: 10.1371/journal.pone.0104029
284. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3784-3788. DOI: 10.1093/nar/gkg563
285. Lee, B.; Sun, S.; Jiménez-Moreno, E.; Neves, A. A.; Bernardes, G. J. L. Site-selective installation of an electrophilic handle on proteins for bioconjugation. *Bioorg. Med. Chem.* **2018**, *26* (11), 3060-3064. DOI: 10.1016/j.bmc.2018.02.028
286. Kelly, S. M.; Jess, T. J.; Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **2005**, *1751* (2), 119-139. DOI: 10.1016/j.bbapap.2005.06.005

287. Pentelute, B. L.; Kent, S. B. H. Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation. *Org. Lett.* **2007**, *9* (4), 687-690. DOI: 10.1021/ol0630144
288. Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem. Int. Ed.* **2007**, *46* (48), 9248-9252. DOI: 10.1002/anie.200704195
289. Gao, X.-F.; Du, J.-J.; Liu, Z.; Guo, J. Visible-light-induced specific desulfurization of cysteinyl peptide and glycopeptide in aqueous solution. *Org. Lett.* **2016**, *18* (5), 1166-1169. DOI: 10.1021/acs.orglett.6b00292
290. Lee, M.; Neukirchen, S.; Cabrele, C.; Reiser, O. Visible-light photoredox-catalyzed desulfurization of thiol- and disulfide-containing amino acids and small peptides. *J. Pept. Sci.* **2017**, *23* (7-8), 556-562. DOI: 10.1002/psc.3016
291. Jin, K.; Li, T.; Chow, H. Y.; Liu, H.; Li, X. P-B desulfurization: an enabling method for protein chemical synthesis and site-specific deuteration. *Angew. Chem. Int. Ed.* **2017**, *56* (46), 14607-14611. DOI: 10.1002/anie.201709097
292. Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007**, *128* (5), 1003-1012. DOI: 10.1016/j.cell.2006.12.041
293. Attar, N.; Campos, O. A.; Vogelauer, M.; Cheng, C.; Xue, Y.; Schmollinger, S.; Salwinski, L.; Mallipeddi, N. V.; Boone, B. A.; Yen, L.; et al. The histone H3-H4 tetramer is a copper reductase enzyme. *Science* **2020**, *369* (6499), 59-64. DOI: 10.1126/science.aba8740
294. Luger, K.; Rechsteiner, T. J.; Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. In *Chromatin Protocols*, Becker, P. B. Ed.; Methods in Molecular Biology, Vol. 119; Humana Press, 1999; pp 1-16.
295. Kapust, R. B.; Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **1999**, *8* (8), 1668-1674. DOI: 10.1110/ps.8.8.1668
296. Duong-Ly, K. C.; Gabelli, S. B. Chapter Two - Affinity purification of a recombinant protein expressed as a fusion with the maltose-binding protein (MBP) tag. In *Laboratory Methods in Enzymology: Protein Part D*, Lorsch, J. R. Ed.; Methods in Enzymology, Vol. 559; Academic Press, 2015; pp 17-26.
297. Raran-Kurussi, S.; Keefe, K.; Waugh, D. S. Positional effects of fusion partners on the yield and solubility of MBP fusion proteins. *Protein Expr. Purif.* **2015**, *110*, 159-164. DOI: 10.1016/j.pep.2015.03.004
298. Subirana, J. A. Specific aggregation products of histone fractions (presence of cysteine in F2a1 from echinoderms). *FEBS Lett.* **1971**, *16* (2), 133-136. DOI: 10.1016/0014-5793(71)80351-4
299. Gelbart, M. E.; Rechsteiner, T.; Richmond, T. J.; Tsukiyama, T. Interactions of Isw2 chromatin remodeling complex with nucleosomal arrays: analyses using recombinant yeast histones and immobilized templates. *Mol. Cell. Biol.* **2001**, *21* (6), 2098-2106. DOI: 10.1128/MCB.21.6.2098-2106.2001
300. Getz, E. B.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* **1999**, *273* (1), 73-80. DOI: 10.1006/abio.1999.4203
301. Baba, S. P.; Bhatnagar, A. Role of thiols in oxidative stress. *Curr. Opin. Toxicol.* **2018**, *7*, 133-139. DOI: 10.1016/j.cotox.2018.03.005
302. Oregioni, A.; Stieglitz, B.; Kelly, G.; Rittinger, K.; Frenkiel, T. Determination of the pK_a of the N-terminal amino group of ubiquitin by NMR. *Sci. Rep.* **2017**, *7*, 43748. DOI: 10.1038/srep43748
303. Li, Z.; Kono, H. Distinct roles of histone H3 and H2A tails in nucleosome stability. *Sci. Rep.* **2016**, *6*, 31437. DOI: 10.1038/srep31437
304. Izumi, Y.; Matsuo, K.; Fujii, K.; Yokoya, A.; Taniguchi, M.; Namatame, H. Circular dichroism spectroscopic study on structural alterations of histones induced by post-translational modifications in DNA damage responses: lysine-9 methylation of H3. *J. Radiat. Res.* **2018**, *59* (2), 108-115. DOI: 10.1093/jrr/rrx068
305. Miles, A. J.; Janes, R. W.; Wallace, B. A. Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. *Chem. Soc. Rev.* **2021**, *50* (15), 8400-8413. DOI: 10.1039/d0cs00558d

306. Kelly, S. M.; Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **2000**, *1* (4), 349-384. DOI: 10.2174/1389203003381315
307. Gil, R.; Barth, S.; Kanfi, Y.; Cohen, H. Y. SIRT6 exhibits nucleosome-dependent deacetylase activity. *Nucleic Acids Res.* **2013**, *41* (18), 8537-8545. DOI: 10.1093/nar/gkt642
308. Vamisetti, G. B.; Satish, G.; Sulkshane, P.; Mann, G.; Glickman, M. H.; Brik, A. On-demand detachment of succinimides on cysteine to facilitate (semi)synthesis of challenging proteins. *J. Am. Chem. Soc.* **2020**, *142* (46), 19558-19569. DOI: 10.1021/jacs.0c07663
309. Zhang, Y.; Zhou, X.; Xie, Y.; Greenberg, M. M.; Xi, Z.; Zhou, C. Thiol specific and tracelessly removable bioconjugation via Michael addition to 5-methylene pyrrolones. *J. Am. Chem. Soc.* **2017**, *139* (17), 6146-6151. DOI: 10.1021/jacs.7b00670
310. Khan, K. A.; Ng, M. K.; Cheung, P. The use of mononucleosome immunoprecipitation for analysis of combinatorial histone post-translational modifications and purification of nucleosome-interacting proteins. *Front. Cell Dev. Biol.* **2020**, *8*, 331. DOI: 10.3389/fcell.2020.00331