Development of chemical biology tools to reveal the function and organization of the silenced chromatin state

Présentée le 4 septembre 2020

à la Faculté des sciences de base
Laboratoire de chimie biophysique des macromolécules
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Nora GUIDOTTI

Acceptée sur proposition du jury

Prof. S. Gerber, présidente du jury
Prof. B. Fierz, Prof. V. Hatzimanikatis, directeurs de thèse
Prof. C. Hackenberger, rapporteur
Dr L. Walport, rapporteuse
Prof. P. Rivera Fuentes, rapporteur

École
polytechnique
fédérale
de Lausanne

2020

Abstract

Effector proteins are recruited to chromatin *via* transient interactions between their specialized reader domains and histone post-translational modification (PTM) patterns. These interactions form signaling pathways that control gene expression or repression and which can lead to cancer when dysregulated. Moreover, combinations of histone PTMs define particular chromatin states and determine 3-dimensional genome organization. Chromatin represents a key intersection point in the complex signaling network of a eukaryotic cell and the mechanisms underlying the readout of PTMs are of great interest, but still not well understood.

Recent studies revealed that histone PTMs can be installed in an asymmetrical configuration, that is, the two copies of histone proteins within a nucleosome do not obligatorily carry the same PTMs. Such asymmetry greatly expands the combinatorial space of histone marks, yet little is known about their biological relevance. In addition, dependent on histone PTM patterns, particular chromatin states - such as the repressive heterochromatin - compartmentalize into membraneless condensates *via* liquid-liquid phase separation (LLPS). Although it is not clear whether LLPS is a general organizational principle, it certainly constitutes a new paradigm in understanding critical chromatin-dependent mechanisms and diseases.

Chemical biology and biophysics are key for the investigation of chromatin organization and to monitor chromatin-effector interactions in time and space. First, for *in vitro* studies of the molecular processes encrypted in the histone code, the facile chemical access to pure modified nucleosomes is crucial. However, their preparation is not straightforward. Second, investigations of chromatin dynamics on the molecular scale depend on the ability to site-specifically label protein effectors and chromatin. Thus, new and versatile protein modification methods together with bright and stable dyes are of key importance. Finally, cell permeable probes to investigate the *in vivo* dynamic behaviour and localization of the genome in live cells are currently lacking.

This PhD project approaches these multiple challenges. First, I extended a recently developed crosslinking strategy, which gives control over the supramolecular assembly of nucleosomes, to the synthesis of nucleosomes with asymmetrical PTM patterns on histone H3 and H4. These precious substrates revealed mechanistic details of stem-cell regulation by key methyltransferases, PRC2 and Set8. Second, I report the use of hypervalent iodine reagents for the effective cysteine-selective labeling of proteins, such as histone octamers. The strategy enables to doubly functionalize the substrates *via* formation of an azide-containing vinylbenziodoxolone intermediate. Using this method, a triplet state quencher was introduced in proximity to a fluorophore, reducing the photobleaching rates and enhancing the fluorescence lifetime in single-molecule studies. Finally, I describe the design of peptide-based LLPS probes for live cell chromatin

visualization. Conceived to stain heterochromatin foci, I expect the probe to be easily tunable to other chromatin states, enabling the 3D dynamic detection of chromatin organization in living cells.

Together, this thesis contributes to the deciphering of the multifaced chromatin PTM readout and provides the chemical biology and biophysics community with novel tools that will allow to further investigate chromatin-dependent processes.

Riassunto

Le proteine vengono reclutate dalla cromatina tramite interazioni transitorie che si stabiliscono tra i loro domini specializzati e le moficazioni post-traduzionali (PTM) degli istoni. Queste interazioni formano segnali cellulari che controllano l'espressione genica e, se alterate, possono provocare il cancro. Inoltre, combinazioni di PTMs definiscono differenti stati della cromatina e determinano l'organizzazione tridimensionale del genoma. La cromatina è quindi un punto d'intersezione chiave nella complicata rete dei segnali cellulari. Nonostante i meccanismi che descrivono l'interpretazione di queste PTMs generino molto interesse, non sono stati ancora completamente compresi.

Studi recenti hanno dimostrato che queste PTMs possono essere asimmetriche. In un singolo nucleosoma, le due copie di istoni non sempre presentano le stesse PTMs. Ciononostante, la rilevanza biologica si tale asymmetria è incerta. Inoltre, alcuni stati della cromatina - come l'eterocromatina - possono subire un processo chiamato "separazione di fase liquido-liquido" (LLPS) e formare degli agglomerati di molecole privi di membrana. Non è chiaro se l'LLPS sia un principio generale, tuttavia, rappresenta un nuovo punto di vista dal quale partire per elucidare i meccanismi e le malattie derivanti dai processi dipendenti dalla cromatina.

Innanzitutto, per capire i processi molecolari che sono determinati dal cosiddetto "codice degli istoni", è essenziale avere a disposizione metodi che permettano di ottenere nucleosomi modificati in modo specifico. Tuttavia, la sintesi di tali substrati non è banale. In secondo luogo, per poter investigare a livello molecolare le dinamiche della cromatina, è importante disporre di tecniche per l'aggiunta specifica di tag sia sulle proteine effetrici, che sulla cromatina. Quindi, metodi nuovi e versatili per modificare le proteine, insieme a fluorofori brillanti e fotostabili, sono di fondamentale importanza. Infine, non esistono sonde capaci di permeare nelle cellule viventi e che permettano di monitorare il comportamento dinamico e la localizzazione del genoma all'interno del nucleo.

Recentemente, è stata sviluppata una strategia di crosslinking per controllare l'assemblaggio supramolecolare dei nucleosomi. In questo lavoro, ho esteso questa tecnica applicandola alla sintesi di nucleosomi con PTMs asimmetriche sugli istoni H3 e H4. Questi preziosi substrati hanno rivelato dettagli meccanicistici della regolazione delle cellule staminali da parte delle metiltransferasi PRC2 e Set8. Secondariamente, descrivo l'uso di reagenti iodati ipervalenti per la modificazione efficace di proteine tramite le cisteine. La strategia consente di aggiungere due gruppi funzionali ai substrati attraverso la formazione di un intermedio benziodoxolone vinilico contenente un'azide. Grazie a questo metodo, un agente stabilizzante è stato introdotto in prossimità di un fluoroforo, riducendo così il tasso di fotobleaching ed estendendo la durata del segnale fluorescente. Infine, espongo il design di sonde

peptidiche per la visualizzazione della cromatina nelle cellule viventi. Anche se concepite per marcare gli agglomerati di eterocromatina, mi aspetto che queste sonde possano essere facilmente adattate al riconoscimento di altri stati della cromatina.

In conclusione, questa tesi contribuisce alla decodificazione del codice degli istoni e fornisce alla comunità scientifica nuovi strumenti che consentiranno di approfondire i processi dipendenti dalla cromatina.

Parole chiave: Epigenetica, modificazioni post-traduzionali degli istoni, eterocromatina, nucleosomi asimmetrici, modificazioni specifiche della cisteina, coniugazione bio-ortogonale, separazione di fase liquido-liquido.

Acknowledgements

I am very grateful to my thesis director Prof. Beat Fierz for giving me the opportunity to join LCBM as a PhD student after my master thesis. Thank you for always being supportive, available for discussion and an inexhaustible source of new ideas. Your creativity and passion for science has been truly inspiring. I would also like to thank Prof. Vassily Hatzimanikatis for being an esteemed thesis co-director.

I thank the committee members Prof. Hackenberger, Dr. Walport and Prof. Rivera-Fuentes for taking the time to review this thesis and attend my private defense. I also thank Prof. Gerber for being the president of the jury. I would like to thank Anne Lene Odegaard for helping me deal with doctoral school rules and requirements, Yoan Dind for rescuing my laptop many times and Marie Jirousek and Jacques Gremaud for their dedication and friendly professionalism.

Heartfelt thanks go to our secretary Marie Munoz for her positive attitude, patience and support. Your warm smile has been extremely helpful both in personal and work-related difficult situations. I wish to thank all former and current LCBM members for the joyful atmosphere they have created in the lab. The daily discussions and funny moments we spent together made me grow a lot both as a person and as a scientist. Thanks to Aurore for being an admirable master thesis supervisor and for introducing me to the lab life with her honest and generous attitude. I wholeheartedly thank Carolin for sharing her knowledge and passion for peptide chemistry with me. I would like to thank our "mommy" Iuliia for being a valuable reference in the lab and for her sweet and fair nature. I really enjoyed our long discussions about science, life and cleaning (!). Many thanks go to Rebeca for having been a dynamic and receptive student with tireless enthusiasm for our project and little Nacho. I really would like to thank Eduard for his contagious Dutch humor and for being a caring friend. Thanks for taking the time to listen to my scientific and non-scientific thoughts with no complaints (!). I also thank Maxime for his precious advice on any imaginable topics. I am sure that with your unique curiosity, enthusiasm and optimism you will achieve great things. To Eduard and Maxime also go many thanks for reading and reviewing part of the-the-the thesis. Thanks to Ruud for being the most altruistic person I have ever met. His funny originality always put me in a good mood, even when our spoiled confocal microscope was not behaving properly. I would like to thank Luc for his quiet but explosive presence and for keeping me company while waiting for peaks at the HPLC. I am grateful to my bench neighbor Harsh for his infinite patience when his pipettes were mysteriously disappearing. I thank Kristina for her diligence and the efforts she puts into everything she does, especially for making (hi)*story* of LCBM at every lab event. I particularly appreciated Anne's expertise with PowerPoint when she helped me print a last-minute poster. To the new PhD students and office mates Alexandra, Pauline and Sim and to the new entries Neige and Tim, I wish that they enjoy their time in the lab as much as I did. In particular, I am happy to pass the COSEC crown to Alex, as I know that with her

organized and resolute personality she will do a great job. I really thank Pauline for bringing cookies and yummy cakes to the office, and Sim for sharing with me the love for losing keys. I am grateful to Neige for amplifying my plasmid while I was busy with this thesis and to Tim for the nice chats and discussions we had in front of the peptide synthesizer. Finally, I also thank Sinan, Andreas, Ninad, Karthik, Louise, Horst, Maeva and Jade for the kind help and advice they provided whenever I had doubts or needed assistance.

Many thanks certainly go to the "Batochime people", including Vanessa, Tamara, Patrick, Davide, Ale, Sevan, Yann and Carlito among many others, for adopting me on the other side of the road. I spent with you some of the most memorable times. Thanks for always being there whenever I needed to take a break from my little nucleosomes. In particular, I would like to thank Vanessa and Tamara for our true and sporty friendship and for cheering me up during hard times. I also want to thank Vanessa and Davide for helping me reduce the "Ticino dialect" I initially put into my Italian abstract.

Importantly, I would like to say thank you to the people (and poodle!) who were waiting for me back at home almost every weekend, making me feel loved and understood. First of all, to my best friend and sister So', for her big heart and for always standing by my side. It is great to know I can always count on you, no matter what. Thanks for showing a sincere interest in my research and after hours of explanation, concluding that I am playing with "fluorescence and DNA and proteins, right??". Not so wrong after all. Many thanks also go to Marika for her contagious laughter and for the fun we had while saving the world with our top-secret missions. I have no words to say how grateful I am to my parents, for always believing in me and giving me a hand, whether I needed it or not. You have been simply essential to me and, together with Sofi and Illy, we make a wonderful family and a great fighting team. Last but not least, I would like to thank Fosco for his unconditional patience and for sharing with me the best adventures, trips around the world and his life.

List of Figures

List of Tables

Abbreviations

(^{3}H-)SAM	(Tritium-labeled) S-adenosylmethionine
(Ch)IP	(Chromatin) immunoprecipitation
lnc-tag	link-and-cut tag
(t)/(m)/(snc)/(lnc)RNA	(Transfer)/(messenger)/(short coding)/(long coding) ribonucleic acid
3D	Three-dimensional
53BP1	p53-binding protein 1
5cac	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Ac	Acetylation
Acm	Acetamidomethyl
ADP	Adenosine diphosphate
AgOAc	Silver acetate
Alloc	Allyloxycarbonyl
ATP	Adenosine triphosphate
BD	Bromodomain
BET	Bromo- and extra-terminal domain
Boc	tert-Butyloxycarbonyl protecting group
bp	base pair
Brd4	Bromodomain containing 4
Cas9	CRISPR-associated protein 9
Cbz	Carboxybenzyl group
CD	Chromodomain
CGI	CpG island
ChIP-seq	ChIP-sequencing
cMAP	Chromatin-sensing multivalent probe
COMU	1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-Carbenium hexafluorophosphate
CPP	Cell penetrating peptide
CRISPR	Clustered regularly interspaced short palindromic repeats
CSD	Chromoshadow domain
CTE	C-terminal extension
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition
DBCO	Dibenzocyclooctynes
Dbz	Diaminobenzoic acid
DCC	N,N'-Dicyclohexylcarbodiimide
DDR	DNA damage response

Dha	Dehydroalanine
DIC	N,N′-Diisopropylcarbodiimide
DNA	Deoxyribonucleic acid
DSB	Double strand break
DTNB	5,5′-Dithiobis(2-nitrobenzoic acid)
DTT	Dithiothreitol
EBX	Ethynylbenziodoxolone
EED	Embryonic ectoderm development
EM	Electron microscopy
EPL	Expressed protein ligation
ESC	Embryonic stem cell
ESI-MS	Electron spray ionization mass spectrometry
EZH2	Enhancer of zeste homolog 2
FLIM	Fluorescence lifetime imaging microscopy
Fmoc	Fluorenylmethoxycarbonyl
FRAP	Fluorescence recovery after photobleaching
GFP	Green fluorescent protein
GmdCl	Guanidinium hydrochloride
GSH	Glutathione
HAT	Histone acetyltransferase
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HCTU	O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEK	Human embryonic kidney
HF	Hydrofluoric acid
Hi-C	High-throughput chromosome conformation capture
His6	Hexahistidine
HMT	Histone methyltransferase
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	Hydroxybenzotriazole
HP1	Heterochromatin protein 1
HSA	Human serum albumin
ICeChIP	Internal Standard Calibrated ChIP
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IPL	Intein-mediated protein ligation
K*cr	Crotonyl lysine
Kac	Lysine acetylation
KCL	Kinetically controlled ligation

K_D	Dissociation constant
KDM	Lysine demethylase
Kme1	Lysine monomethylation
Kme2	Lysine dimethylation
Kme3	Lysine trimethylation
Kub	Lysine ubiquitylation
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LLPS	Liquid-liquid phase separation
MBT	Malignant brain tumor
Me	Methylation
MESNa	Mercaptoethanethiolate sodium salt
MLL1	Mixed-lineage leukemia 1
MPAA	Mercaptophenylacetic acid
MST	Microscale thermophoresis
MTG	Methylthio glycolate
NChIP	Native ChIP
NCL	Native chemical ligation
NK1	Neurokinin 1
NLS	Nuclear localization sequence
NMR	Nuclear magnetic resonance
NTE	N-terminal extension
pBPA	p-benzoyl-L-phenylalanine
PcG	Polycomb group
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PhSeCys	Phenylselenocysteine
PML	Promyelocytic leukaemia
PNBK*	p-nitrobenzyl-caged lysine
POI	Protein of interest
PPI	Protein-protein interaction
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PTM	Post-translational modification
Re-ChIP	Sequential ChIP
Rme	Arginine methylation
RNA pol II	RNA polymerase II
RNAi	RNA interference
RP-HPLC	Reverse phase high performance liquid chromatography
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEA	Bis(2-sulfanylethyl)amido
SeEA	Bis(2-selenylethyl)amido
SeL	Diselenolane

SiR(-mal)	Silicon-rhodamine maleimide
smTIRFM	Single-molecule total internal reflection fluorescence microscopy
SN2	Nucleophilic substitution type 2
SP	Substance P
SPAAC	Strain promoted azide-alkyne cycloaddition
Sph	Serine phosphorylation
SPPS	Solid phase peptide synthesis
TAD	Topological associated domain
TAF3	TATA-Box
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TF	Transcription factor
TFA	Trifluoroacetic acid
TFET	Trifluoroethanedithiol
Thz	Thiazolidine
Trolox	6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
Trt	Trityl
TSQ	Triplet-state quencher
TSS	Transcription start site
Uaa	Unnatural amino acid
VA-044	2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride
VBX	Vinylbenziodoxolone
wt	Wild type
XChIP	Cross-linking ChIP

Table of Contents

Chapter 1: Introduction

1.1 Chromatin organization and post-translational modifications

1.1.1 DNA organization in eukaryotes

The human genome contains approximately 6 billion base pairs (bp) packaged into a cell nucleus with an average diameter of 5-10 µm. Since a base pair spans around 0.34 nm, each human nucleus contains about 2 meters of DNA.[1] How can this long polymer fit into such a microscopic compartment? The answer lies in its tight folding and structured packaging. In addition, despite this high level of compaction, the DNA must be continuously accessed and interpreted in a highly synchronized manner to allow the cells to accomplish defined functions and to respond to environmental changes. In eukaryotic cells this is achieved by organizing the genome in a dynamic polymeric complex called chromatin.
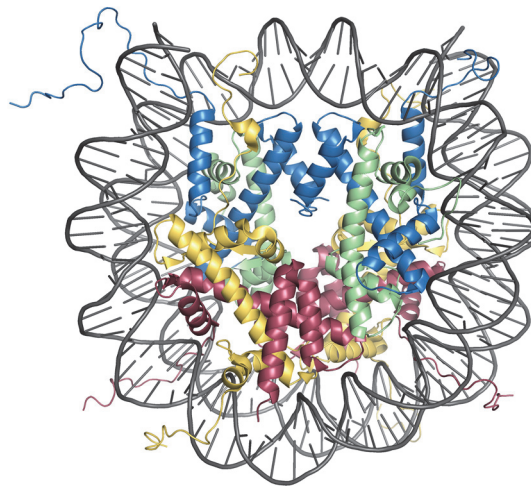
Figure 1: Crystal structure of nucleosome core particle. H2A: yellow. H2B: red. H3: blue. H4: green. DNA: grey. (PDB: 1EQZ*).*

The nucleosome is the fundamental building block of the chromatin structure and the first high resolution crystal structure was obtained by Richmond et al. in 1997.[2] It is composed of 147 bp DNA wrapped around a core histone octamer. The histone octamer consists of two copies of each core histone (H2A, H2B, H3, H4 or their variants, i.e. H2A.X), more specifically two H2A-H2B dimers and one H3-H4 tetramer, arranged in a pseudo twofold symmetry particle (**Figure 1**).[3] Far from being a rigid and static structure, the nucleosome is subjected to a variety of transient and reversible fluctuations that temporarily expose DNA binding sites to transcription factors (TFs) and are strongly dependent on DNA sequence and histone variants.[4,5,6,7,8,9,10] In addition to these spontaneous DNA movements, nucleosomes can also be repositioned along the genome by chromatin remodeling complexes. These large complexes use the energy of ATP hydrolysis to create a force to slide and/or evict nucleosomes or to exchange histone variants, thus altering the chromatin

structure.[11] Therefore, nucleosomes are highly complex particles subjected to a multitude of dynamic processes, which in turn tightly regulate the exposure of DNA TFs binding sites. Thereby, nucleosomes play an important role in modulating the thermodynamics and kinetics of the effector-DNA interactions.

However, nucleosomes only constitute the first level of the chromatin compaction (**Figure 2**), as they are connected by a 20 to 80 bp linker DNA. *In vitro* studies show that, at low ionic strengths, the self-repulsion of DNA negative charges from the phosphodiester backbone forces the chromatin to assume a "beads-on-a-string" conformation.[12,13] On the other hand, under physiological salt conditions, internucleosomal contacts cause the formation of higher-order structures, such as 30-nm chromatin fibers. Moreover, an additional histone protein, the linker histone H1, can associate with nucleosome core particles and promote further compaction.[14] The structure of the 30-nm fibers is currently under debate and different conformational models have been proposed. The two-start model, where contacts between next-neighbor nucleosomes (i, i+2) form a superhelix with the linker DNA zig-zagging between the two stacks;[15,16,17] and the solenoid model, with neighboring nucleosome directly interacting (i, i+1).[18,13] Although different *in vitro* studies support either of the models, Routh et al. demonstrated that the DNA linker length and the presence of H1 greatly influence the predisposition to form one or the other structure.[19] These *in vitro* studies may not be representative of the dynamic chromatin structure in living cells. In fact, many further studies highlighted the presence of two-start-like contacts in interphase chromatin and particularly at the compacted heterochromatin regions,[20,21,22,23] however, other recent *in vivo* investigations using 3D high resolution EM[24] or fluorescence super-resolution imaging,[25] are consistent with a highly irregular DNA folding and thus with the absence of the 30-nm fiber in cells.
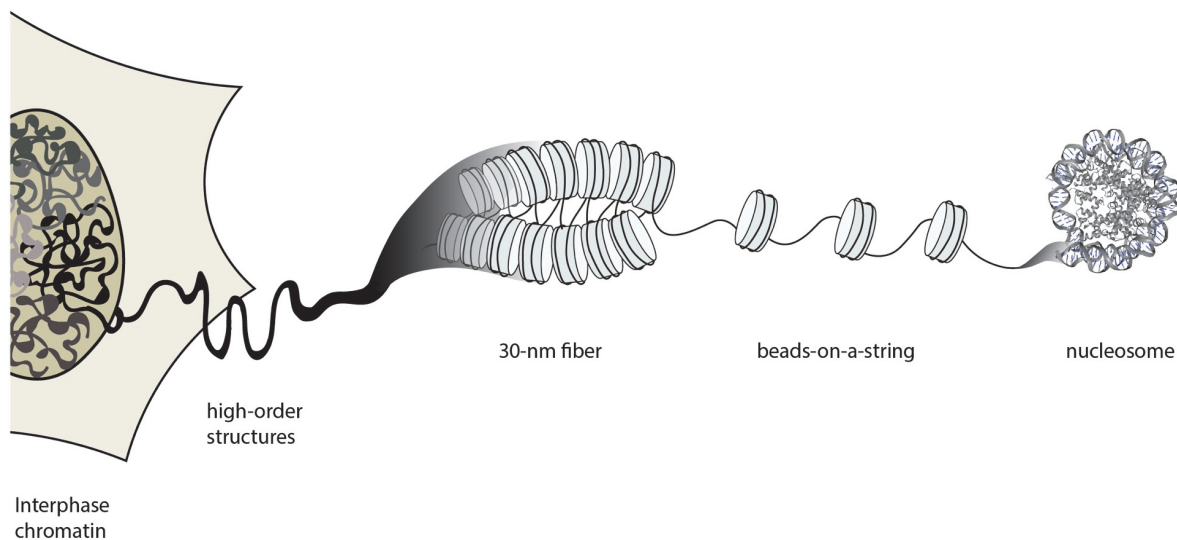
Figure 2: Hierarchy of chromatin folding. Chromatin is organized on multiple levels of compaction inside the nucleus. From left to right: interphase chromatin, high-order structures, the 30-mn fiber, beads-on-a-string conformation, single nucleosome.

Nevertheless, on larger scales, whole-genome chromosome conformation capture (Hi-C) studies[26] suggest that chromatin forms alternating Mbp-long compartments (defined as A- and B-compartments). While A-compartments show transcriptional activity, B-compartments are transcriptionally inactive. Within these

compartments, extensive loops that cluster genes by means of architectural proteins such as cohesin are formed, and the chromosome folds into genomic domains that show strong self-interaction and are insulated from neighboring regions (topologically associated domains, TADs).[27] A growing body of evidence indicates that the driving force of this three-dimensional arrangement is phase separation,[28,27] a physical process that will be described in detail in **Chapter 1.2**. These architectural chromatin units shape the genome landscape, suggesting a fundamental role in chromatin organization and regulation. In addition, chromatin organization varies tremendously along the cell division cycle. In interphase, chromatin is less condensed than during mitosis and no distinct chromosomes can be discriminated. However, the chromosomes occupy distinct territories within the nucleus and their 3D organization depends on long-range intrachromosomal interactions.[29] Maintaining insulated neighborhoods among chromosomes is key for the determination of cell specificity and must be conserved all along the cell cycle, while dramatic chromatin structural dynamic changes occur. Thus, ranging from local nucleosomes flexibility to global scale chromosome genomic domain organization, these processes collaboratively orchestrate gene transcription.

1.1.2   Epigenetics, chromatin states and post-translational modifications

Although the somatic cells of a multicellular organism are genetically identical, they are characterized by heterogeneous morphologies, specific localization within the body and can exert different functions. This is due to their differential gene transcription profiles and consequently to the diverse expression rates of proteins. Many of the variations in gene expression appear upon development and cell proliferation, and are maintained through mitosis. Epigenetic modifications contribute to these heritable cell-to-cell variations. Although none of them alter the primary sequence of the DNA, they have consequences on gene transcription and thus on cell lineage, function and fate.

Epigenetic marks involved in the regulation of gene expression include DNA methylation,[30] non-coding RNA-mediated silencing[31] and histone post-translational modifications (PTMs)[32]; all of which considerably expand the information potential of the genetic code.

- DNA methylation

DNA methylation was the first proposed epigenetic modification[33] and mainly consists in the addition of a methyl group to a cytosine nucleotide (5mC, **Figure 3A**). While this mark suppresses the expression of the downstream gene when located in promoter regions, it is linked to transcriptional activation when found in gene bodies.[34] An example of DNA-methylation controlled gene transcription mechanisms is imprinting,[35,36] where some genes are expressed in a parental-origin-specific manner rather than from both chromosome homologues. Apart from 5mC, other cytosine modifications exist and derive from the successive oxidation of 5mC (**Figure 3A**).[37] These DNA modifications lower 5mC levels and thus contribute to the active cytosine demethylation processes.[38]

3

- Non-coding RNAs

Non-coding RNA-mediated silencing is another powerful mechanism that regulates gene expression and genome stability.[39] These RNA molecules lack protein coding capacity and are subdivided in two major categories: short and long noncoding RNAs (sncRNA and lncRNA, respectively). sncRNA are associated to RNA interference (RNAi) pathways, that in turn induce mRNA degradation and translational repression.[40] On the other hand, lncRNAs are a highly heterogeneous class of noncoding RNA sequences with > 200 nucleotides. They can be transcribed from different regions of the genome (introns, exons, enhancer regions)[41] and control gene expression by different mechanism, including RNA sequestration, RNA transport and chromatin remodeling.[42,43]

- Histone PTMs

Finally, another factor that is of fundamental importance for the epigenetic regulation of gene activation and repression is the wide palette of PTMs that are established on the core histone proteins. Core histone proteins are composed of a central alpha-helical region flanked by N- and C-terminal extensions. These extensions, called histone tails, are mainly unstructured and flexible, and can be subjected to a wide variety of chemical modifications, such as methylation, acetylation, phosphorylation, ubiquitination and ADP-ribosylation (**Figure 3B**).[44] These copious and generally reversible PTMs regulate essential chromatin-dependent processes. Moreover, PTMs normally do not act in isolation but work in a combinatorial manner: they occur concomitantly within the same or neighboring nucleosomes, providing a tight control over the chromatin structure and altering DNA accessibility for effector proteins, such as regulators of gene expression and DNA repair. Biological outcomes are thus specified by combinatorial patterns of histone modifications, a theory also known as the "histone code".[45,46]
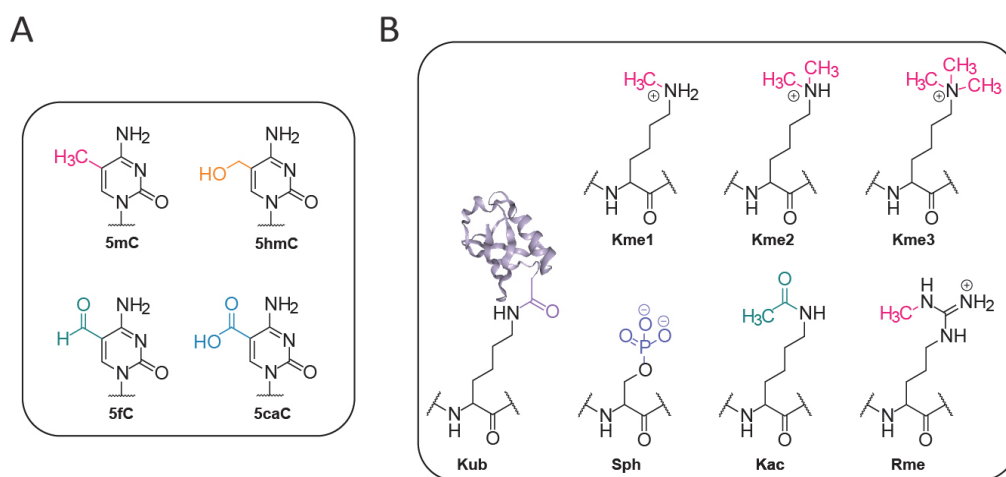
Figure 3: Epigenetic modifications. A) Chemical structure of DNA modifications: 5-methylcytosine (5mC); 5-hydroxymethylcytosine (5hmC); 5-formylcytosine (5fC); 5-carboxylcytosine (5caC). **B**) Chemical structure of example histone PTMs: mono-, di- and tri-methylation of lysine (Kme1, Kme2, Kme3); lysine ubiquitylation (Kub); serine phosphorylation (Sph); lysine acetylation (Kac) and arginine methylation (Rme).

1.1.3   Defined histone post-translational modifications identify specific chromatin states

According to the classical view, eukaryotic genomes can be generally arranged into two different chromatin types. The first one is a decondensed and transcriptionally accessible state, where genes are predominantly transcribed and the histones are acetylated. This is referred to as "euchromatin". The second one, called "heterochromatin", is densely packed, histone acetylation is replaced by repression-associated PTMs and most of the genes are inactive (**Figure 4**). Heterochromatin can further be subdivided into two domains. Facultative heterochromatin - regulated by the Polycomb Group (PcG) protein - contains genes that are differentially expressed through development and/or differentiation. On the other hand, constitutive heterochromatin - maintained by Heterochromatin Protein 1 (HP1) - is essentially devoid of genes and permanently inactive. Of note, recent findings suggest the involvement of liquid-liquid phase separation in the formation of heterochromatin and other chromatin states (see **Chapter 1.2.2**). Both heterochromatin and euchromatin are enriched and depleted of certain characteristic combinations of histone modifications, which define their structural state and regulatory mechanisms (**Figure 4**).[47] For instance, promoters in euchromatin regions are enriched with the trimethylation of lysine 4, which is found on the N-terminal tail of H3 (H3K4me3). Conversely, the trimethylation of lysine 9 on histone H3 (H3K9me3) is a hallmark of the gene-poor constitutive heterochromatin. Further, most of the genes are decorated with the H3 trimethylation of lysine 27 (H3K27me3) in facultative heterochromatin regions.[48]

Euchromatin	Heterochromatin
H3S10ph, H3K4me3, H3K9ac, H3K14ac, H3K36me3, H4K5ac, H4K8ac, H4K12ac, H4K16ac	H2AK119ub, H3K9me3, H3K27me3, H4K20me1
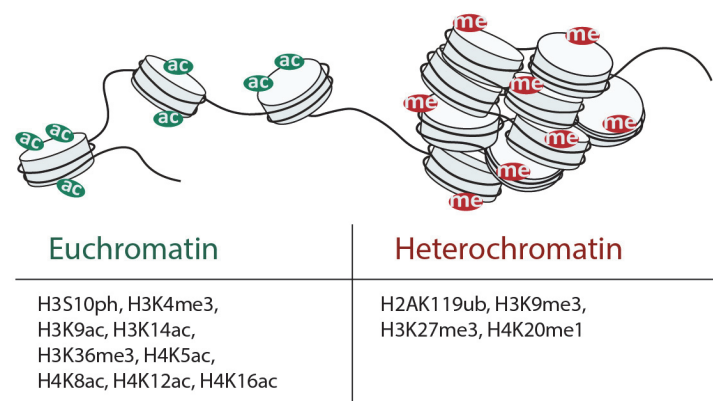
Figure 4: **Euchromatin and heterochromatin are defined by specific PTMs**. Some examples of such PTMs are reported in the table.

The histone PTM code affects gene regulation by means of two main mechanisms, either by directly influencing the overall structure of chromatin (the "direct" model) or by mediating the recruitment of downstream effector molecules (the "effector-mediated" model). Examples of the direct model are phosphorylation and acetylation of core histones that attenuate the favorable DNA-histone and histone-histone interactions. For instance, the positively charged H4 N-terminal tail interacts with the acidic patch region on the surface of neighboring nucleosomes. Acetylation of H4K16 prevents this contact and disrupts higher order chromatin folding.[49] On the other hand, examples of the effector mediated model are H3K4me3[50] and H3 lysine 14 acetylation (H3K14ac)[51] predominantly found on euchromatic regions and

responsible for the recruitment of effector proteins such as RNA polymerase II, transcription factors and remodeling complexes; H3K9me3 which provides a binding site for proteins (such as HP1[52,53]) that maintain the compact state of the constitutive heterochromatin; H3K27me3 which induces silencing of genes in the facultative heterochromatin by means of Polycomb repressive complex 1 and 2 (PCR1 and PCR2) binding.[54]

1.1.4 Chromatin effectors: readers, writers and erasers

Effector proteins decode histone PTMs by means of their specialized protein domains ("reader" domains). Both the modification state and position within the histone sequence is pivotal for the specific recognition of individual or combinations of marks. The first reported interaction of a PTM and a reader domain was between a bromodomain (BDs) and an acetylated lysine.[55] BDs are protein modules found in nuclear histone acetyltransferases (HATs), transcription factors and other regulatory proteins, and they play key roles in transcriptional activation and chromatin remodeling. By contrast, methyllysines interact with other reader modules, such as chromodomains (CD),[56] Plant HomeoDomain (PHD) fingers,[57] WD40 repeats,[58] Malignant Brain Tumor (MBT)[59] and Tudor domains.[60] Lysine methylation is an especially interesting and complex PTM, since this residue can be mono-, di- and tri-methylated and is one of the most stable histone marks. Binders of methylated lysine typically have an aromatic cage, formed by two to four aromatic residues, that engages the quaternary ammonium with a cation - π interaction (**Figure 5**).[61,62]
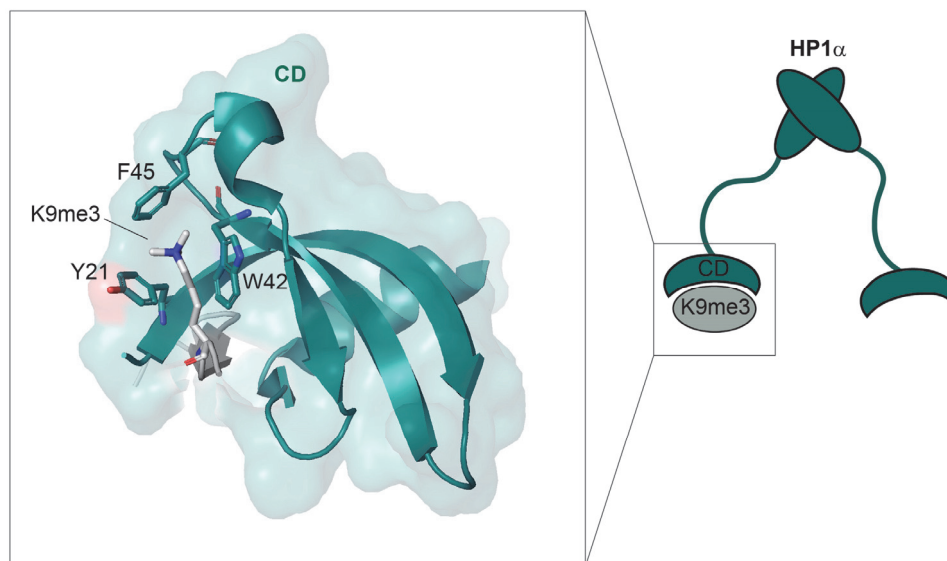
Figure 5: **Crystal structure of H3K9me3 peptide bound to HP1α chromodomain**. CD residues Tyr21, Trp42 and Phe45 form an aromatic cage that interacts with the trimethylated lysine through cation-π interaction (PDB: 3FDT).

Methyllysine effectors can recognize this PTM either by means of a cavity-insertion recognition mode (mostly for mono and di-methyl recognition) or using a surface-groove recognition mode (di- and tri-methylation).[63] In the first case (i.e. tandem tudor) the methylation mark is buried in a deep cleft that potentially performs a size-dependent selection.[64,65] In the second case (i.e. HP1α chromodomain –

H3K9me3, **Figure 5,** or the WD40 domain of PRC2 EED subunit – H3K27me3) the methyllysine lies along a surface groove, which is much wider and more accessible, leading to a less stringent preference for specific lysine methylation states.[66,67,68,69]

Beyond reader domains that interpret PTM combinations, catalytically active units that deposit ("writers") or delete ("erasers") specific marks also exist. Their opposing role is crucial in the regulation of histone marks turnover and homeostasis. Among the writer domains, histone methyltransferases (HKMTs) are a class of enzymes that catalyze the (mono, di or tri-) methylation of lysine or (mono, di-) methylation of arginine residues using S-adenosyl methionine (SAM) as a methyl donor. Although several families of HKMTs exist, the general mechanism for the methyl transfer is a nucleophilic substitution (SN2). The deprotonated nucleophilic nitrogen atom of the lysine/arginine residue attacks the methyl group on the sulfur atom of the SAM molecule (**Figure 6**).[70]
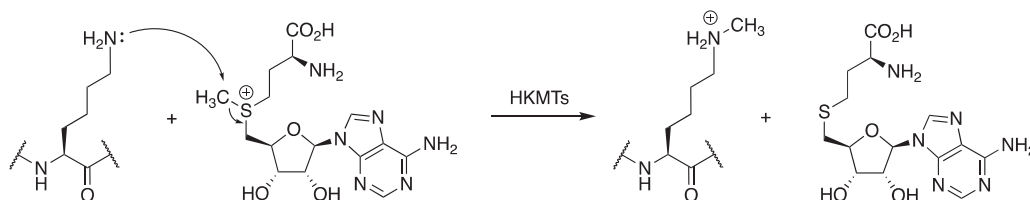
Figure 6: Mechanism of lysine methyl transfer reaction catalyzed by HKMTs using SAM.

The first lysine methyltransferase domain, SUV39H1, was identified by Jenuwein et al. in 2000.[71] This enzyme is specific for depositing H3K9me3 (starting from the monomethylated state), the key PTM defining the constitutive heterochromatin state and recognized by HP1.[72] Since then, dozens of KMTs with different substrate specificity and methylation degrees capabilities were identified. Some examples include MLL1 (mixed-lineage leukemia) a complex that preferentially installs the trimethylation of H3K4; PRC2 which catalyzes the trimethylation of H3K27 through its EZH2 subunit (Enhancer of zeste homolog 2);[73] Set8, a KMT that exclusively monomethylates H4 at position K20;[74] and SETD2, the sole enzyme responsible for H3K36me3 establishment.[75] Among the PTMs, methylation is one of the marks with the lowest turnover, indeed, the existence of lysine demethylases had been a long-standing subject for debate, until the first lysine demethylase (KDM), LSD1, was identified in 2004.[76]

Other examples of writer domains are histone acetyltransferases (HATs). These enzymes transfer an acetyl group from the acetyl-CoA to the side chain of a lysine residue.[77] The first reported example of a HAT is GCN5, a member of the GCN5-related N-acetyltransferase (GNAT) superfamily, that specifically acetylates H3K14, H4K8 and H4K16.[78] HAT activity is counterbalanced by histone deacetylases (HDACs) which, in turn, erase the acetyl mark. Their antagonistic role controls the genome-wide levels of acetyl groups on histones and thus regulate gene transcription.[79] Finally, histone phosphorylation is installed by histone kinases from ATP and removed by histone phosphatases. It is involved in diverse nuclear events, such as

DNA damage repair,[80] transcription regulation[81,82] and chromatin condensation during mitosis and meiosis.[83]

1.1.5 Multivalent chromatin effector complexes

In nature, most of the chromatin effectors are large complexes formed by the association of multiple domains that recognize, install or remove a certain histone PTM. In fact, the affinity between a substrate and an isolated reader domain, such as the ones described above, is generally in the micromolar range and the selectivity is relatively modest.[63] However, when multiple subunits are brought together through complex formation, both the affinity and the specificity for their targets undergo a dramatic enhancement. This phenomenon is known as "multivalency", and it is caused by thermodynamic and kinetic effects.[84] In fact, the prearrangement of the domains in a complex significantly decrease the loss of degrees of freedom derived from the binding event. These reduced entropic costs result in a gain of free energy which consequently make the affinity stronger (lower dissociation constant K_D).[63] In addition, in kinetic terms, the preorganization of these binding modules induces an increase in their local concentration that, besides raising the association rates, it also favors rebinding after dissociation, thus decreasing the dissociation rates (**Figure 7**).
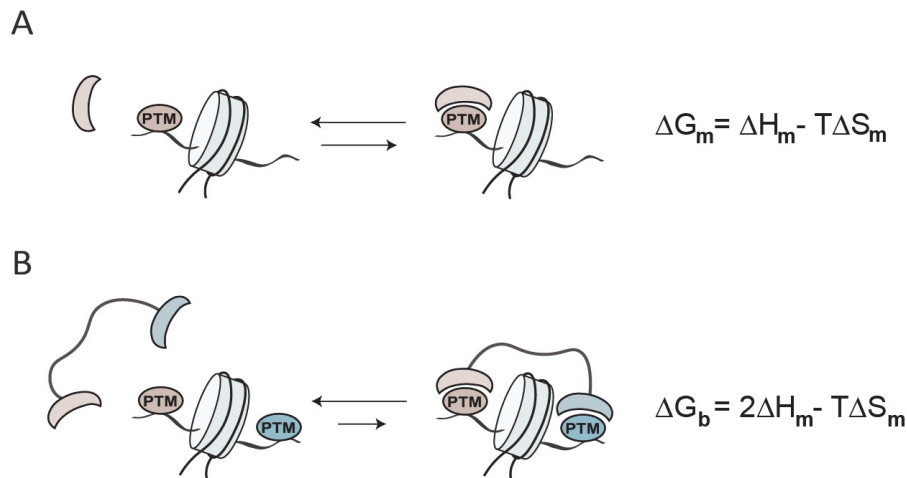
Figure 7: Chromatin-effector multivalency enhances the affinity and specificity for the target nucleosome. A-B) Comparison of monovalent (**A**) and bivalent (**B**) binding to nucleosomes. The prearrangement of the reader domains into a complex lowers the entropic costs of the second binding event. m: monovalent, b: bivalent.

Some of these multivalent effectors increase the binding affinity to a specific mark by combining repeats of the same binding domain in their structure, simultaneously engaging two substrates. Examples of such multivalent chromatin effectors are the members of the HP1 family, which can dimerize through interaction between their Chromoshadow domain (CSD).[85] By contrast, other multivalent effectors bear different domains that can potentially interact with distinct marks on the same or on different histone tails. The tumor suppressor p53-binding protein 1 (53BP1) is an example of trivalent chromatin effector protein that plays essential roles in the cellular response to Double Strand Breaks (DSBs).[86] Further, PRC2 is a key

epigenetic regulator that contribute to the maintenance of gene silencing[87] and is organized as a large multiprotein complex, where each subunit has specific functions and catalytic activities.[88]

In summary, all these examples show how the specific combination of multiple domains, either identical or with a different function, can enhance the strength of the interaction between effectors and chromatin PTMs and, importantly, enables the interpretation of the histone code.

1.1.6   Bivalent domains and asymmetrically modified nucleosomes

As described in the previous chapters, nucleosomes are the basic building blocks of the chromatin structure and they can be subjected to a multiplicity of PTMs that regulate crucial chromatin-dependent processes. The intrinsic symmetry of the nucleosome particle has long been appreciated, however, the presence of two copies of each histone protein raises some rational questions: Are histone PTMs exclusively established in a symmetric manner? Are sister histones in individual mononucleosomes modified at identical positions with exactly the same type of modification? If not, are there any defined asymmetries and what is their function?

These questions had remained unanswered for a long time, mainly because the most widespread methods used for the investigation of histone mark localization throughout the genome cannot prove PTMs colocalization on the same nucleosome nor distinguish between the tails of histone pairs within a nucleosome (**Figure 8**).
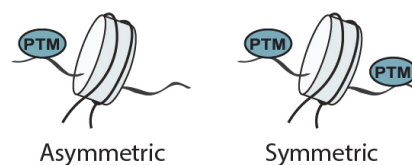
Figure 8: Nucleosomes can be asymmetrically modified. Asymmetric (left) and symmetric (right) nucleosomes.

These technologies include chromatin immunoprecipitation (ChIP) combined with high-throughput sequencing (ChIP-Seq)[89]. The ChIP experiment can either rely on the fixation of chromatin with formaldehyde followed by fragmentation by sonication (XChIP) or make use of native chromatin prepared by micrococcal nuclease digestion (NChIP). The chromatin fragments are then selectively immunoprecipitated with antibodies against the desired histone PTMs and analyzed by next-generation sequencing. Although ChIP data are very informative regarding single histone modification genomic enrichment, they do not provide an ultimate answer when considering multiple combinations of PTMs. In fact, apparent colocalization might derive from different cell populations or PTMs lying on different alleles. Sequential ChIP (Re-ChIP) solves this problem by sequentially immunoprecipitating the same chromatin population with antibodies specific for the different PTMs of interest. A second drawback of ChIP is the relative nature of the measurement, making the comparison between experiments difficult. A method called Internal Standard Calibrated ChIP (ICeChIP) addresses this issue by spiking in native chromatin with

barcoded semisynthetically modified nucleosomes prior to IP, allowing the quantitative detection of histone modification densities.[90] Regarding the genomic localization of histone modifications, these technologies are highly instructive. However, all these methods cannot establish whether the modifications originate from the same or distinct nucleosomes.

1.1.7 Embryonic stem cells and the bivalent domains

Extensive efforts have been made to map and understand the PTM distribution on embryonic stem cell (ESC) chromatin. ESCs are cells derived from the inner cell mass of an early-state preimplantation blastocyst. They are pluripotent and capable of self-renewal: they can divide while maintaining the undifferentiated state, yet they have the potential to differentiate into any of the lineages of the developing and adult organism.[91] These characteristics make ESCs especially interesting from a biological and medical point of view and triggered thorough investigations of the mechanisms controlling these unique properties. The ESCs state is primarily modulated by transcription factors (TFs) which, by binding to specific DNA sequences in proximity to the promoter, ensure the correct transcription rate of the genetic information. Although this regulation is mostly driven by the DNA sequence, chromatin PTMs and its associated factors play a fundamental role in controlling gene expression. Indeed, genome-wide mapping studies of PTMs in ESCs have unveiled the presence of defined histone modifications at particular genomic domains; while some of these correlations are recurrent in all mammalian cell types (i.e. acetylation of H3K27 at active enhancers, H3K4me3 at active promoters, H3K27me3 at repressed genes), others are specific to ESCs and few other cell lines.
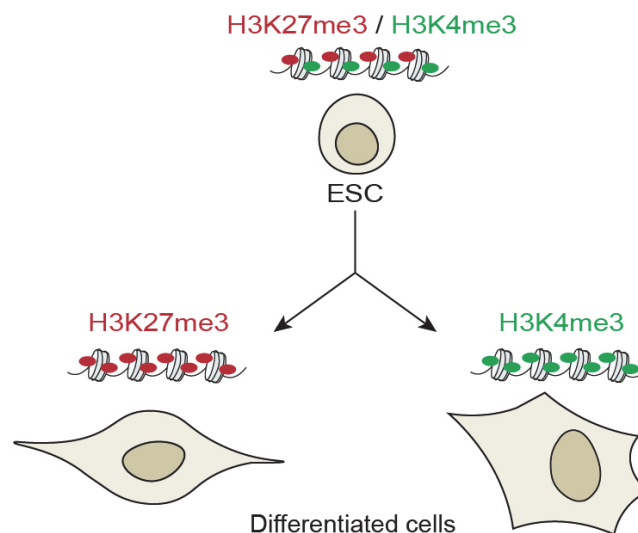
Figure 9: Fate of the bivalent domains upon pluripotent cells differentiation. Developmental genes in ESCs are temporarily repressed by the presence of the bivalent mark (H3K27me3/H3K4me3). When the cell differentiates, the bivalent domain is lost and the genes are either expressed (H3K4me3) or repressed (H3K27me3). The differential expression of developmental genes upon ESCs differentiation leads to different somatic cell types.

Using ChIP based methods, it was found that many promoters of developmentally regulated genes in ESCs and potentially in cancer cells are characterized by the presence of a distinctive histone modification

signature, called a "bivalent domain".[92,93,94,95] These regions are simultaneously marked by histone modifications of seemingly opposing roles: H3K27me3, a PTM associated with repressive chromatin and gene silencing and established by PRC2,[73] and H3K4me3 or H3K36me3, activation marks deposited by the Mixed-Lineage Leukemia (MLL) complex[96] and SETD2,[75] respectively, and found at promoters of active genes.[92] Bivalent domains are frequently found at promoters of developmental transcription factor genes expressed at low level. Importantly, upon cell differentiation this bivalent pattern tends to be resolved by selectively losing either the activating or repressing mark (**Figure 9**). Therefore, it was postulated that bivalent domains maintain the "stemness" of the cells by silencing developmental genes and preserving their potential to be rapidly activated upon initiation of specific developmental pathways.[92]

Recently, our laboratory has engineered genetically encoded multivalent probes (cMAPs) for the *in vivo* detection of bivalent domains (H3K4me3/H3K27me3), enabling the visualization of this combinatorial histone PTM pattern in living stem cells.[97] cMAPs consist of a fluorescent protein reporter fusion to two reader domains, each specific for a PTM (**Figure 10A**). The presence of the bivalent mark at defined nuclear foci was corroborated by confocal microscopy (**Figure 10B**), and changes in ESCs chromatin state were monitored over time upon treatment with epigenetic modulators.
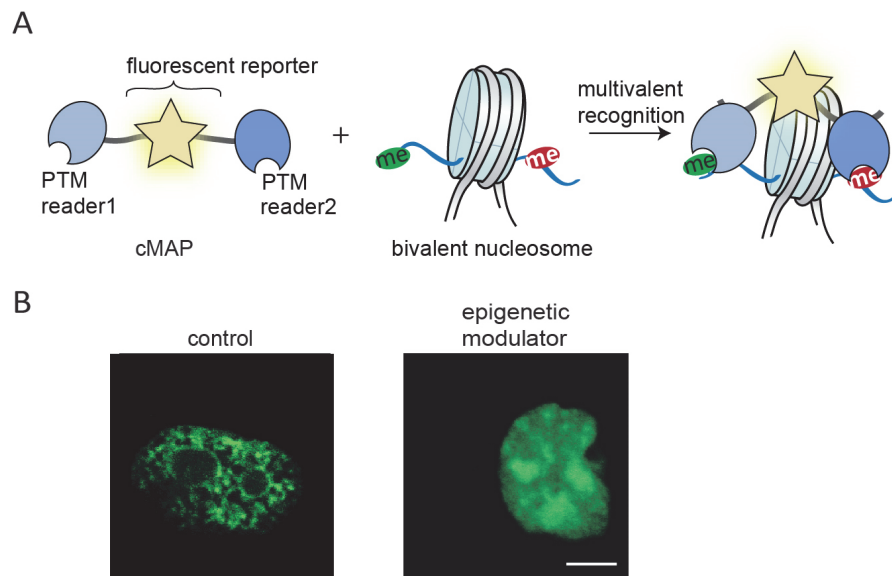
Figure 10: cMAPs reveal that bivalent domains are organized in clusters in living stem cells. A) Scheme of the cMAP concept. Star: fluorescent reporter, yellow fluorescent protein (YFP, Venus).[98] PTM reader 1: Polycomb chromodomain (Pc CD) reader of H3K27me3.[99] PTM reader 2: Transcription initiation factor TFIID subunit 3 (TAF3) PHD, reader of H3K4me3.[100] **B)** Localization of cMAP in live mouse ESCs treated and untreated with UNC1999, an epigenetic modulator and inhibitor of PRC2 methyltransferase activity.[101] Upon treatment the bivalent domains disappear and the probe diffuses homogenously in the nucleus. Scale bar: 5 μm. Figure adapted from Delachat et al. *Cell Chemical Biology* (2018).[97]

Thus, for the first time, bivalent chromatin was visualized in live cells and was found to form localized clusters in the nucleus. However, the promoter configuration of these bivalent sites was still unknown. In

fact, the activating and repressive mark may be present on adjacent nucleosomes within the promoter, on the same nucleosomes or even on the same histone tail.

1.1.8   Nucleosome post-translational modification configuration: symmetric or asymmetric?

The pioneering work by Voigt et al. shed some light on this intriguing question.[102] They developed a method based on successive affinity pulldown of micrococcal nuclease-digested chromosomal mononucleosomes from ESCs, cancer cells and fibroblasts using modification-specific antibodies, followed by tryptic digestion and Liquid Chromatography coupled to Tandem Mass Spectrometry (LC - MS/MS) analysis.[102] This methodology allowed to quantify the relative abundance of histone PTMs and proved that sister histones are not imperatively identical within a nucleosome. In particular, it was demonstrated that: 1) both symmetrically and asymmetrically modified H3K27me3 and H4K20me1 nucleosomes exist, and 2) the bivalent mark, i.e., H3K4me3 or H3K36me3 along with H3K27me3, lie on opposite tails. However, due to technical complexities, the symmetry of bivalent domains could not be definitively ruled out. Only a more recent single-molecule TIRF study of fluorescently labeled bivalent mononucleosomes confirmed that 0.5% of the nucleosomes in mouse ESCs are bivalent: the two opposing marks H3K4me3/H3K27me3 are predominantly deposited in an asymmetric configuration (94%), rather than co-occurring in the same H3 molecule (6%).[103]

The proof of existence of bivalent asymmetric nucleosomes triggered the investigation of PRC2 trimethylation activity in these domains. Methyltransferase assays showed that asymmetric H3K4me3 modification blocks the *in vitro* deposition of H3K27me3 by PRC2 on the same tail (in "*cis*") but not on the opposite one (in "*trans*") within a nucleosome, and that PRC2 preferentially modifies asymmetric H3K27me3 nucleosomes that already contain one H3K27me3 mark.[102,104] While the bivalent mark has been established as an essential requirement for ESCs maintenance and differentiation, and the recognition of asymmetrical H3K27me3 is thought to be important for heterochromatin spreading, little is known about the biological relevance of most of the other asymmetrical epigenetic marks.

Another example of the consequence of nucleosome asymmetry was described by Liokatis et al.[105]   They established a procedure relying on the same tandem affinity purification scheme[102] to generate ^{15}N or ^{13}C isotope-labeled asymmetrically modified nucleosomes that were subsequently incubated with H3-modifying enzymes. This approach allowed them to discriminate the two histone proteins and to selectively monitor single PTMs and the crosstalk between modifications in *cis* and in *trans* by time-resolved NMR. As a case study, they focused on the phosphorylation/methylation crosstalk between H3T3 and H3K4; the first being a PTM found at mitotic chromatin while the second is a hallmark of promoters of active genes.[106,107] By generating ^{13}C labeled H3K4me1 histones and combining them with ^{15}N unmodified H3 molecules, they reconstituted single asymmetric nucleosomes that were incubated with Haspin kinase, the enzyme responsible for H3T3 phosphorylation. While efficient phosphorylation of the unmodified H3 was detected,

none of the H3K4me1 sister histones were modified, thus highlighting the negative regulatory effect of the methylation mark on phosphorylation activity in *cis*.

These *in vitro* studies provided evidence for the importance of nucleosome asymmetry in modulating chromatin effector function and mode of action, but the *in vivo* outcomes of such asymmetric arrangements were yet to be investigated, as the natural symmetry of the H3-H3 surface made impossible to control the histone dimer formation in cells. An interesting approach to probe the functional consequence of nucleosomes asymmetry *in vivo* was introduced shortly later by Ichikawa et al.[108] Using rational design combined with *in vivo* optimization and selection, they were able to develop a pair of H3 proteins that form heterodimers *in vivo* but cannot homodimerize. This allowed them to introduce asymmetric sequence mutations and explore their effects on transcriptional regulation and on histone crosstalk. Although innovative, this *in vivo* methodology can only deal with natural amino acid sequence mutations. Indeed, to generate asymmetric PTMs – carrying heterodimers, this strategy must be combined with unnatural amino acid incorporation by amber codon suppression,[109] which adds further complexity to the system, or with peptide ligation schemes (see **Chapter 1.3.3**), that consequently restrain its use to *in vitro* applications. In addition, the unnatural H3-H3 contact surface might influence the histone octamer/DNA interactions and movements, thus affecting the highly dynamic behavior of the chromatin-effector proteins interactions.

These results, together with the evidence that histones can be asymmetrically substituted with histone variants and that nucleosomes can lack one H2A-H2B dimer,[110] indicate that asymmetry could be a general feature of nucleosomes that greatly expands the combinatorial cross-talk of histone marks and plays important roles in the mechanisms of chromatin effectors enzymatic activity and recruitment. Therefore, asymmetrically modified nucleosomes may expose an altered binding surface for chromatin effector proteins and the multivalent recognition of differentially modified histone sisters may provide a new way to accommodate inter- and intranucleosomal interactions. The existence of symmetric and asymmetric modification populations adds an additional layer of complexity to the readout of the PTM histone code. Moreover, asymmetry may calibrate the affinities of single chromatin-binding domains for the specific PTM pattern and thus regulate the general recruitment of effector proteins to chromatin.

Although evidence for such interplay between nucleosome asymmetry and gene regulation exists, the mechanisms underlying the establishment, maintenance and function of the majority of the known asymmetric PTMs are still unclear. This is mainly due to the fact that such nucleosomes are not easily available for detailed *in vitro* mechanistic studies. Indeed, the fundamental prerequisite for studying such molecular processes is the facile chemical access to defined and pure modified asymmetric nucleosomes. In **Chapter 2** and **Chapter 3**, I will explain in detail the modular strategy that we developed and adopted to synthesize asymmetric nucleosomes carrying either the H4K20me1 asymmetric mark or the bivalent signature H3K27me3/H3K36me3.

1.2 Liquid-liquid phase separation in biology

1.2.1 <u>Overview</u>

Eukaryotic cells must organize complex and specific biochemical reactions in space and time. In fact, the organization of macromolecules within a cell is vital for many cellular regulatory processes such as development, division and homeostasis. This is accomplished by creating compartments within the cell and thus generating distinct chemical environments. Besides physically separating the biological matter, compartmentalization also enhances the efficiency of many subcellular processes by selectively enriching certain species while excluding others.[111]

A biological compartment has two important properties: it is delimited by a boundary and the components within it must be free to diffuse, react and interact. Many such compartments are delimited by membranes: the nucleus, for instance, provides a physical barrier that separates transcriptional-related processes from translation.[112] Other examples include lysosomes, which provide the chemical environment necessary for protein degradation, or the mitochondria, where ATP is produced.[113] However, cells also harbor organelles that lack a delimiting membrane and are thus defined as "membraneless" (**Figure 11**).[114]
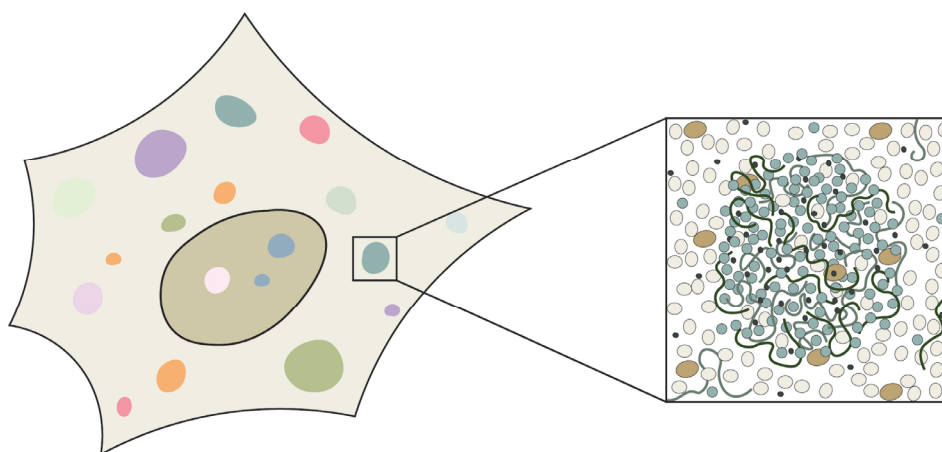
Figure 11: Phase separated biomolecular compartments in cells. Membraneless compartments concentrate specific molecules and exclude others, thus creating selective microenvironments that delimit specific chemical reactions in space and time.

Examples are the nucleoli, sub-compartments of the nucleus responsible for generating the ribosomes,[115] or the centrosomes, located in the cytoplasm and nucleating the microtubules.[116] How can these bodies demix from their surrounding? Why do their components not simply diffuse and combine with the neighboring environment? Many other cytoplasmic and nuclear membraneless organelles (i.e. P bodies,[117] PML bodies,[118] Cajal bodies,[119] nuclear speckles[120]) had been studied for a long time, yet the physical nature of these assemblies was elucidated only recently. In 2009, P granules were shown to fuse, to be deformed by flows, to be relatively viscous and to be able to exchange their components rapidly with the surrounding milieu.[121] Such liquid-like properties suggest a phase-separation mechanism for the formation of droplet-like organelles. Indeed, two years later, the same features were shown for nucleoli.[122]

Phase separation is a physical process that describes the stable coexistence of two liquid phases (a dense phase and a diluted phase) which occurs when a supersaturated solution of components spontaneously separates. It is driven by the physical interactions between the molecules within a phase and creates a favorable low-energy system that counterbalances the higher entropy of the unmixed state.[123] As a result, a condensed phase suspended in the bulk solution is formed. Thus, to create a stable droplet, two physical requirements have to be met: supersaturation of the self-interacting molecule and critical size of the drop.
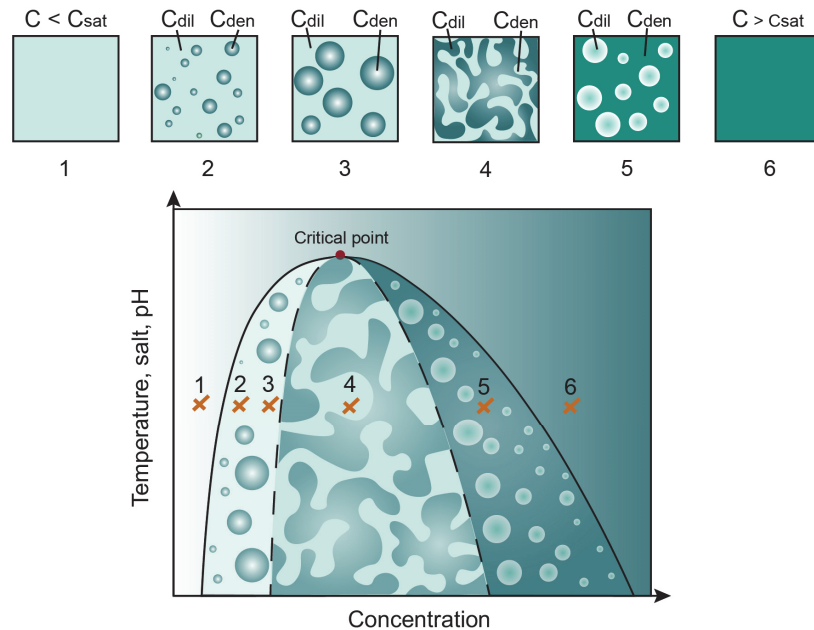
Figure 12: Schematic phase diagram. Phase diagrams are generated experimentally and subdivide the conditions (of concentration vs. another parameter such as temperature, salt and pH) that result in a single and homogeneous phase and conditions that induce phase separation (binodal line, solid). Within the binodal line, another line defines the region of instability where the system undergoes phase separation by spinodal decomposition (spinodal line, dashed). In between the spinodal and binodal line, the solution demixes when nucleated. Beyond the critical point, no phase separation occurs. C: concentration. C_{sat}: concentration of saturation. C_{dil}: concentration of the diluted phase. C_{den}: concentration of the dense phase. For $C < C_{sat}$ the protein diffuses in solution. When $C > C_{sat}$ dense droplets form (C_{den}). As the total C increases further, C_{dil} and C_{sat} stay constant but the size of the droplets increases to accommodate the additional molecules.

Indeed, the energetic costs necessary for creating an interface (directly proportional to the droplet surface) must be lower than the energy gained from the interactions within the droplet (proportional to the droplet volume). While this phenomenon had been known and applied for long time to others fields, such as polymer chemistry, and some proteins had been reported to phase-separate *in vitro* at high concentrations,[124,125] the concept that links liquid-liquid phase separation (LLPS) to biomacromolecules organization is relatively new. In 2012 the Rosen lab provided the first evidence for the role of LLPS in governing the formation of biomolecular condensates, by reconstituting protein- and RNA-containing bodies that could induce nucleation of actin polymers.[126] They reported that NCK and N-WASP, two signaling proteins that mediate actin polymerization, can undergo phase separation *in vitro* and that protein concentration and valency of the interacting partners are crucial in determining the phase separation threshold. From then on, copious works supported this theory[127] and other environmental

conditions such as the presence of specific PTMs,[128,129,130] temperature and pH[131,132,133] were found to control the assembly and disassembly of these biomolecular compartments (**Figure 12**). Phase separation could then constitute a mechanism by which cells respond to external stimuli and stress.

These works prompted further research into the molecular characteristics that promote condensates formation. LLPS appears not to be accessible for many proteins under the conditions in living cells, thus some specific molecular signatures might be responsible for the phase separating behavior at physiological conditions. The concept of "scaffold" and "client" molecules highlights some of these important properties. Scaffolds are molecules that drive the phase separation process, while clients partition into the condensates formed by the scaffolds; both are required to create a network of dynamic interactions (**Figure 13**).
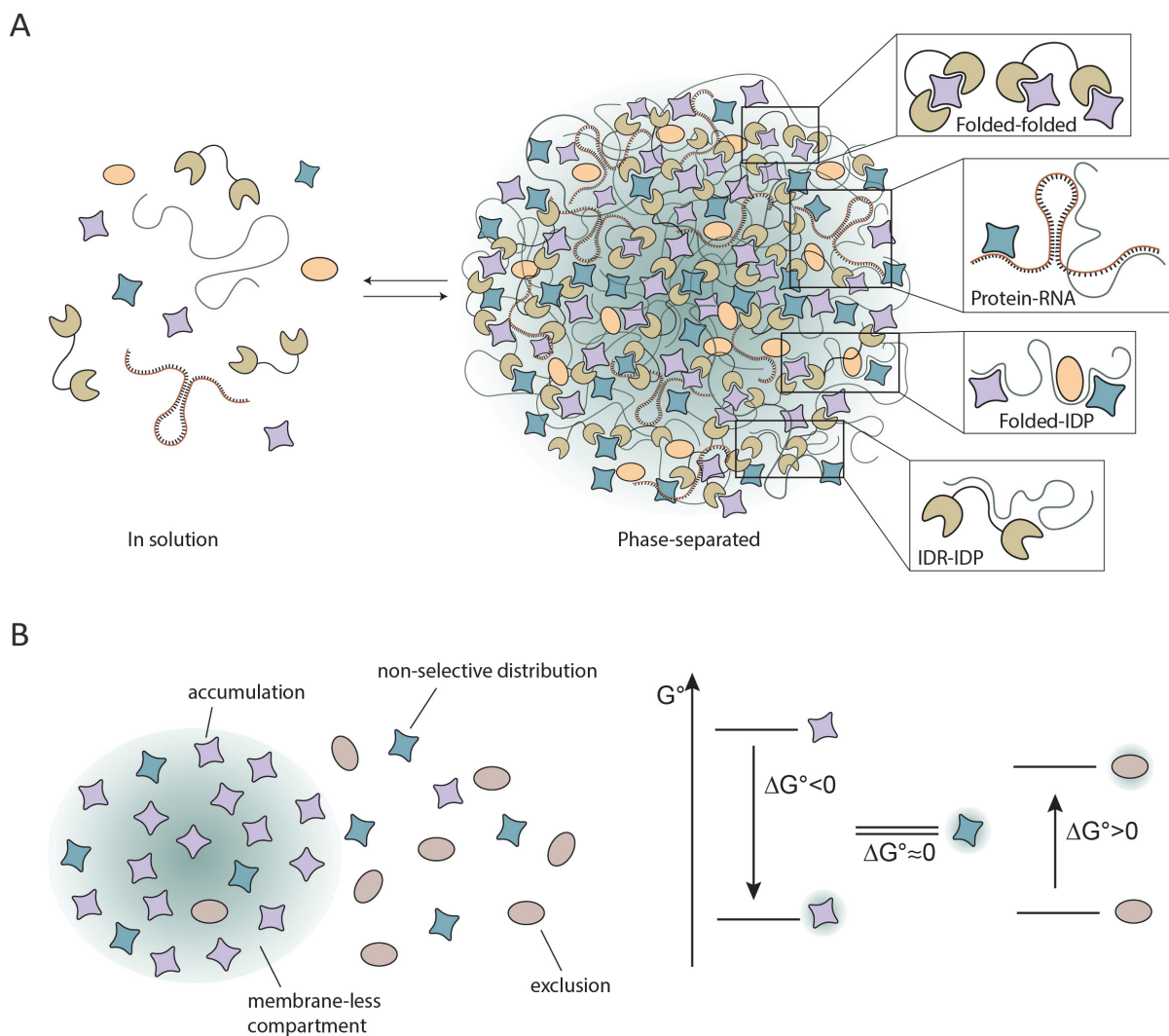
Figure 13: Liquid-liquid phase separation organizes the biomolecules in cells. A) Multivalency, intrinsically disordered domains and protein-protein interactions support the network of dynamic interactions that drive phase separation. **B**) Client molecules accumulate into the droplet ($\Delta G° < 0$), are excluded from the compartment ($\Delta G° > 0$) or do not show any significant preferences ($\Delta G° \approx 0$).

This dynamic interplay is supported by three main features: multivalency, intrinsically disordered regions (IDRs) or intrinsically disordered proteins (IDPs), and protein-protein interactions (PPIs). Multivalency

consists in the repetitive display of one or more domains within a protein and genetic manipulation of protein valency revealed a strict relationship between saturation concentration and valency level.[134,134,126] The repetitive modules can either be folded domains or low complexity disordered segments; multidomain proteins often carry several folded protein-protein interaction motifs connected by flexible linkers. The folded domains are implicated in specific PPIs or in protein-nucleic acid interactions and provide the organizational platform for assembly. On the other hand, IDRs and IDPs promote more dynamic interactions with a broader palette of binding partners; they have no fixed secondary and tertiary structure and multiple low-energy conformational states coexist. More specifically, their sequence lacks aromatic and aliphatic residues, normally responsible for the folding of globular domains.[135,136,137] The combination of IDRs and globular domains within multivalent proteins enables the synergy between the different components of the droplet, provides a high degree of conformational flexibility that is required to keep the binding events dynamic and uncoupled, and contributes to the liquid-like properties of phase separated molecular bodies.

1.2.2   Liquid-liquid phase separation in the chromatin context

A myriad of chemically and structurally distinct macromolecules coexist at high concentration inside the nucleus of a eukaryotic cell and high concentrated solutions of macromolecules tend to undergo phase separation.[138] If we consider that huge amounts of chromatin are densely packed within the nuclear membrane, yet complex and independent processes are executed with precise spatial and temporal control inside this crowded environment, one could rationally hypothesize that chromatin sub-compartments are arranged as phase-separated assemblies.

As already described in **Chapter 1.1.1**, the genome is organized on multiple levels of complexity; clusters of genes occupy distinct spatial territories and are surrounded by different environments. These subcompartments are enriched or depleted of specific enzymes and other macromolecules and the accumulation of specific proteins is key for the determination of a defined chromatin state, such as the highly compacted and transcriptionally silenced foci (heterochromatin) or the relaxed and active genomic regions (euchromatin). As mentioned in **Chapter 1.1.2**, heterochromatin is a chromatin structure that is crucial for the clonal inheritance of cell identity, chromosome segregation, regulation of genes expression and genome stability.[139] The prevalent and widely accepted view of heterochromatic silencing relies on the decoration of histone H3 N-terminal tail with the trimethylation of lysine 9 (H3K9me3). This PTM is established by the methyltransferase SUV39H1[71] and provides binding sites for HP1 and other factors,[52,140,53] which in turn trigger chromatin compaction and sterically hinder the binding of regulatory proteins, such as RNA polymerase, with consequent inhibition of gene transcription.[141] However, compaction alone cannot explain the formation of defined, membraneless heterochromatin sub compartments in the nucleus. These domains are accessible by regulatory factors in a highly dynamic way and proteins within the domain exhibit fast diffusion. Furthermore, it is known that heterochromatin-related repressive functions can "spread" directionally along the genome, that is, expand in a sequence-

independent and stochastic manner leading to the inactivation of the neighboring genes.[142] It was shown that *S. pombe* HP1 oligomerization actively participates in gene silencing and that the SUV39H1 methylase homologue Clr4 is positively regulated by its own product.[143] A two-state mechanism, involving anchoring of SUV39H1 to H3K9me3 and subsequent allosteric activation, was found to promote methylation in cells.[144] In addition, human HP1 can directly bind to SUV39H1,[145] thus suggesting a positive feedback loop between the H3K9 methylation mark and HP1 binding. Nevertheless, the mechanisms of such large-scale spreading of silencing are not well understood. Similarly to heterochromatin spreading, also gene activation exhibits such a long-distance control: the binding of transcription factors at enhancers determines the function of polymerases found at promoters at a distance of hundreds or thousands of bp.[146] Several plausible mechanisms were proposed, such as looping, twisting, sliding, oozing or hopping.[142] However, gene transcription and repression seem to be dynamic processes that involve local diffusion of chromatin effectors and the creation of a platform for the recruitment of additional functions. Demixing and phase separation could provide a new way of conceiving and interpreting chromatin states formation, function and responses to the changing environment.

1.2.3 HP1 as driving force for phase separation and heterochromatin formation

HP1 is a central component of the constitutive heterochromatin and is present in three isoforms in mice and humans: HP1α, HP1β, HP1γ.[147,148] While their CD (residues 21-71)) recognizes the H3K9me3 modification,[66] the CSD (residues 120-179) allows dimerization, enables the cross-bridging of neighboring nucleosomes and acts as a platform for recruiting other effectors (**Figure 14A**).[149,150] While the CD is similar to other CDs of different methyllysine binders, the CSD is unique to HP1.[151] In addition to these two structured domains, HP1 proteins exhibit three unstructured regions: a C- and a N-terminal extensions and a positively charged hinge region linking the CD to the CSD and responsible for non-specific DNA and RNA binding.[152] Although their CD and CSD sequences are conserved for more than 90%, the three paralogs perform different functions and exhibit distinct localizations inside the nucleus.[153] While HP1α is commonly associated with silent heterochromatin, the β and γ play roles in both activation and repression of genes.[154] Since phosphorylation of human HP1α N-terminal extension (NTE) is known to boost heterochromatin foci formation by increasing the affinity and specificity of HP1α for the H3K9me3 mark,[155,156] and these phosphorylation sites are absent on HP1β and HP1γ paralogs, the different roles of the HP1 proteins might stem from variations in the unstructured regions rather than the conserved domains. Further, the combination of multivalency and the presence of repetitive folded and disordered regions within the protein structure make HP1α an excellent candidate for driving phase separation. Indeed, two independent recent studies found that HP1α can demix from the rest of the nucleoplasm and form liquid droplets (**Figure 14A**).[157,158]
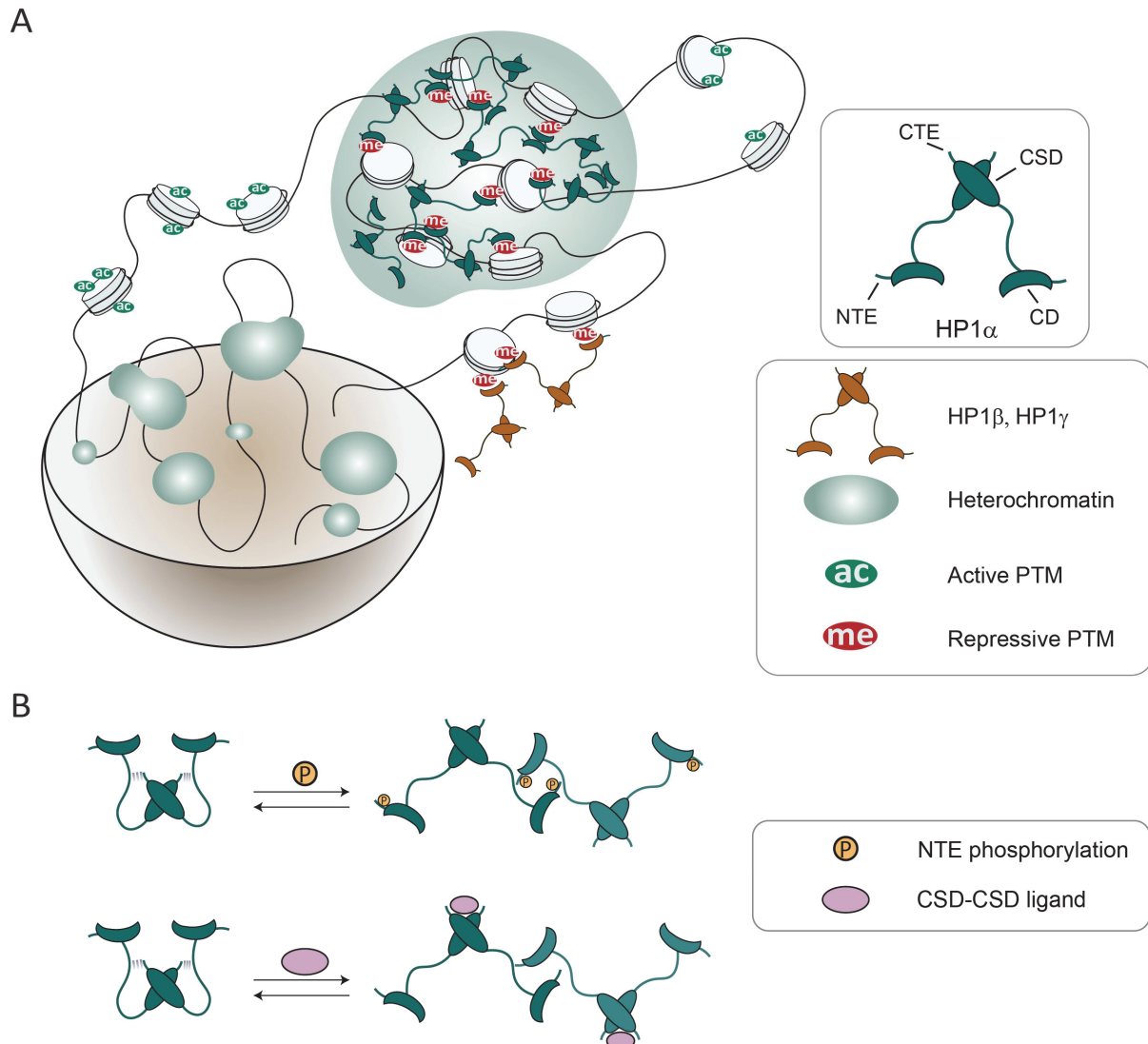
Figure 14: Heterochromatin foci are formed by liquid-liquid phase separation. A) Scheme of the heterochromatin biomolecular condensates (green droplets) in the nucleus (brown) including HP1α and the H3K9me3 modified nucleosomes. Acetylated nucleosomes are excluded from the droplets. HP1 isoforms β and γ cannot oligomerize and thus are not involved in the phase separated heterochromatin compartment. **B**) HP1α CTE interacts with the hinge region within the same molecule when it is not phosphorylated, thus inhibiting multivalent interactions. Upon NTE phosphorylation, oligomerization occurs. In addition, the dimer CSD-CSD interface is home to the binding of a great variety of HP1α binders. These ligands can either stabilize the elongated dimer conformation (phase separation induction) or, by contrast, promote the close state (dissolution of the phase separate droplets).

In one of the studies, Larson and coworkers showed that phosphorylation of the NTE or DNA binding by the positively charged hinge region are antagonistic and necessary prerequisites for triggering *in vitro* phase separation of human HP1α through an electrostatically driven process.[157] The N-terminal phosphate groups interact with the positively charged hinge region and promote oligomerization by fostering inter-dimer connections (**Figure 14B**). On the other hand, the CTE region favors a closed dimer conformation by interacting with the hinge region on the same dimer (**Figure 14B**). Such an inhibited state is unable to make multivalent interactions and consequently has a higher saturation concentration.[157] While DNA had a stimulatory effect on unmodified HP1α droplets formation, phosphorylation disturbed the cooperative

binding between DNA and HP1α and slowed down DNA compaction. Thus, DNA compaction activity by HP1α seems to be the main contributor to heterochromatin packaging and HP1α phosphorylation serves as a means to calibrate chromatin density and the size of the heterochromatin bodies.[157]

The partitioning properties of some nuclear components into the phosphorylated HP1α phase was then investigated *in vitro*.[157] Proteins and macromolecules that interact with HP1α, such as H3K9me3 chromatin fibers and nucleosomes, accumulated into the droplets, while other inert compounds were either excluded or did not show a significant preference (**Figure 13B**). Furthermore, recent findings by Wang and co-workers highlighted the primary role of histone PTMs in regulating chromatin compartmentalization. H3K9me3 nucleosome arrays induced robust droplets formation when interacting with multivalent CD-containing complexes such as HP1β-SUV39H1.[159] Therefore, PTMs seem to be a critical force for phase separation rather than simple participants to the process.

Another perspective on HP1a phase separation was concomitantly provided by Strom et al.[158] They first showed that *Drosophila* HP1a is able to phase separate *in vitro* and that the process does not rely on DNA binding or NTE phosphorylation. They then uncovered some intriguing features of HP1a droplets formation, growth and dissolution performing *in vivo* experiments on early stage *Drosophila* embryos and mammalian cells. GFP-HP1a fusion proteins could nucleate, form spherical bodies that grow by fusion and dissolve during mitosis or when the weak hydrophobic interactions among macromolecules were disrupted. These observations make *in vivo* heterochromatin foci plausible liquid-liquid phase separated compartments. Furthermore, inert macromolecules were excluded from heterochromatin foci in a way dependent on HP1a density rather than nucleosome compaction, indicating that their inability to penetrate heterochromatin foci is predominantly determined by the selective permeability of the heterochromatin phase rather than a result of steric hindrance and reduced accessibility of the compacted chromatin fiber.

Interestingly, the sphericity of the heterochromatin droplets decreased with maturation of *Drosophila* embryos. FRAP analysis revealed that loss of circularity was accompanied by an increase in the ratio of immobile versus mobile HP1a fractions. Further evidence of partial HP1a immobility was provided by the incomplete displacement of HP1 by the aliphatic alcohol 1,6-hexanediol, a molecule that interferes with weak hydrophobic protein-protein interactions. These findings agree with previous observations of the differential dynamics of HP1α in cells, that range from fast to immobile.[160,161] Most recently, the liquid-to-gel transition of HP1α was further corroborated and investigated by solid-state NMR studies, where the molecular interactions and dynamics responsible for the in *vitro* HP1α gelation events were monitored and identified.[162]

Taken together, the enrichment of heterochromatin-specific molecules and exclusion of gene activation-related compounds at heterochromatin compartments suggests to affect the physico-chemical features of the chromatin environment, such as binding affinities and equilibria, kinetics and enzymatic rate constants.

Phase separation provides a mean to sequester and compact chromatin and correlated repressive factors. Such a mechanism represents an alternative explanation to the canonical view of how heterochromatin is established and can spread along the genome. These new discoveries lead to a revisit of the classical view underlying heterochromatin function, formation, maintenance and spreading in cells, from a phase-separation point of view:

1) First, the coexistence of soluble, liquid and gel-like HP1α phases might represent a means of controlling different functions inside the cell. One can speculate that the soluble HP1α molecules are located at developmentally regulated regions of the genome, where the chromatin fiber is more permissive and accessible and genes are not constitutively repressed. During development, the HP1α proteins condensate and form phase-separated droplets that reduce the access of the transcription machinery to developmentally repressed genes. Finally, the most inaccessible gel state, where the strong interactions among molecules hinder diffusion and free rotation, represents heterochromatin in fully differentiated cells (**Figure 15**).[163] In addition, a solid HP1α network might be involved in pathogenesis, as observed for other phase-separating proteins that can undergo gelation, such as the FUS protein.[164]
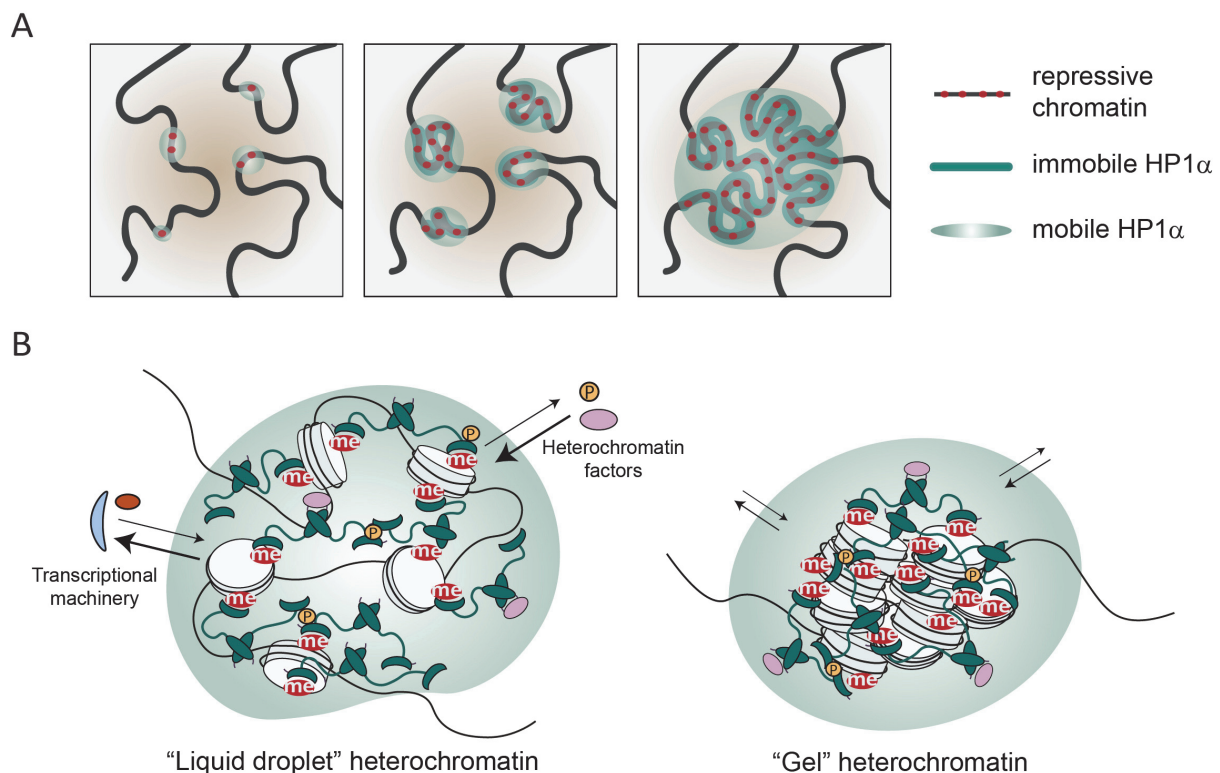
Figure 15: Heterochromatin undergoes liquid-liquid phase separation and gelation. A) Scheme of the formation and spreading of heterochromatin liquid droplets. The droplet growth is accompanied by an increase in the ratio of immobile versus mobile HP1α molecules. **B**) HP1α liquid-liquid phase separation (left). Ligands that bind the CSD-CSD interface, H3K9 methylases and other repressive effectors facilitate this process. Their ability to interact with heterochromatin-related components acts as an "entry ticket" that ensures their accumulation in the droplet. On the other hand, members of the transcriptional machinery are excluded. Certain heterochromatin components might stabilize a gel-like heterochromatin state (right). Such a state is less permeable and more compacted and might play a structural role.

2) The transition between states has to be tightly regulated to ensure a dynamic response to differentiation and the changing environment. The ligand-dependent conformational switch between close and open states reported for HP1α dimers (**Figure 14B**) might represent one way of modulating the phase-separation ability of HP1α. On the other hand, PTMs play a fundamental role in the regulation as well: serine phosphorylation of the NTE competes with DNA for binding to the hinge region, boosts HP1α oligomerization and thus might control the size and density of the heterochromatin bodies.

3) Phase separation might stand in as a sensor of the changing environment. If the HP1α concentration is kept close to the threshold for phase separation, even small variations in concentration cause dramatic perturbations of the phases equilibrium and consequently of the physical state of heterochromatin.[131,165] In addition, the enrichment/depletion of specific compounds in heterochromatin assemblies provides a method to create a selective microenvironment where only molecules needed for the function of a particular compartment are allowed to access, while others are excluded. Thus, phase separation keeps opposite functions spatially separated.

4) The existence of conserved CSD among HP1 paralogs implies that heterodimers can arise from the combination of HP1α with HP1β or γ. Since HP1β does not phase-separate when phosphorylated or in the presence of DNA, nor compacts DNA,[157] one could hypothesize that such heterodimers reduce the phase-separating tendency of HP1α and consequently act as a barrier to limit heterochromatin spreading and DNA compaction (**Figure 14A**).

5) Finally, heterochromatin spreading in cells could arise from chromatin regions of the genome that partition into a phase separated heterochromatin droplet. Thus, complementary to the classical view where HP1α oligomerization drives heterochromatin spreading throughout the genome, phase separation provides an additional valid explanation to this puzzling mechanism. Regarding compaction, HP1α can compact DNA into dense puncta, yet phosphorylation of the NTE region slows down the process.[157] Phosphorylation could then be a regulator of DNA compaction rate. An exciting recent work by Sanulli et al. uncovered the link between chromatin compaction and HP1-mediated phase separation.[166] The *S. pombe* HP1 analogue Swi6 was found to trigger nucleosome reshaping with consequent exposure of buried histone residues. Such conformational change is thought to increase the likelihood of multivalent interactions among nucleosomes, thus promoting phase separation.[166]

1.2.4   Liquid-liquid phase separation controls other chromatin states

As a result of these inspiring new discoveries regarding the link between phase separation and gene silencing control, a general increasing interest in phase separation and chromatin related events stimulated researchers to reconsider the behavior of other chromatin states with a phase separation point of view. Indeed, transcriptionally active chromatin regions came into sharp focus. The regulation of gene expression

is based on the ability of the transcription machinery to get together at defined genomic loci. Three independent studies by Chong et al., Sabari et al. and Cho et al. revealed that the transcription machinery is organized in biomolecular condensates that form liquid phase-separated droplets in living cells.[167,168,169] Transcription factors, the complex Mediator, Brd4 and RNA pol II all formed defined puncta in living stem or cancer cells through their IDRs. The phase separation behavior of these compartments was assessed by FRAP analysis and addition of 1,6-hexanediol and further validated by live cells imaging, where foci were shown to fuse and form bigger condensates. In addition, phase separation provides a framework able to explain the simultaneous engagement of super-enhancers with their cognate gene promoter and consequent gene activation.[170] Most recently, new evidence for the liquid-like nature of another chromatin state, the damaged chromatin regions, was brought to light by Kilic et al. Live cell microscopy and CRISPR/Cas9-mediated endogenous tagging were employed to show that accumulation of 53BP1 in DNA repair compartments triggers phase separation and spatially and temporally coordinates the DNA damage response (DDR) and the downstream gene regulation.[171]

1.2.5 Need for new liquid-liquid phase separation detection tools

LLPS thus represents an attractive model for nuclear compartmentalization and spatio-temporal control of biochemical reactions in cell nuclei and particularly at chromatin foci. The investigation of the biophysical, biochemical and structural principles underlying the formation of such biomolecular bodies is key to understanding the physiology and pathogenicity of the corresponding biological processes and systems. The common workflow for identifying LPPS in cells includes a whole set of distinct experiments that rely on different technologies and techniques, such as FRAP, 1,6-hexanediol addition, analysis of cooperative movement at boundaries, detection of fusion and coalescence events. Each is informative of a specific feature of LLPS, but has to be supported by the others to drive a clear conclusion. Although the research in this field is developing tremendously fast, a method that can unequivocally asses liquid-liquid phase separation in live cells and discriminate among other phase-separation mechanisms is currently lacking.[172] Simple methods that allow to define LLPS in an unambiguous way are needed. In **Chapter 5**, I describe how we approached this problem by designing and synthesizing multivalent peptide probes for the detection of LLPS in live cells.

1.3 Peptide chemistry and semisynthesis of modified histone proteins

Having facile access to large amounts of homogeneously labeled or post-translationally modified histone proteins is the fundamental prerequisite for investigating the mechanisms underlying PTM establishment and their influence on chromatin effector-mediated processes, such as gene expression control and disease. Furthermore, specific attachment of small molecules, including fluorophores, affinity tags or crosslinkers finds broad applications in a wide variety of fields, ranging from microscopy studies for *in vivo* tracking of

protein-fluorophores conjugates,[173] to the functionalization of therapeutic antibodies,[174] or for dissecting the mechanisms of pathological enzymes.[175]

Solid phase peptide synthesis (SPPS) is a powerful technique that allows the synthesis of peptides carrying most of the common PTMs, as the modified amino acid building blocks can be inserted at any desired positions of the sequence. However, the synthesis of peptides longer than 50 amino acids is tricky and suffers from low yield (histone proteins length is > 100 amino acids). Although synthetic peptides derived from histone tails have revealed to be precious substrates for deciphering some PTM functions,[176,177,178,179] they cannot answer questions that involve the complex interplay existing between histones, DNA, and neighboring nucleosomes. Indeed, the consequences of most PTMs can only be fully elucidated when the PTM is in the context of a nucleosome or a chromatin fiber. Therefore, to overcome this length constraint, SPPS is typically combined with other chemistry-based strategies, such as fragments condensation or ligation and cysteine-directed modification, yielding full-length modified histones that can be incorporated into nucleosomes.

1.3.1 Unnatural amino acid incorporation

Modified histones can be obtained by purifying endogenously expressed and modified histone proteins; however, the position, type and number of PTMs cannot be controlled, leading to heterogeneous histone populations.[180,181] Furthermore, enzymatic reactions on purified unmodified histones have also been widely used; yet the enzyme responsible for a certain PTM might not be known, not sufficiently specific or may generate a mixture of modified and unmodified histones *in vitro*. Importantly, both methods are limited to the 20 genetically encoded natural amino acids. An alternative method relies on genetic code expansion by amber codon suppression.[109] In this case, unnatural amino acids (Uaa) with a unique side chain can be incorporated specifically into proteins by engineering an orthogonal tRNA/aminoacyl-tRNA synthetase pair. This enzyme loads the unnatural amino acid onto the CUA-anticodon tRNA, and the ribosomal machinery consequently introduces the modification into the growing protein at the UAG stop codon position (**Figure 16**). Using this strategy, various PTMs and other moieties were successfully introduced into histone proteins, such as acetyllysines (**Figure 16A**),[182] precursors of methyllysines (**Figure 16B**),[183,184,185] photoreactive Uaa (pBPA),[186,187] photoaffinity crotonyl and photoaffinity caged lysines[188] (**Figure 16C**) and δ-thiol containing lysines (**Figure 16D**).[189] This last unnatural amino acid deserves particular attention as it can react with a thioester-containing molecule to yield a native isopeptide bond (i.e. traceless ubiquitylation[189] (**Figure 16D**)). However, the amber codon suppression method presents several drawbacks: it is far from being modular, since a new tRNA/aminoacyl-tRNA synthetase needs to be developed for each Uaa, and the incorporation of multiple Uaas on the same protein remains very challenging.[190]
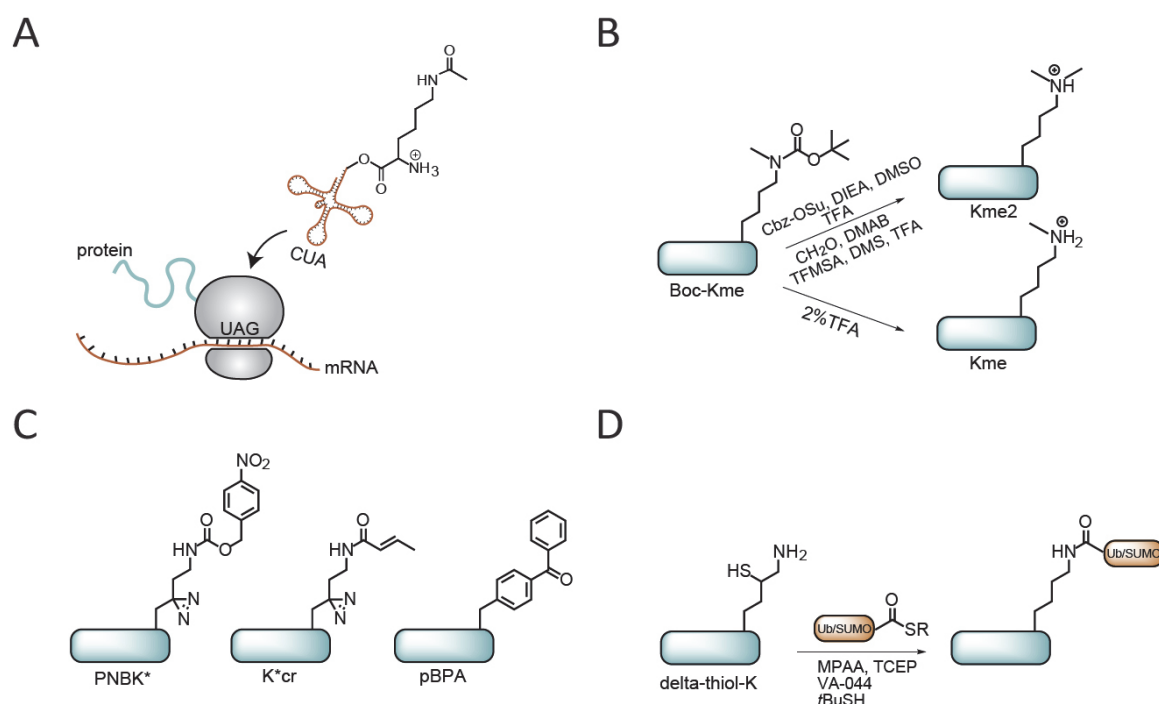
Figure 16: Amber codon suppression and PTMs. A) Scheme of the site-specific incorporation of Uaas (i.e. acetyl lysine) by amber codon suppression. **B**) Precursors of PTMs can be introduced into proteins and subsequently converted into the desired PTM (i.e. methyllysines). **C**) Structure of p-nitrobenzyl-caged lysine (PNBK*), photoaffinity analogue of crotonyl lysine (K*cr) and the UV-inducible cross-linker amino acid p-benzoyl-L-phenylalanine (pBPA). **D**) δ-thiol containing lysines for traceless ubiquitylation.

1.3.2 Cysteine-directed modifications

Cysteine-directed chemistry has been widely applied to the synthesis of histones containing PTM mimics or other tags. The sequence of the protein of interest is strategically mutated to recombinantly incorporate a cysteine residue at the desired position. The sulfhydryl group unique nucleophilicity is then exploited to perform site specific cysteine alkylation with electrophilic cognate partners.[191] A key requirement for such a strategy is the absence of other cysteines in the native sequence of the protein that needs to be modified. While H2A, H2B and H4 meet this prerequisite, H3 presents a cysteine at position 110; yet mutation of this residue to alanine does not compromise neither the structure nor nucleosome function. Such procedure allowed the synthesis of methylated[192] and acetylated[193] histone mimics providing insights into nucleosome structural features,[194] crosstalk between PTMs[195] and chromatin effector-PTMs interactions (**Figure 17**).[196] Alternatively, Davis and coworkers reported the oxidative elimination of cysteine to dehydroalanine (Dha) mediated by O-mesitylenesulfonyl hydroxylamine (**Figure 17**).[197] Following addition of a thiol-containing molecule, the PTM mimic was efficiently installed at the desired position. Furthermore, the combination of amber codon suppression technology and specific conversion to Dha of the unnatural amino acid phenylselenocysteine (PhSeCys) allowed Guo et al. to selectively modify proteins bearing multiple cysteines in their sequence.[198] Similarly, a pyrrolysine incorporation machinery was evolved by Liu et al. for encoding the Nε-Cbz-lysine that, after mild oxidation to Dha, was converted to analogs of methyllysine, acetylated lysines and phosphocysteines.[199]
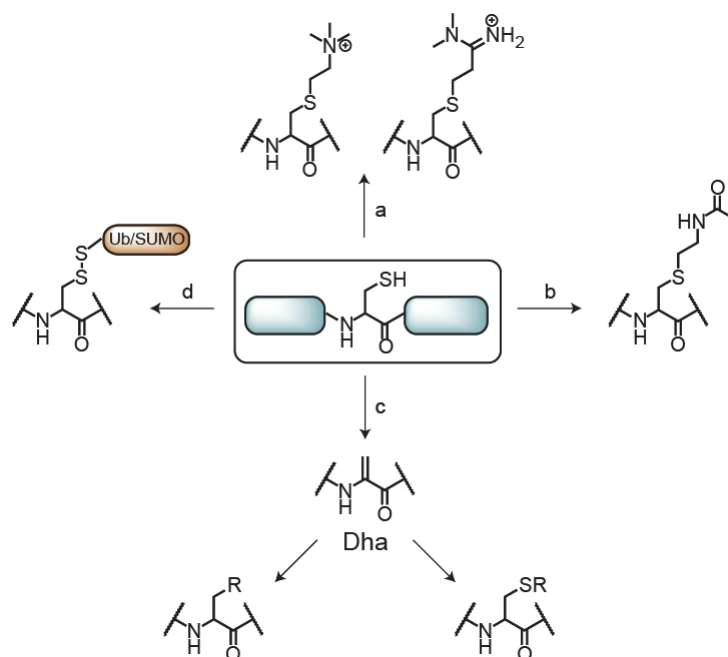
Figure 17: Cysteine analogs of PTMs. a. Trimethyl lysine and dimethyl arginine. b. Acetyl lysine. c. Dehydroalanine conversion of cysteine allows the synthesis of a variety of PTMs. d. Reversible ubiquitylation/SUMOylation by installation of cysteine S-S linked ubiquitin/SUMO.

Finally, cysteine was used as a handle for thiol-disulfide exchange, thus producing reversible ubiquitylation and sumoylation modifications (**Figure 17**).[200,201,202] It is important to underline that the produced modification is not identical to the native PTM, as the reaction introduces a sulfur atom at the gamma position of the side chain residue. Although these analogues have proved not to alter the biological activity of the corresponding native PTM in most cases,[192,203,201] the binding affinity for their partners might be altered.[204] Moreover, this approach cannot mimic certain modifications (i.e. phosphorylated serine and threonine) and the incorporation of multiple PTMs is challenging.

Other than histones and PTMs, the cysteine residue is the most widely employed amino acid for protein labeling and protein-small molecule bioconjugation. As mentioned above, due to the high propensity of the thiol group to form the thiolate ion under physiological condition with consequent high nucleophilicity, their low abundance in nature, and the easiness in introducing them in proteins by mutagenesis, cysteines are the perfect substrate for selective bioorthogonal reactions.

One of the earliest chemistries used for cysteine modification was the substitution reaction with haloalkyl reagents, such as iodoacetamides (**Figure 18A**).[191] Although fast and efficient, this labeling method lacks selectivity, as lysine and histidine residues can crossreact.[205] Indeed, maleimide-thiol coupling has become the preferred method among biochemists, as the reaction features high kinetics, is very selective towards cysteine residues and many maleimide-derivatives are commercially available. Maleimides are Michael acceptors that upon reaction with a thiol group form a thiosuccinimide bond (**Figure 18B**).[206] Even though succinimide thioethers are generally considered stable moieties, Retro-Michael addition can occur, leading

to loss of label and yielding a free maleimide in solution that can further react with off-target thiol molecules.[207] In addition, both the reagent and conjugate are prone to hydrolysis with concomitant ring opening and formation of a stable product.[208] While hydrolysis preserves the conjugate and prevents maleimide release, it also renders the maleimide unreactive (**Figure 18B**).
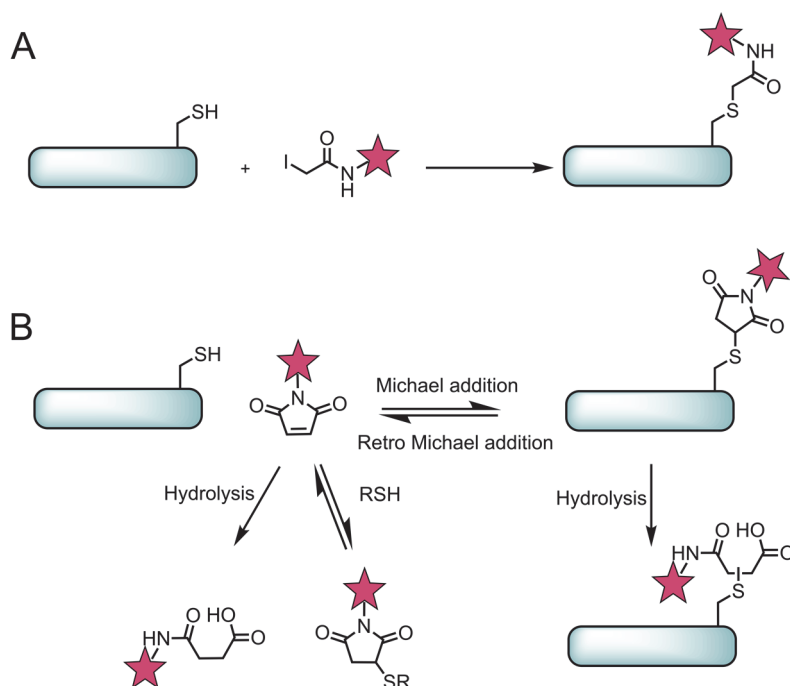
Figure 18: Most widespread cysteine labeling methods. A) Scheme of the cysteine labeling reaction using iodoacetamide reagents. **B**) Maleimide reagents react with cysteine residues by reversible Michael addition. Maleimide hydrolysis and the reaction with external thiols compete with the labeling reaction. Hydrolysis of the product irreversibly leads to ring opening.

These drawbacks prompted the development of a range of metal-free and metal-catalyzed new cysteine-labeling strategies. Examples are the thiol-ene[209] or thiol-yne[210] reactions, the conversion to dehydroalanine,[197] the use of C-substituted maleimides,[211,212] the cysteine arylation reaction by perfluoroaromatic molecules,[213] palladium organometallic reagents[214,215] or gold,[216,217] and the use of phosphonamidate electrophiles.[218] Despite these advances, the development of new highly efficient and chemoselective methods that rely on the use of easily accessible reagents that do not require challenging preparations and careful storage is in high demand. In **Chapter 4**, I will explain in detail the use of hypervalent iodine reagents for the specific labeling of cysteine-containing peptides and protein complexes, such as histone octamers.[219,220,221,222]

1.3.3 Solid phase peptide synthesis

Peptide synthesis refers to the sequential amide bond formation between the carboxylic group of a first amino acid and the amino group of a second amino acid. At the beginning of the 20th century, Fisher published the first synthesis of a peptide in solution.[223] The technique, however, was low-yield and time-consuming as the intermediate peptide needed to be purified and characterized after each coupling step

and suffered from poor solubility. In 1963, Bruce Merrifield introduced the concept of peptide synthesis on solid phase:[224] A N-α protected amino acid was linked through its C-terminus to an inert and insoluble polymeric resin, and sequential amino acid coupling steps allowed the peptide to grow on the resin from the C-to-N direction (**Figure 19**). The implementation of a solid support gave the system remarkable advantages: the reactions could be forced to completion by employing excess of reagents, since the intermediate peptide could be isolated from by-products and excess reagents by simple washing steps. This reduced the loss of material during the purification steps and dramatically improved the yield. Although extensively reviewed and optimized, this technique set the basis for the most commonly applied peptide synthetic strategies nowadays.
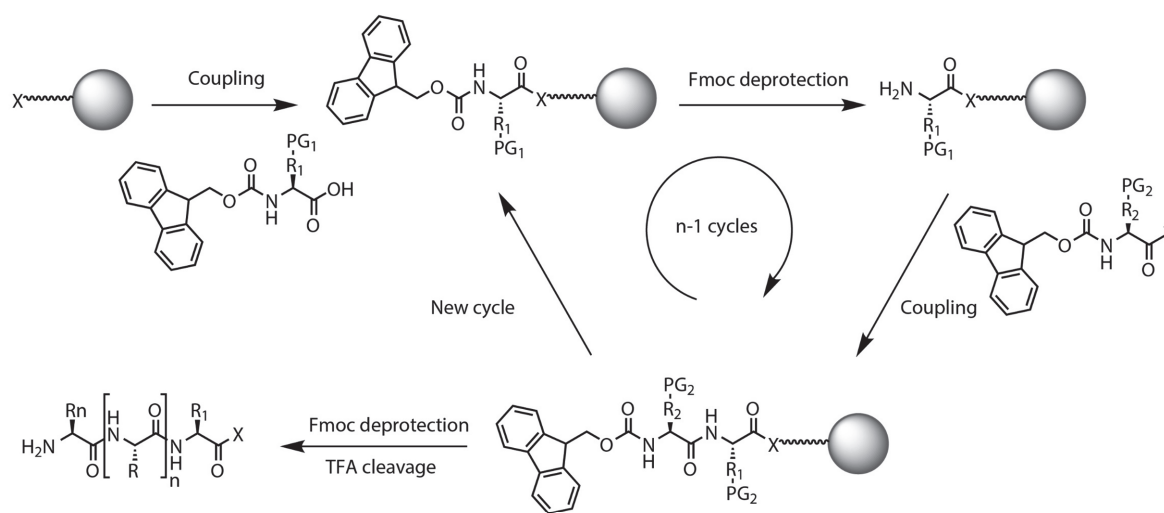
Figure 19: Cycles of Fmoc-SPPS. At the beginning of each cycle, Fmoc deprotection is performed by treatment with piperidine. Subsequently, the incoming Fmoc-protected amino acid is activated at the C-terminus and coupled to the growing peptide. At the end of the synthesis, Fmoc deprotection is performed by treatment with piperidine and the peptide is cleaved from the resin using TFA. X: resin linker. PG: protecting group. Y: activating reagent. Grey sphere: resin.

In general, SPPS is a procedure that involves a stepwise incorporation of N-α-protected amino acids into a peptide attached to a resin via a cycle of N-terminal deprotection and C-terminal activation reactions, separated by washes (**Figure 19**). The most widely used solid supports are synthetic polymers bearing functional reactive groups that can temporarily link the first amino acid and thus defines the peptide C-terminus upon cleavage from the resin. Examples are rink amide,[225] 2-chloro trityl hydrazine[226,227] or Wang resins[228] that yield an amide, a hydrazide or a carboxylic acid, respectively. Importantly, the linker needs to remain intact during the whole synthetic procedure and thus be resistant to all SPPS conditions. The amino acid C-terminus of the incoming amino acid is activated with a coupling reagent that renders the hydroxyl of the carboxylic acid a better leaving group upon nucleophilic substitution reactions. Indeed, this favors the nucleophilic attack by the amino group of another residue over the acid-base reaction that would yield an unreactive ammonium salt. A wide range of coupling reagents have been reported and include carbodiimides (such as DCC, DIC[229,230] in duet with the additives HOBt or HOAt [231,232]) phosphonium salts (BOP and their derivatives[233,234,235]) and uronium salts (among others, HBTU,[236] HCTU,[237] HATU,[238]

COMU[239]). The choice of the most appropriate coupling reagent aims at maximizing the coupling rate while minimizing racemization.[240]

Orthogonal protecting schemes usually need to be adopted to prevent undesirable side reactions on the various side chains and avoid multiple amino acids of the same type to be coupled during the same cycle. While side chain protecting groups are defined as "permanent", N-terminal protection is temporary as it must be removed at each coupling step. Importantly, both cleavage from the resin and side chain residues deprotection procedures depend on the N-α protection scheme. The two main N-α protecting groups used today are tert-butyloxycarbonyl (Boc) and fluorenylmethyloxycarbonyl (Fmoc). Boc is an acid labile moiety that is removed by trifluoroacetic acid (TFA) and the final peptide cleavage from the resin, along with the removal of permanent side chain protecting groups, is achieved by treatment with hard acids, such as anhydrous hydrofluoric acid (HF). Beyond the safety concerns arising from HF toxicity and the special equipment needed to handle it, repetitive TFA acidolysis and the harsh cleavage conditions make the Boc strategy incompatible with acid sensitive peptide modifications (i.e. phosphorylation, glycosylation). Fmoc, on the other hand, is a base labile protecting group that can easily be removed under mild conditions, by treatment with organic bases such as piperidine.[241] This allows acid labile protecting groups that are stable under basic conditions, such as benzyl and Boc protecting groups, to be used on the side-chains of amino acid residues. The peptide can then be cleaved from the resin using TFA, which will also remove the permanent protecting group, leading to the full-length deprotected peptide in solution. Importantly, the absence of corrosive acids during the repetitive synthetic process allows the automatization of deprotection, coupling and washes with peptide synthesizers.

The concept of orthogonal protecting groups can be extended and exploited for the synthesis of complex or modified peptides.[242,243] In fact, identical residues can be protected at their side chains with different functional groups that are removed by differing chemical mechanisms. Thus, one side chain can be deprotected specifically for further functionalization while leaving the other residue protected. One example is the derivatization of lysines. While Boc is the protecting group of choice for lysines in routine Fmoc-SPPS, allyloxycarbonyl (Alloc) group is resistant to acidic conditions and is removed by nucleophiles in the presence of Pd.[244] Such a strategy allows the on-resin functionalization of specific lysine side chains and in particular the synthesis of peptides containing isopeptide bonds. Orthogonal protection schemes for the highly reactive thiol function of cysteines were developed as well. Acm and Thz represent alternative acid-resistant protecting groups with respect to the canonical acid-labile trityl (Trt).[245] Such schemes can be employed, for instance, in the specific formation of multiple intra- or intermolecular disulfide bridges.[246,247]

Although automated peptide synthesis, together with the implementation of technical and chemical advances aiming at improving the overall yield and reduce peptide aggregation on the resin (i.e. double couplings, high-swelling and low loading resins, pseudoprolines[248]) allows to synthesize complex peptides up to 50 amino acids in a reasonable time and purity, the synthesis of longer peptides remains challenging.

Importantly, most biologically relevant protein domains are longer than this. To overcome this limitation, ligation strategies that allow the assembly of polypeptide chains were developed, and will be described in detail in the next section.

1.3.4 Native chemical ligation

Many different ligation strategies have been continuously developed and optimized over the last few years. Examples include the cysteine-aziridine ligation,[249] traceless Staudinger ligation,[250] the KAHA ligation,[251] Serine/Threonin ligation (STL),[252] the oxime ligation,[253] KAT ligation,[254] imine ligation[255], Diels-Alder ligation[256] and sortase-mediated ligation[257] among others. However, the most widespread strategy is by far the native chemical ligation (NCL). The NCL strategy was introduced by Kent et al. in 1994 and describes a highly efficient and robust chemoselective ligation reaction between two unprotected peptides or protein segments under mild conditions in aqueous solution.[258] One peptide fragment is functionalized with a C-terminal thioester, while the second peptide bears a N-terminal cysteine residue. The ligation proceeds through an initial *trans*-thioesterification step, where the thioester carbonyl carbon undergoes nucleophilic attack by the cysteine thiol. Rapid and irreversible rearrangement of this intermediate *via* S-to-N acyl shift through a 5-membered ring transition state then yields a native amide bond at the junction site and restores the original thiol side chain of the cysteine (**Figure 20**).
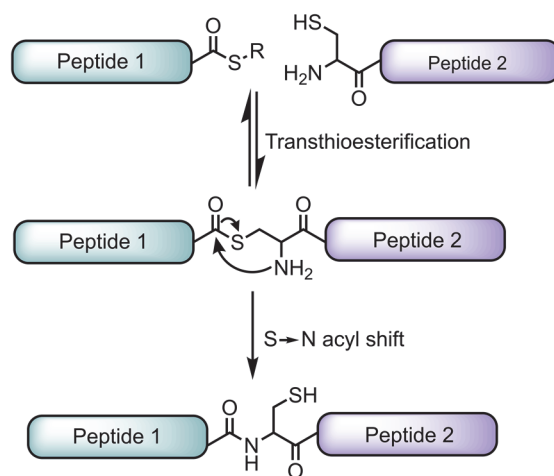
Figure 20: Schematic representation of native chemical ligation reaction. The native amide bond between a C-terminal thioester peptide and a peptide carrying a N-terminal cysteine is achieved through trans-thioesterification and S- to N-acyl shift**.**

Since the first report of this powerful and innovative reaction, hundreds of previously inaccessible proteins could be synthesized by means of this method. However, several drawbacks made the use of NCL not always straightforward: 1) α-thioesters were not readily accessible by Fmoc-SPPS. 2) Due to the rarity of the cysteine residue in the native sequence of most natural proteins, this methodology often left a "scar" in the assembled peptide. 3) The rate of the reaction was low, especially for sterically hindered C-terminal amino acid α-thioesters, such as valine, isoleucine or proline, which required considerably prolonged reaction times.[259] 4) Due to the fragment size limitation dictated by SPPS, and before the development of appropriate

orthogonal protection schemes at the C-terminal thioester and N-terminal cysteine, sequential ligations of more than two fragments were prevented by polymerization reactions. Thus, only relatively short proteins were accessible. Indeed, enormous efforts have been made aiming at improving the accessibility, efficiency, speed and applicability of NCL. The next sections will give and overview of these recent advances.

- *Thioester formation*

The preparation of α-thioesters by Fmoc-SPPS is non-trivial. Thioesters are sensitive compounds that undergo hydrolysis in basic environments and do not survive the multiple deprotection cycles mediated by piperidine in the Fmoc-SPPS schemes.[260] Conversely, they are highly resistant towards strong acids; thus an α-thioester linkage between peptide and resin support is totally compatible with Boc-SPPS (i.e. 3-mercaptopropionic acid[261]). However, as mentioned above, the use of HF in the final deprotection step of a Boc-SPPS restricts its usage to only specialized laboratories and does not allow to deal with labile functionalities, such as phosphorylated and glycosylated residues. These considerations prompted efforts towards the development of Fmoc-SPPS-compatible strategies for the generation of peptide α-thioesters. Most of these approaches involve the use of a thioester "surrogate", that is, a non-thioester linker that is stable to the whole set of reagents used in Fmoc-SPPS. Upon completion of the synthesis or cleavage from the resin, the peptide C-terminus is chemically converted to thioester.
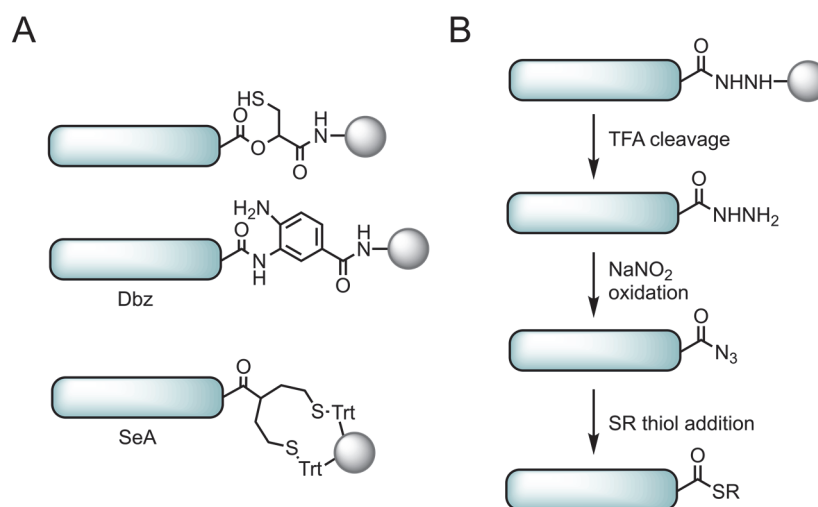
Figure 21: Crypto thioesters. A) From top to bottom: oxy-(2-mercapto-1-carboxyamide)-ethyl linker, diaminobenzoic acid (Dbz) linker and bis(2-sulfanylethyl)amido (SEA) linker. **B**) Hydrazide mediated thioester formation.

In 2004, Botti et al. reported a oxy-(2-mercapto-1-carboxyamide)-ethyl linker (**Figure 21A, top**) that generated an α-thioester during ligation by intramolecular O to S acyl shift .[262] Another successful example is diaminobenzoic acid (Dbz, **Figure 21A, middle**). This linker is converted on resin to an aromatic N-acylurea (Nbz) moiety that is stable to acidic cleavage conditions but rapidly undergoes thiolysis in aqueous buffers, enabling thioester formation either prior to purification or *in situ* during ligation.[263] Furthermore, tunable auxiliaries such as bis(2-sulfanylethyl)amido (SEA, **Figure 21A, bottom**) can be incorporated during SPPS.[264] Addition of tris(2-carboxyethyl)phosphine (TCEP) reduces the disulfide bond present on

the bisthiol, triggering N-to S acyl shift and generating the desired thioester under tight control. Complementary bis(2-selenylethyl)amino (SeEA) auxiliaries were also developed.[265] However, due to the ease of use and excellent efficiency, the hydrazine linker has become the method of choice for preparing peptide crypto-thioesters by Fmoc-SPPS (**Figure 21B**).[266] Upon cleavage from the resin by TFA treatment, the peptide C-terminus is converted to hydrazide. Subsequent oxidation by nitrite produces an acyl azide moiety that can be further reacted with a suitable thiol to generate a thioester. Importantly, the ability to perform the conversion to thioester *in situ* enables the sequential ligation of peptide-hydrazide fragments by NCL in the N-to-C direction.[267]

- *Improving the rate of NCL*

The rate limiting step of the ligation reaction is the thiol-thioester exchange.[268] Alkyl thioesters were often used in ligation chemistry, as they could be easily prepared by SPPS, yet their reactivity is quite poor. Instead, a more reactive peptide thioester can be generated through the addition of an exogenous aryl thiol. In fact, aryl thiols can rapidly exchange with alkyl thioesters and their lower pKa makes them excellent leaving groups upon reaction with a N-terminal cysteine and thus they considerably enhance the rate of the ligation.[269] Johnson and Kent in 2006 reported that the aryl thiol additive mercaptophenylacetic acid (MPAA)[270] improved the NCL kinetics by one order of magnitude compared to the commonly employed mercaptoethanethiolate sodium salt (MESNa).[271] On the other hand, selenium moieties have gained increasing attention in the field of ligation chemistry. The stronger nucleophilicity of selenolates with respect to thiolates, together with the low pKa of selenols,[272] makes selenocysteine a great candidate for NCL. Indeed, it was shown that replacement of cysteine with the selenium counterpart significantly boosted the reaction rate, especially at low pH.[273,274] Further, selenoesters were found to be better acyl donors than their sulfur analogues and to accelerate the ligation at sterically hindered C-terminal sites.[275] Finally, flow chemistry systems have been recently exploited to drive to completion ligations in substantially shorter times than batch reactions, down to a few minutes.[276]

- *Coupling of ligation and desulfurization*

Cysteine is a low abundant amino acid in nature and the requirement for a N-terminal cysteine at the ligation site hindered the applicability of NCL. The most powerful synthetic advancement to overcome such a need was made by Dawson et al. in 2001, when they reported the coupling of NCL with desulfurization.[277] Pd/Al$_2$O$_3$ or Raney nickel-mediated desulfurization of cysteine to yield an alanine residue at the junction allowed ligations at the more naturally abundant X-Ala sites and thus dramatically expanded the scope of NCL. Inspired by this pioneering work, numerous groups have thereafter extended the strategy to several cysteine surrogates, introducing mercaptoamino acids as N-terminal residues. By now, 15 out of the 20 natural amino acids can be employed in NCL.[278] However, the synthesis of such derivatives involves multistep strategies and only a few precursors are currently commercially available. In addition, the poor recovery of the desulfurized product, the risk for secondary alcohol epimerization and side reactions on

thiols, thioethers and thioesters[279,280] prompted the investigations of new desulfurization strategies. Based on previous works on desulfurization reactions between mercaptan and trialkylphosphite derivatives,[281,282,283] in 2007 Wan and coworkers established a mild and highly versatile metal-free radical-based desulfurization method (**Figure 22**).[284] In an initial step, the radical initiator 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) generates an alkylthiyl radical at the cysteine side chain, which subsequently undergoes reversible addition to the chain propagator TCEP to form a phosphoranyl radical intermediate. Homolytic cleavage of the C-S bond then provides an alkyl radical that rapidly abstracts a hydrogen from the thiol moiety of another cysteine. The radical chain reaction is thus propagated until a hydrogen donor, such as the thiol additive tBuSH, provides a hydrogen to terminate the chain (**Figure 22A, B**). With yields above 80% and its inertness towards sensitive groups, such as methionine, thioesters and thiazolidine-protected cysteines, metal-free desulfurization rapidly became the method of choice.
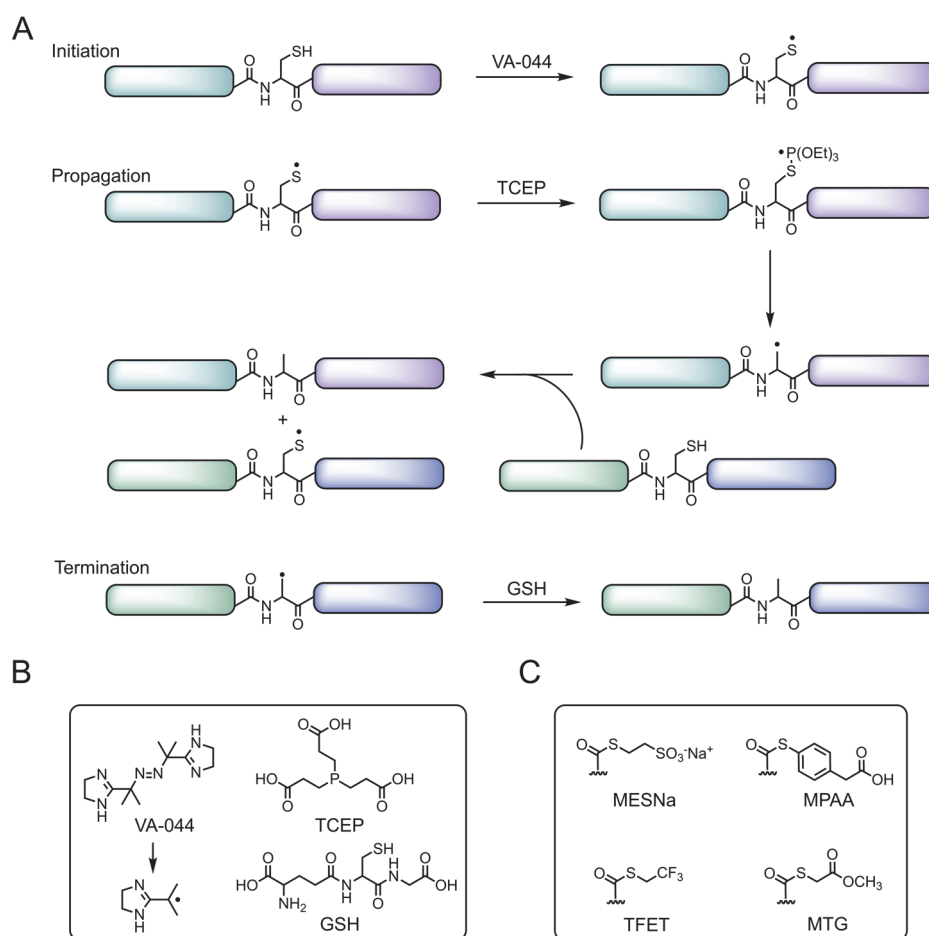
Figure 22: Radical desulfurization. A) Reaction mechanism, including initiation, propagation and termination steps. **B**) Radical desulfurization reagents. **C**) Peptide thioesters obtained with different thiol additives.

The need for purification of intermediates lowers the overall yield of any chemical reaction. Unfortunately, the above-mentioned desulfurization strategy is not compatible with the aryl thiol additives widely employed in NCL (such as MPAA), as they act as radical scavengers and quench the radical propagation.

This prevents ligation and desulfurization reactions from being performed in one-pot. Although thiol additives such as MESNa were shown not to interfere with the desulfurization step, the low reactivity of the MESNa peptide thioesters significantly lowered the kinetics of the ligation reaction.[285] An advance in this area was provided by Thompson et al. in 2014, as they identified a novel thiol additive, trifluoroethanedithiol (TFET), that provides NCL kinetics comparable to MPAA, while not interfering with the *in situ* desulfurization reaction.[286] In addition, a less volatile and malodorous thiol additive methylthio glycolate (MTG) was reported by Huang et al. (**Figure 22C**).[287] Although the kinetics are slightly slower than MPAA (1.8 fold), MTG peptide thioesters have remarkable higher resistance towards hydrolysis than TFET thioesters, which allows to reduce the thioester excess in the reaction mixture and makes the transformation less pH sensitive.

- *Protein semisynthesis and EPL*

The limitation of NCL to the synthesis of relatively small proteins has encouraged the combination of the NCL strategy to recombinant protein production. This powerful approach combines the strength of both chemical synthesis and recombinant DNA technology: the ability to access (almost) any modified peptide fragments and the capability of producing proteins of any sizes. In case the modification is placed at the N-terminus of the protein of interest (POI), a synthetically produced peptide thioester is ligated to the N-terminal cysteine of the recombinantly expressed protein core; conversely, a chemically synthesized C-terminal peptide fragment can be ligated through its N-terminal cysteine to a recombinantly produced protein α-thioester to install a modification at the POI's C-terminus (**Figure 23**).

The installation of a N-terminal cysteine by recombinant protein expression is non-trivial. In fact, the requirement for an initiating methionine codon prevents the encoding of cysteine as the N-terminal residue. However, since the first methionine is often co-translationally cleaved *in vivo*, the requisite cysteine can be installed next to the initiating methionine. Although examples of histone proteins synthesized by this strategy were reported,[288,289] the Met cleavage is not always complete and secondary modification of the free N-terminal cysteine can result from reaction with cellular metabolites.[290] Fusion protein strategies circumvent these issues by directly linking the protein N-terminus to a protease-cleavable domain that also serves as affinity purification tag. Upon cleavage, the fusion tag is released affording a mature N-terminal cysteine-containing fragment that can undergo NCL. Importantly, the protease has to be permissive to a cysteine residue at the S1 position in the substrate. Factor Xa protease,[291] SUMO protease,[292] and tobacco etch virus (TEV) protease[293] are examples of such enzymes and have been widely employed in NCL-mediated protein semisyntheses. (**Figure 23A**) On the other hand, the generation of a C-terminal thioester from recombinant protein expression is also not straightforward, and takes advantage of protein fusion strategies, too. First introduced in 1998 by the Muir and Xu groups independently, the expressed protein ligation (EPL)[294] (or intein-mediated protein ligation (IPL)[295]) relies on the fusion of the protein fragment to a special domain called intein. Inteins are autoprocessing protein

fragments that are naturally self-excised from a protein precursor *via* a process known as protein splicing, where two amide bonds are cleaved and the flanking external proteins (exteins) are ligated by formation of a new peptide bond (**Figure 23B**).[296]
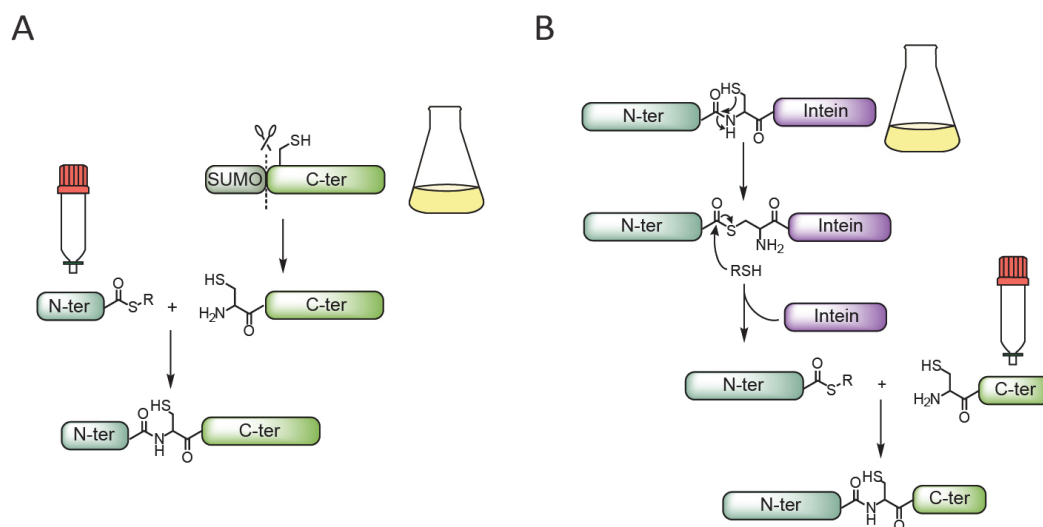
Figure 23: Protein semisynthesis. A) A synthetically produced peptide thioester is ligated to the N-terminal cysteine of recombinantly expressed protein core. The recombinant fragment is produced by expression as a fusion protein followed by cleavage of the tag to reveal the free N-terminal cysteine. **B**) Expressed protein ligation. A C-terminal recombinant thioester is generated by thiolysis of the N-terminal fragment fused to a mutated intein and ligated to a synthetic C-terminal fragment

Mutations at the C-terminal conserved asparagine in the intein sequence prevent the splicing event, and the initially formed thioester can undergo thiolysis when exposed to small thiol molecules. [297] Indeed, following expression and purification, the protein of interest is cleaved from the mutated intein domain by treatment with thiols, to afford the α-thioester derivative, which can then undergo ligation with the synthetic cysteine-containing fragment. Although widely used, many common inteins have long splicing half-time (hours to days) with consequent slow thiolysis rates.[298] The discovery of ultrafast inteins, with splicing times in the range of seconds,[299] has tremendously enhanced the potential of these protein domains in the protein chemistry field. In addition, these inteins are naturally split: the two inactive and disordered segments have no thiolysis activity, yet when they spontaneously assemble (with nanomolar affinity) and refold they trigger *trans*-splicing, thus providing a means to control the thioester generation and avoid premature cleavage of the fusion protein.

- *Sequential and kinetically controlled NCL*

Although the use of NCL for the preparation of proteins from only two peptide segments is rather straightforward, for syntheses involving multiple peptide segments the dual reactivity of the middle fragments Cys-peptide-thioester must be carefully controlled in order to avoid oligomerization or cyclization side reactions. In the case of an assembly in the C-to-N direction, the middle fragment N-terminal cysteine must be masked with a protecting group that can be removed chemoselectively (i.e. Acm and Thz) (**Figure 24A**).[300] The opposite case, the N-to-C direction, is more challenging due to the difficulty

in directly protecting thioester groups. However, the differential reactivity between aryl and alkyl thiols can be exploited, resulting in a kinetically controlled ligation (**Figure 24B**).[301] Otherwise, a convergent synthesis only relying on peptide hydrazides represents a valid alternative, since the hydrazide is an unreactive moiety unless activated by $NaNO_2$ (**Figure 24C**).[302]
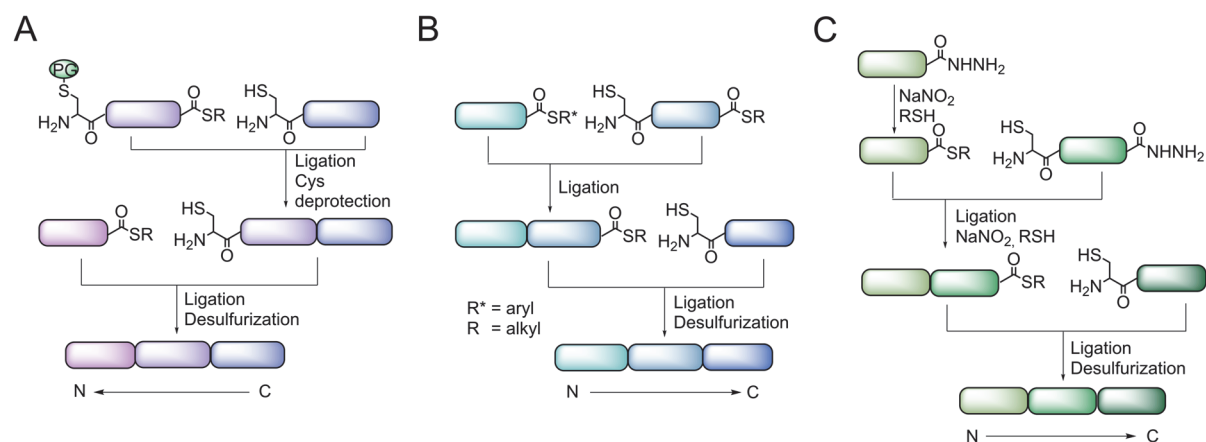
Figure 24: Strategies for multi-segments ligations. A) C- to N-direction. The N-terminal cysteine of the middle fragment is temporarily protected. PG = protecting group. **B**) N-to C-direction. Kinetically controlled ligation (KCL). Ligations involving aryl thiols are kinetically favored over alkyl thiols. **C**) N-to C-direction. Peptide hydrazides at the middle fragment C-terminus are unreactive unless activated.

1.3.5   Semisynthesis of chromatin by native chemical ligation

The ensemble of these technologies and methodologies can be applied to the semisynthesis of specifically modified nucleosomes. In fact, histone proteins are ideal candidates for semisynthetic strategies: most of their PTM are found either at the N- or C- extensions and their ability to refold from a denaturing environment allows the use of chaotropes in the reaction mixture, which in turn improves the solubility of the peptide reagents and the kinetics of the reaction. In general, the N-terminal histone tail peptide, carrying the PTMs of interest at the desired positions and a thioester surrogate at the C-terminus is synthesized by SPPS, ligated to the N-terminal cysteine of a recombinant truncated core histone *via* EPL and desulfurized at the ligation site. Subsequently, refolding and concomitant assembly of semisynthetic and recombinant core histones generate core histone octamers. Octamers are then combined with a DNA segment containing the high-affinity 601 nucleosome positioning sequence[303] (or any other suitable sequences) to reconstitute intact nucleosomes by a gradual dialysis from high to low salt concentration (**Figure 25**).

These strategies now allow us and other groups to produce a wide variety of symmetric nucleosomes carrying single or combinations of PTMs. Examples are the preparation of specifically ubiquitylated H2A and H2B (H2AK199ub, H2BK120ub, H2BK34ub)[304,305,306] the synthesis of multiply acetylated H3 and H4 N-terminal tail histones,[288] the generation of acetylated H4K16ac.[307] In addition, we have recently developed a synthetic strategy for asymmetric nucleosomes reconstitution,[104,308,309] which will be discussed in detail later in **Chapter 2** and **Chapter 3**.
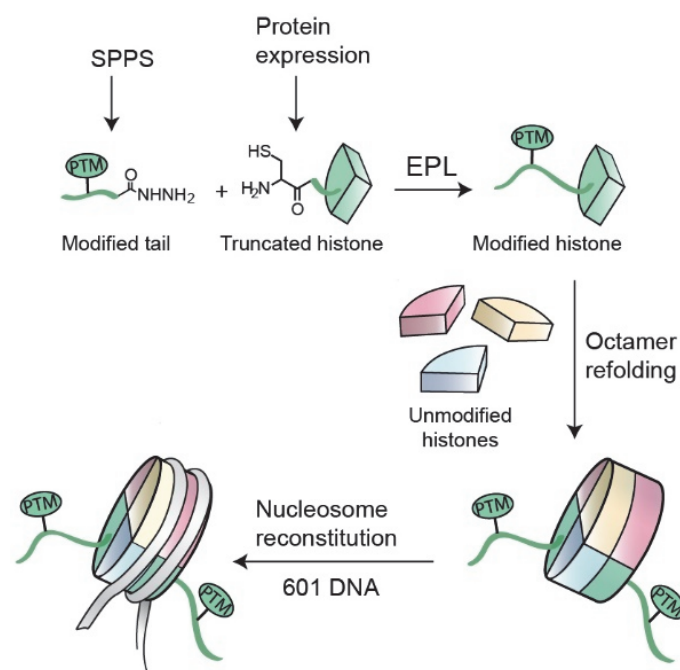
Figure 25: Semisynthesis of modified nucleosomes. Scheme of the process used to obtain nucleosomes, that involves: semisynthesis of H3 modified histones, production of recombinant core histones, octamer refolding and nucleosome reconstitution.

1.4   Aims of the project

Chromatin effector proteins regulate cell signaling pathways, such as cell growth and differentiation, and result in disease such as cancer when dysregulated. Effectors are recruited to histone PTM patterns through combinations of reader domains. These multivalent interactions are highly dynamic and sensitive for the spatial arrangement of histone PTMs. In addition, the detection of these dynamic interactions in live cells is fundamental for understanding their role in the regulation of these cellular processes. Chromatin thus represents a critical signaling node and the mechanisms underlying its PTMs readout as well as its dynamic organization in living cells are of great interest, although currently poorly understood.

Recent studies have revealed that nucleosomes can be modified in an asymmetric fashion, carrying two differentially modified variants of a histone. In particular, in stem cells, cancer cells and fibroblasts, nucleosomes are asymmetrically modified for H3K4me3, H3K36me3, H3K27me3, as well as H4K20me1 in genomic regions termed "bivalent domains".[102] Such asymmetric nucleosomes might present distinct binding sites for multivalent effectors. Moreover, growing evidence suggests that the formation of specific chromatin states (i.e. the repressive heterochromatin), involves liquid-liquid phase separation and is mediated by multivalent effector - chromatin interactions.[158,157]

Chemical biology methods are critical for the understanding of chromatin function and organization and recent work in the Fierz laboratory had opened new avenues for such investigations.

37

First, preparing defined nucleosomes and chromatin fibers carrying specific and multiple PTMs at defined sites is non-trivial. Recently, our laboratory developed a new procedure to control the supramolecular assembly of nucleosomes.[104] However, the approach was only applied to one set of histone PTMs in one histone protein type.

Second, an *in vitro* single-molecule total internal fluorescence microscopy (smTIRFM) platform for the quantitative measurement of the interaction dynamics between effector proteins and chemically defined chromatin fibers was established.[310] However, for such biophysical studies of effector – chromatin interactions, the binding partners need to be site-specifically labeled with highly bright fluorophores with low photobleaching rates, which remains a challenging task.

Finally, our group had generated genetically encoded probes to detect a specific bivalent chromatin state in living cells,[97] yet general methods to image and investigate LLPS-dependent chromatin arrangements in live cells are lacking.

Therefore, in order to untangle the multiple molecular processes that result in the interpretation of the histone code by effector proteins, new chemical approaches and reliable biophysical methodologies are required. During my thesis, I thus developed novel protein chemistry and chemical biology tools to shed light upon the mechanisms defining the dynamic interpretation of histone PTMs both *in vitro* and *in vivo*.

The specific aims of this PhD project were the following:

- **Extend the strategy for the synthesis of asymmetrically modified H3 nucleosomes to other combinations of PTMs on H3 and H4 (Chapter 2 and Chapter 3)**
  Control over the chromatin state is an indispensable prerequisite for studying the molecular mechanisms that regulate effectors-chromatin interactions. The recently developed crosslinking strategy was applied to the generation of H3 nucleosomes carrying the bivalent mark (H3K27me3 and H3K4me3) on opposite tails. Here, I developed methods for the synthesis of nucleosomes asymmetrically modified on H4, in particular for methylated H4K20, key PTM for bivalency and DNA damage repair. Moreover, I synthesized asymmetric nucleosomes with H3K27me3 and H3K36me3. Together these complexes constitute all asymmetric nucleosomes found in ES cells and were used for mechanistic investigations of stem-cell regulation.

- **Implement a chemoselective, "doubly orthogonal" cysteine tagging strategy, using hypervalent iodine reagents, for the site-specific labeling of peptides and proteins (Chapter 4)**
  To gain insights into gene regulation and for the biophysical study of chromatin binding dynamics and structural organization, selective chromatin fluorescent labeling techniques are in high demand. Here, I employed ethynylbenziodoxolone (EBX) reagents to specifically install a dual

handle at cysteine residues on histone octamers, and showed that our novel labeling strategy enables the enhancement of cyanine dyes photostability.

- **Develop a peptide-based LLPS probe for live cell dynamic chromatin visualization (Chapter 5)**
  Chromatin is organized at multiple scales in live cells and appears to possess liquid droplet-like features. The formation of such membraneless biomolecular compartments might help the cells to orchestrate in space and time the multiple chemical reactions that occur inside the nucleus. Here, I designed and synthesized modular peptide probes that accumulate at specific chromatin states and enable the dynamic *in vivo* detection of chromatin domains.

In conclusion, I believe that by combining our chemical and biophysical efforts, we have contributed to the understanding of the mechanisms underlying chromatin PTMs redout and provided the community with novel tools to further investigate chromatin-dependent processes.

Chapter 2: Semisynthesis and Reconstitution of Bivalent Asymmetric H3K27me3/ H3K36me3 Nucleosomes

2.1   Outline

In nature, individual histones in the same nucleosome can carry identical (symmetric) or different (asymmetric) PTM patterns, increasing the combinatorial complexity. Embryonic stem cells exhibit "bivalent" nucleosomes, some of which are marked by an asymmetric arrangement of H3K36me3 (an activating PTM) and H3K27me3 (a repressive PTM). Here I describe a modular semisynthetic method to access such asymmetrically modified nucleosomes and show that H3K36me3 inhibits the activity of the methyltransferase PRC2 locally while still prolonging its chromatin binding time.

2.2   Contributions

This chapter is based on the following publication:

Guidotti, N., Lechner, C. C., Bachmann, A. L., & Fierz, B. (**2019**). A Modular Ligation Strategy for Asymmetric Bivalent Nucleosomes Trimethylated at K36 and K27. *ChemBioChem*, 20(9), 1124-1128.

Contributions by other authors are described, where appropriate, along the text and in figure legends. In particular:
- G. N. performed most of the reactions on peptides, reconstituted the asymmetric nucleosomes and performed the methyltransferase assays.
- L. C. C. synthesized and ligated some peptide building blocks, tested the reaction conditions for hydrazide ligation in the presence of thiazolidine and conditions for ligation at C-terminal valine sites.
- B. A. L. reconstituted chromatin fibers and performed single-molecule experiments.

2.3   Introduction

Nucleosomes, which form the basic units of chromatin, are pseudo-symmetric particles containing two copies of each histone H3, H4, H2A and H2B. The histone proteins then form an octameric core around which 147 bp of DNA are wrapped.[1] In cells, histones carry complex combinations of PTMs which are associated with the regulation of gene transcription, DNA replication and repair.[311] As in each nucleosome histones are present in pairs, those can further carry the same (symmetric) or different (asymmetric) PTM patterns, increasing the combinatorial complexity of their readout.[102] Trimethylation of distinct lysine

residues, such as lysine 4, 27 and 36 in H3 (H3K4me3, H3K27me3 and H3K36me3) represents a particularly important set of PTMs involved in the regulation of the gene expression patterns.[312] While H3K4me3 and H3K36me3 are associated with active transcription, H3K27me3 is involved in gene silencing, mediated by PcG proteins.[88] PcGs are important chromatin effectors that contribute to the maintenance of transcriptional repression,[87] and are organized as large multimeric protein complexes with different functions and catalytic activities.

Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2) are the most studied members of the family. While PRC1 is a E3 ubiquitin ligase that transfers a mono-ubiquitin to H2AK119 (H2AK119ub) by means of its RING1A or RING1B subunits,[313] PRC2 is responsible for the mono-, di- and trimethylation of H3K27.[88] The core PRC2 subunits include the methyltransferase EZH2 (or its homolog EZH1), the zinc finger-containing SUZ12 and the protein EED; all are required to ensure the catalytic activity of the complex.[73,314,315,316,317,318,319] The products of PRC2 methylation show distinct genomic regions. The monomethylated H3K27 is predominantly found at active genes, H3K27me2 is the most abundant mark and is spread at inter- and intragenic regions and H3K27me3 is enriched at CpG island (CGIs) promoters at repressed genes.[320,321] Together with the histone-binding protein RBBP4/7, the PRC2 core complex associates with substoichiometric subunits that modulate PRC2 recruitment and activity. In addition, the regulation of PRC2 is critically dependent on pre-existing histone PTMs, as active modifications inhibit the methyltransferase activity, while H3K27me3 allosterically activates the enzyme.[102,322,104] In fact, the methylation activity of the EZH2 is enhanced when the H3K27me3 mark binds to an aromatic cage located in the EED subunit. Structural studies have revealed that the PRC2 complex can sit between two nucleosomes; the EED subunit interacts with H3K27me3 while the EZH2 subunit makes contacts with the neighboring nucleosome.[323] The conformational change arising from H3K27me3 binding is proposed to alter the EED-EZH2 interface, that in turn reaches the active site in the EZH2 SET domain.[324,325] This positive feedback loop between H3K27me3 recognition and deposition thus contributes to the enrichment and spreading of H3K27me3. On the other hand, active marks, such as H3K4me3 or H3K36me3 hamper PRC2 activity.[322,326] Collectively, histone PTMs seem to primarily influence PRC2 activity rather than binding, although these mechanisms are incompletely understood. This, together with the fact that genes encoding for PRC2 are often dysregulated in diseases such as cancer, makes the study of the mechanisms underlying its activity, function and regulation extremely intriguing.

Embryonic stem cells exhibit a specific chromatin state where active (H3K4me3 / H3K36me3) and repressive histone PTMs (H3K27me3) coexist on single nucleosomes at developmental genes, keeping their expression state in the balance.[92] Importantly, in such "bivalent" nucleosomes, the histone PTMs are organized in an asymmetric fashion, i.e. H3K4me3 or H3K36me3 on one H3 molecule, whereas H3K27me3 exists on the second H3 copy. Our group has previously developed a chemical approach to synthesize bivalent nucleosomes carrying H3K4me3/H3K27me3 in a traceless manner, which allowed the investigation of PRC2 regulation by these substrates in detail.[104] In contrast, access to nucleosomes bivalent

for H3K36me3/H3K27me3 is still lacking, as the PTM is located further within H3, which complicates its chemical synthesis. We thus set out to develop a strategy to synthesize asymmetric nucleosomes carrying the H3K36me3 modification and evaluate in detail the consequences of such asymmetry on PRC2 catalytic activity.

2.4 Results

2.4.1 Synthesis of asymmetric nucleosomes using the *lnc*-tag

Recent work in our group reported on the development of a traceless transient crosslinking strategy for the synthesis of H3 nucleosomes asymmetrically modified at K4 and K27.[104] The synthetic approach was based on the inclusion of a link-and-cut (*lnc*)-tag at the N-terminus of the H3 protein, and relied on the transient disulfide crosslinking of two differently modified H3 molecules (xlncH3), which forced them into the same nucleosome upon chromatin assembly (**Figure 26A**)
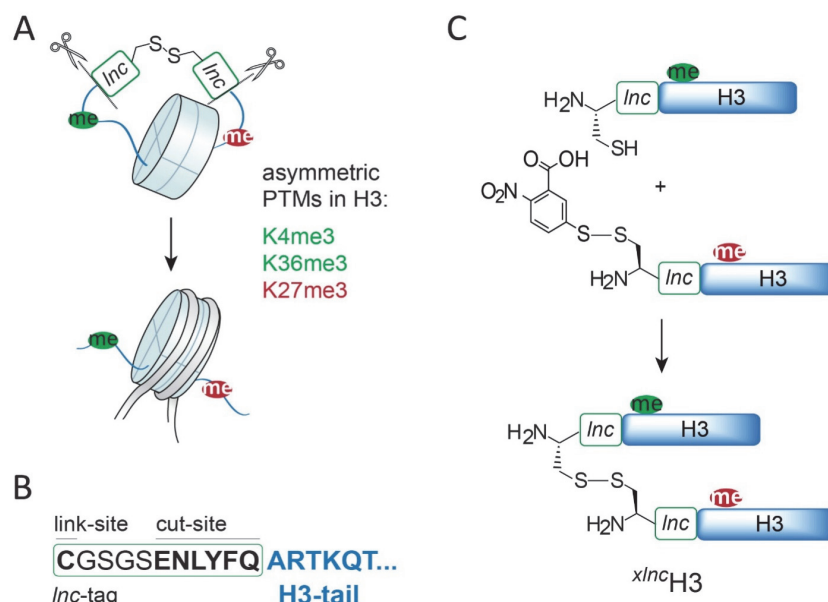
Figure 26: Previously reported synthetic strategy for nucleosomes asymmetrically modified on H3. A) Synthetic strategy for H3, following ref.[104] **B**) Sequence of *lnc*-tag at the H3-tail. **C**) xlncH3 formation, demonstrating the transient crosslinking strategy.

The N-terminal H3 tail containing the modification of interest was designed so as to have at its N-terminus an additional sequence called "*lnc*-tag" which carried an Acm protected N-terminal cysteine and a TEV protease site at its C-terminus (**Figure 26B**). Following EPL of the N-terminal tail to the recombinantly expressed H3 core, and subsequent desulfurization of the cysteine at the ligation site, both the Cys-protected *lnc*-tag and the desired PTMs were incorporated in the native H3 sequence. Afterwards, the *lnc*-tag cysteine was deprotected and, in a subset of H3 molecules (notably only the ones carrying the K27me3 modification), this was followed by activation with Ellman's reagent (5,5'-dithiobis(2-nitrobenzoic acid), DTNB) (**Figure 26C**). Following transient heterodisulfide crosslinking of H3 pairs, histone octamers

refolding and nucleosomes reconstitution, the TEV-cleavable tag ensured the recovery of scarless H3 asymmetric nucleosomes.

2.4.2 Asymmetric H3K36me3 nucleosomes: strategy and design

Here, I report the traceless and efficient synthesis of bivalent nucleosomes containing H3K36me3. Based on the synthesis of H3K4me3/H3K27me3 bivalent nucleosomes[104] and inspired by a three-piece ligation scheme developed for the synthesis of H3K36me3[327] and H3R42me2,[328] we designed a traceless synthetic strategy for asymmetric H3K36me3 nucleosomes (**Figure 27A**). Due to the position of the K36me3 mark, close to the globular domain of the H3 histone protein, we envisioned a three-segment ligation strategy in the C-to-N direction (**Figure 27B, C**).
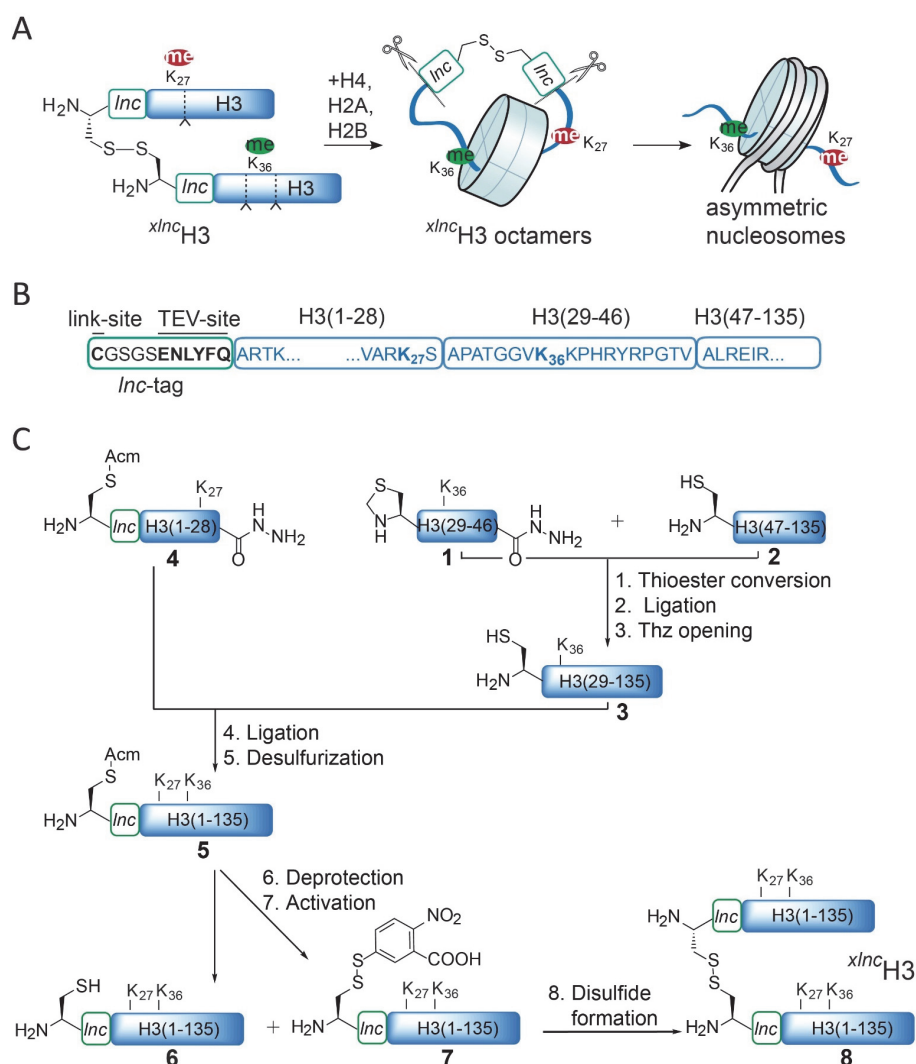
Figure 27: Traceless synthetic strategy for H3K27me3/H3K36me3 asymmetric nucleosomes. A) Schematic view of the assembly of asymmetric nucleosomes using *xlnc*H3 heterodisulfide dimers. **B**) *lnc*-tagged H3 sequence. **C**) Synthetic route for the semisynthesis of heterodisulfide dimers *xlnc*H3. 1. Conversion of middle peptide hydrazide **1** to MTG or MPAA thioester; 2. Ligation of MPAA thioester peptide **1**[SR'] to truncated H3Δ46 protein **2**; 3. *In situ* thiazolidine to cysteine conversion; 4. Ligation of *lnc*-tagged N-terminal peptide hydrazide **4** to truncated H3Δ28 protein **3**; 5. *In situ* desulfurization of Cys28 and Cys47; 6. Acm removal at the N-terminal Cys; 7. *In situ* TNB activation; 8. Heterodisulfide *xlnc*H3 formation.

Such a synthetic route enables the highest modularity for the generation of a library of combinatorially modified H3 proteins, since a multitude of differentially modified histone N-terminal fragments can potentially be ligated to the H3K36me3-containing, semisynthetic H3 core (**Figure 27C**).

In a first step, Dr. Carolin Lechner synthesized the middle fragment H3(29-46)A29C,K36me3 **1** as a C-terminal hydrazide by Fmoc protocol on a hydrazine resin as a thioester precursor (**Figure 28A, B**). Importantly, she used Thz as a convenient N-terminal cysteine protection in the peptide hydrazide **1**. This protecting group had long been reported to be incompatible with the conditions required to convert C-terminal hydrazides to thioesters, due to its oxidation to a nitrosamine side product by NaNO2.[302] Indeed, when the reaction was tested on peptide **1** using 5 eq. NaNO2, the nitrated side product (**1^{SR*}**) was observed as the major species (**Figure 28C, D**). She however realized that this side-reaction can be easily avoided by reducing the amount of NaNO2 in the reaction. By performing the oxidation step with equimolar amounts of NaNO2, followed by the addition of the thiol additive MTG, peptide **1** was cleanly converted into thioester **1^{SR}**. Of note, an alternative approach employing the addition of TFA to the conversion mixture was recently described.[329]
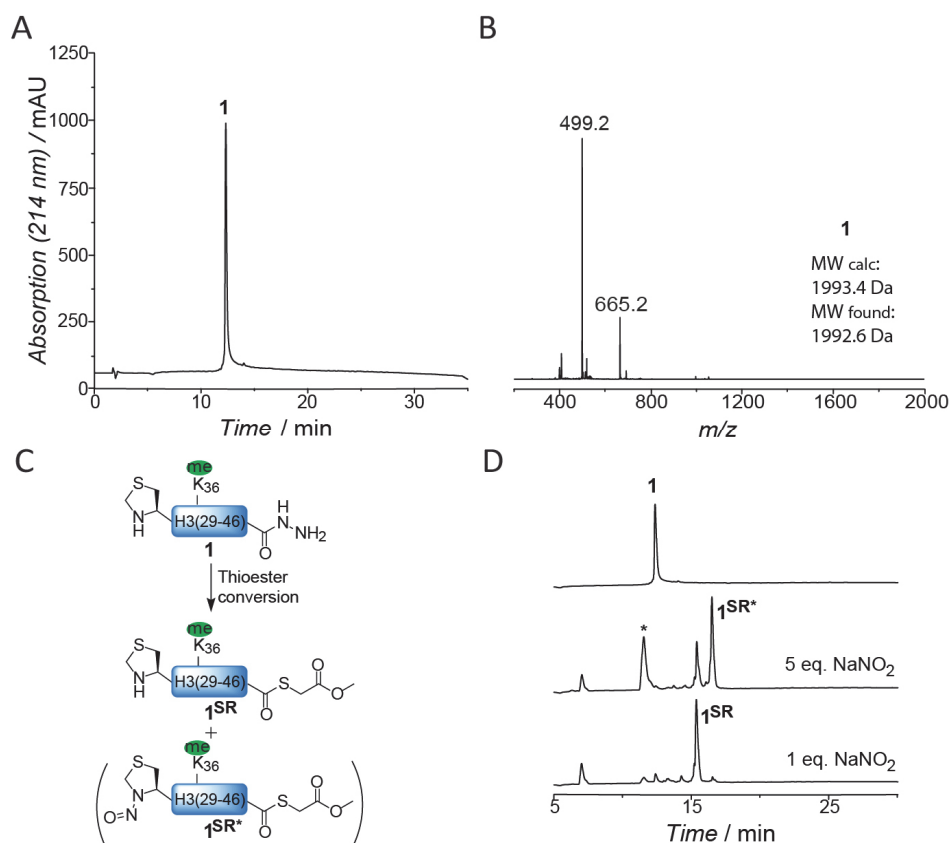
Figure 28. Synthesis of Thz-protected peptide hydrazide 1 and conversion to MTG thioester 1^{SR} employing different equivalents of NaNO2. A) Analytical RP-HPLC chromatogram and **B)** ESI-MS analysis of purified peptide **1** (MW calculated: 1993.4 Da, MW found: 1992.6 Da). **C)** Scheme of the reaction. **D)** Analytical RP-HPLC chromatograms of the reaction, showing the presence of the oxidized peptide side product **1^{SR*}** and a secondary side product (*) when 5 eq. NaNO2 are used, whereas the reaction proceeds cleanly with 1 eq. NaNO2. The side product peak at 12 min (*) most likely corresponds to S-nitroso methyl thioglycolate. Data by Dr. Carolin Lechner.

2.4.3 Synthesis of full-length lncH3

Having established a straightforward hydrazide to thioester conversion in the presence of thiazolidines, Dr. Carolin Lechner proceeded with the first ligation, between peptide **1^{SR}** and the recombinantly produced H3(47-135)A47C (**2**). Initial experiments revealed that the MTG-thioester in **1^{SR}** reacted only very sluggishly, due to the presence of a C-terminal valine in **1**.[259] She therefore opted for a more reactive thioester with MPAA (**1$^{SR'}$**). In addition, the crude peptide **1** could directly be converted after diethyl ether precipitation and resuspension, again using 1 eq. NaNO$_2$ and thus avoiding a material- and time- consuming purification step. When using MPAA as a thiol additive, the ligation proceeded to completion within 16h (**Figure 29A, B, C**). Subsequently, palladium-catalyzed conversion of Thz to Cys was performed in the same pot to afford ligation product **3** with 35% yield (**Figure 29B, D**).
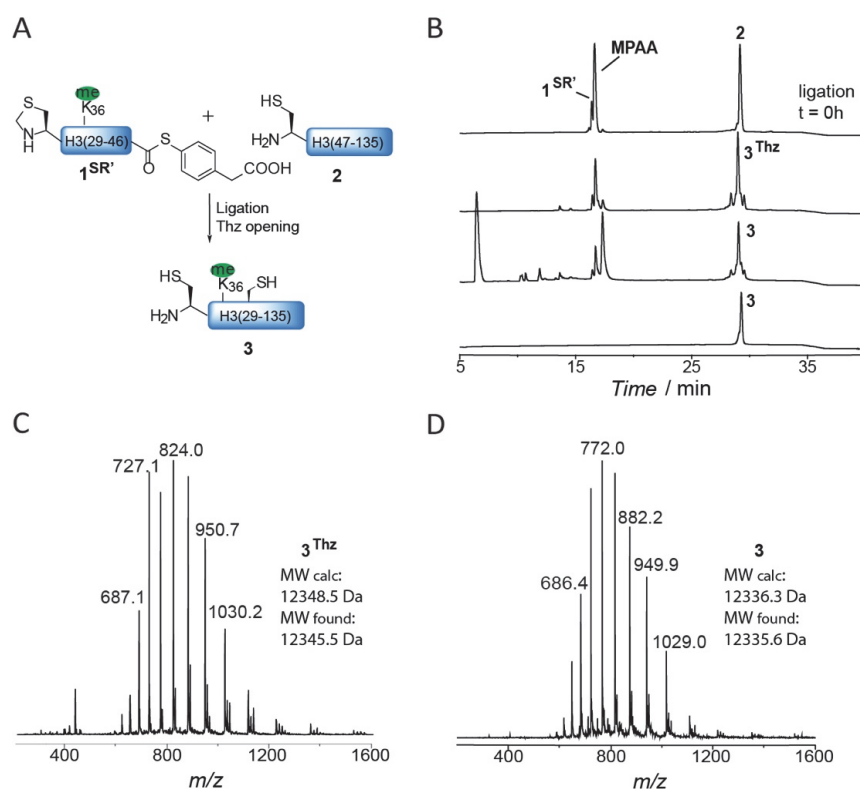
Figure 29: **Synthesis of protein 3 by one-pot ligation and Thz deprotection. A**) Scheme of the reaction. **B**) Analytical RP-HPLC chromatograms of the preparation of protein **3**. **C**) MS analysis of **3^{Thz}** (MW calculated: 12348.5 Da, MW found: 12345.5 Da). **D**) MS analysis of **3** (MW calculated: 12336.3 Da, MW found: 12335.6 Da). Data by Dr. Carolin Lechner.

Crucially, product **3** represents a precious intermediate building block for our modular synthetic strategy as it can be used in combination with different PTM-carrying H3 N-terminal peptides to expand the library of combinatorially modified H3K36me3-containing asymmetric nucleosomes**.** For the purpose of this work, **3** was ligated to the *lnc*-tagged H3 N-terminal peptide **4**. Peptide **4**, corresponding to *lnc*-H3(1-28), was synthesized by Dr. Carolin Lechner as a peptide hydrazide and with its critical N-terminal cysteine protected using Acm (**Figure 30A, B**). Ligation between **3** and **4** was performed using standard hydrazide

ligation conditions (**Figure 30C, D**).[266] Subsequent one-pot radical desulfurization[284] of Cys29 and Cys47 at the ligation junctions yielded the Acm-protected *lnc*-tagged H3K36me3 histone protein **5** in good yield (55 %, **Figure 30C, D, E**). Finally, to liberate the cysteine at the N-terminal *lnc*-site, the Acm group was removed using silver acetate (AgOAc) to afford product **6** (69 % yield, **Figure 30C, D, F**).
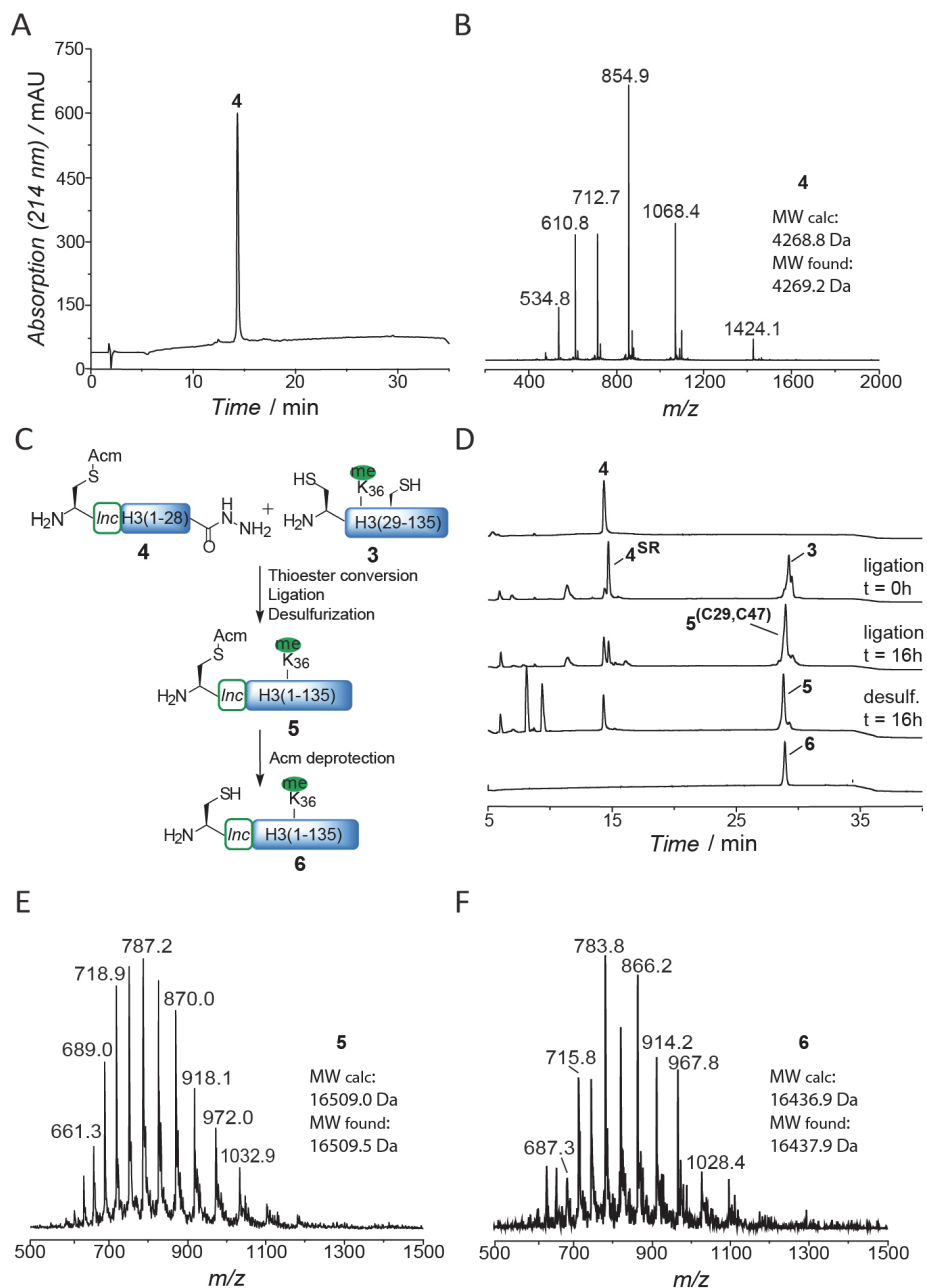
Figure 30: Synthesis of protein 6 by one-pot ligation and desulfurization followed by Acm removal. A) Analytical RP-HPLC chromatogram and **B**) ESI-MS analysis of purified peptide **4** (MW calculated: 4268.8 Da, MW found: 4269.2 Da). Data by Dr. Carolin Lechner. **C**) Scheme and **D**) analytical RP-HPLC chromatograms of the preparation of the full-length *lnc*-tagged H3 protein **6**. **E**) ESI-MS analysis of **5** (MW calculated: 16509.0 Da, MW found: 16509.5 Da). **F**) ESI-MS analysis of **6** (MW calculated: 16436.9 Da, MW found: 16437.9 Da).

2.4.4 Formation of ^{xlnc}H3 heterodisulfide dimers

At this point, the protein can be combined with any other *lnc*-tagged histones to form various asymmetric disulfide dimers. To produce asymmetric, bivalent H3K27me3/ H3K36me3 dimers, Dr. Carolin Lechner reacted **6** with a second H3, carrying both a *lnc*-tag and the H3K27me3 modification (**7, Figure 31A, B, C**).
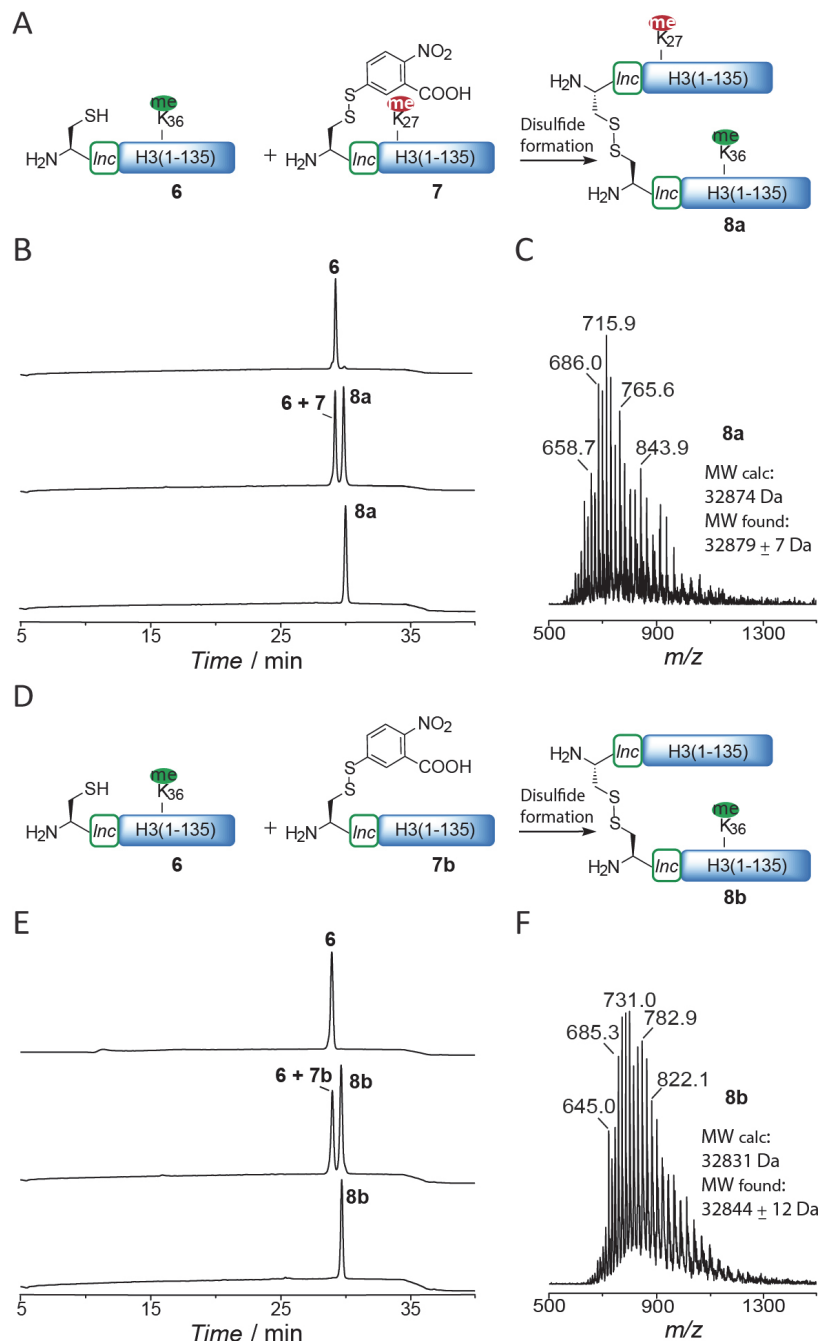
Figure 31: **Synthesis of asymmetric disulfide dimers 8a and 8b**. **A**) Scheme of the reaction yielding bivalent heterodisulfide dimers **8a** carrying asymmetric K27me3 and K36me3 modifications. **B**) Analytical RP-HPLC chromatograms of the preparation of bivalent H3K36me3/H3K27me3 heterodisulfide dimers **8a**. **C**) ESI-MS analysis of **8a** (MW calculated 32874 Da, MW found: 32879 ± 7 Da). Data by Dr. Carolin Lechner. **D**) Scheme of the reaction yielding asymmetric heterodisulfide dimers **8b** carrying asymmetric K36me3 modification. **E**) Analytical RP-HPLC chromatograms of the preparation of asymmetric H3K36me3 heterodisulfide dimers **8b**. **F**) ESI-MS analysis of **8b** (MW calculated 32831 Da, MW found: 32844 ± 12 Da).

Histone **7** was synthesized as previously described[104] and the critical cysteine in the *lnc*-tag sequence was activated as a heterodisulfide with Ellman's reagent (**Figure 31A**). Thiolate-disulfide exchange between **6** and **7** afforded crosslinked ˣˡⁿᶜH3 (H3K27me3 /H3K36me3) **8a** with 40 % purified yield (**Figure 31B, C**). In addition, I reacted **6** with unmodified *lnc*-H3 to obtain ˣˡⁿᶜH3 asymmetrically trimethylated at K36 (asH3K36me3, **8b**, **Figure 31D, E, F**).

2.4.5 Synthesis of H3K36me3 histones for symmetric H3K36me3 nucleosomes

I obtained the control symmetric histones, trimethylated at lysine 36 on both H3 tails, by reacting fragment **3** with peptide **4'**. Peptide **4'** was synthesized by Fmoc-SPPS as C-terminal hydrazide and did not present the *lnc*-tag at the N-terminus. Ligation and subsequent desulfurization reactions afforded histone **5'** (**Figure 32A, B, C**).
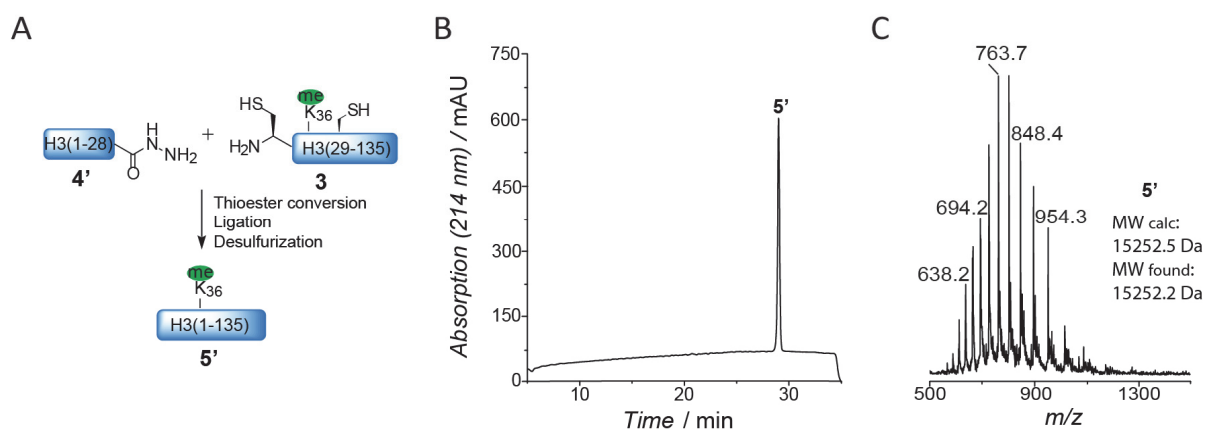
Figure 32: Synthesis of protein 5' by one-pot ligation and desulfurization of peptide 4' and fragment 3. A) Scheme of the reaction. **B)** Analytical RP-HPLC chromatogram of purified **5'**. **C)** ESI-MS analysis of purified **5'** (MW calculated: 15252.5 Da, MW found: 15252.2 Da).

2.4.6 Refolding of histone octamers and reconstitution of nucleosomes

I then proceeded to refold asymmetric histone octamers, by combining **8a** or **8b** with 2 eq. of the recombinant histone H4, and 2.2 eq. of both histones H2A and H2B under denaturing conditions (6 M guanidinium hydrochloride, GmdCl) followed by dialysis to native conditions and size exclusion chromatography purification (**Figure 33**). Concomitantly, symmetric H3K36me3 and unmodified octamers were produced following the same procedure.
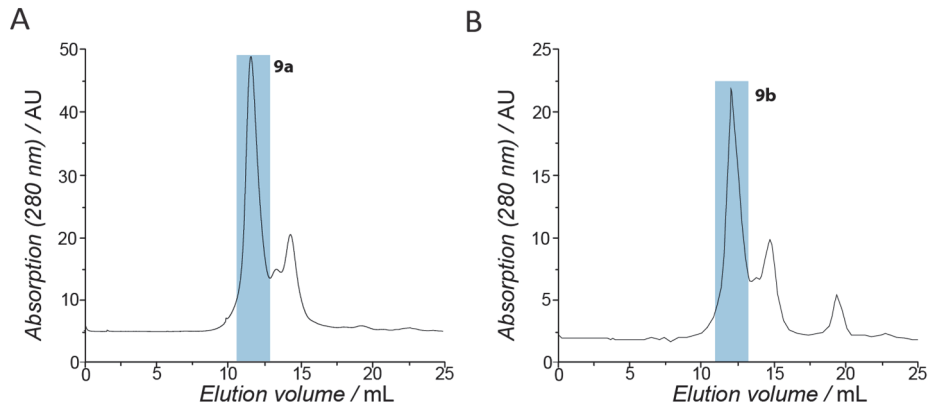
Figure 33: Size exclusion chromatography purification of asymmetric octamers 9a (A) and 9b (B). The blue rectangles indicate the fractions containing the octamers.

Asymmetric nucleosomes **10a** and **10b** were subsequently reconstituted by combining octamers **9a** or **9b** (containing the appropriate *xlnc*H3 molecules) with a DNA fragment consisting of the high-affinity 601 nucleosome positioning sequence[10] under high salt conditions (2 M NaCl). Subsequently, nucleosome assembly was initiated by a gradual dialysis to 10 mM KCl over 12 h. Finally, the disulfide bond in *xlnc*H3 was reduced and the *lnc*-tag was removed using TEV protease digestion (**Figure 34A, B**). To complete our library, we also synthesized and reconstituted nucleosomes containing asH3K27me3 (**10c**, following published protocols[104]) as well as nucleosomes symmetrically modified with H3K36me3 (**10d**) or unmodified (**Figure 34A, B**).
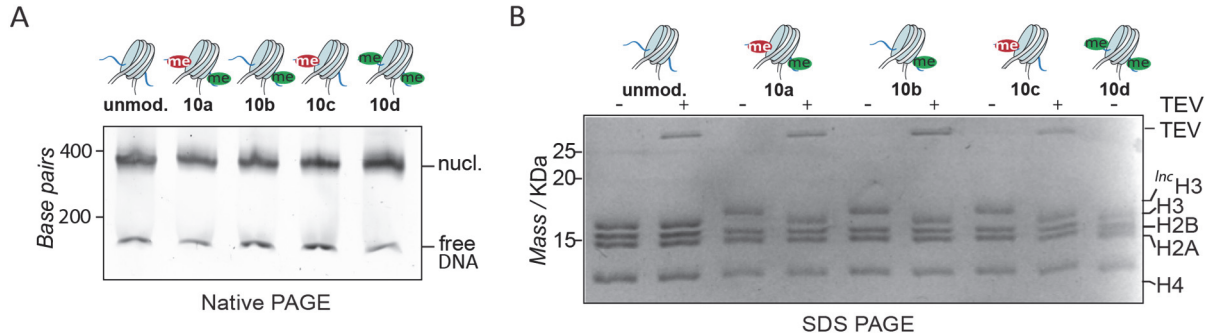
Figure 34: Nucleosome reconstitution and *lnc*-tag cleavage by TEV protease. A) Native-PAGE analysis of the reconstituted nucleosomes. **B**) SDS-PAGE analysis of nucleosomes before and after *lnc*-tag cleavage.

2.4.7 PRC2 activity is stimulated by H3K27me3 in *trans* and inhibited by H3K36me3 in *cis*

Having the library of relevant asymmetric nucleosomes in hand, notably unmodified, bivalent H3K27me3/H3K36me3 (**10a**), asH3K36me3 (**10b**), asH3K27me3 (**10c**) and symH3K36me3 (**10d**) nucleosomes (**Figure 34A and Figure 35A**), I proceeded to investigate the crosstalk of H3K27me3 and H3K36me3 in PRC2 regulation. I probed methyltransferase activity of recombinant PRC2 complex using ^{3}H-S-adenosyl-methionine (^{3}H-SAM) as a cofactor. I performed both endpoint experiments by fluorography (**Figure 35A**, **bottom**) and kinetic studies using liquid scintillation counting (**Figure 35B**). In accordance with previous work,[102,104] asH3K27me3 (**10c**) allosterically stimulated PRC2 activity compared to

unmodified nucleosomes. In contrast, asH3K36me3 (**10b**) significantly hindered the methyl mark deposition and symmetric H3K36me3 (**10d**) completely abolished PRC2 activity within the same nucleosome. These results highlight the inhibitory and stimulating function of H3K36me3 and H3K27me3 marks on PRC2 methylation, respectively.[322,330,324] Importantly, when bivalent nucleosomes (**10a**) were used as substrate, PRC2 methylation activity was reduced by approximately 50% when compared to unmodified nucleosomes. These findings support the conclusion that the transcriptionally activating H3K36me3 modification inhibits PRC2 activity on the same tail (*in cis*), while the heterochromatic mark H3K27me3 extends its stimulatory effect across the nucleosome, partially overriding H3K36me3-based inhibition (**Figure 35C**). In conclusion, PRC2 activity is therefore regulated by H3K36me3 via a similar mechanism as compared to H3K4me3, another critical bivalent PTM.[104]
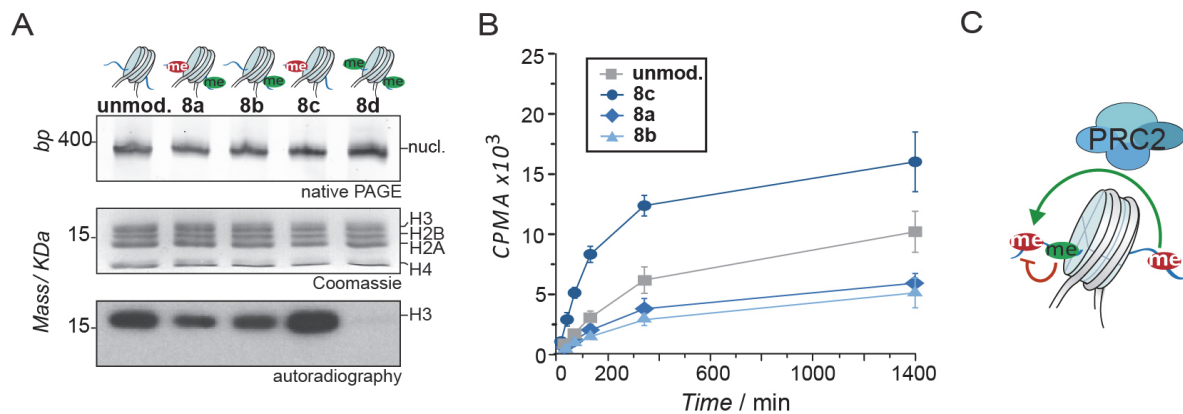
Figure 35: PRC2 activity is stimulated by H3K27me3 in *trans* and inhibited by H3K36me3 on the same histone tail. A) Methyltransferase assays with recombinant PRC2. From left to right: unmodified, bivalent H3K36me3/H3K27me3 **10a**, asymmetric H3K36me3 **10b**, asymmetric H3K27me3 **10c**, symmetric H3K36me3 nucleosomes, **10d**. From top to bottom: native PAGE, SDS-PAGE and fluorography. **B**) Kinetics of PRC2 methyltransferase activity towards indicated nucleosomes. CPMA = counts per minute for alpha radiation. Error bars represent standard deviation. n=3. **C**) Model of PRC2 nucleosome methylation.

2.4.8   H3K36me3 prolongs PRC2 chromatin residence time

The mechanism of how H3K36me3 inhibits PRC2 activity is not clear. We thus wondered if H3K36me3 altered PRC2 binding to chromatin fibers. Using a recently established single-molecule total internal reflection fluorescence microscopy (smTIRFM) assay, Dr. Andreas Bachmann measured PRC2 dwell times (i.e. dissociation kinetics) on synthetic chromatin fibers (**Figure 36A**).[331] Chromatin arrays, carrying either no PTMs (unmodified), symmetric H3K27me3 or symmetric H3K36me3 were reconstituted (using a DNA construct containing 12 repeats of a 177 bp long 601 sequence), fluorescently tagged with a far-red dye (Atto647N) and biotinylated for surface immobilization. Subsequently, the chromatin fibers were immobilized in a PEG-passivated microfluidic channel through biotin-neutravidin interaction. A 2 nM solution of DY-547 labeled recombinant PRC2 complex (obtained with a ybbR-tag based labeling strategy[332]) was then injected. PRC2 binding events to immobilized chromatin fibers were detected as fluorescent spots in the green-orange channel (specific for DY-547) which colocalized with chromatin positions (detected in the far-red channel). Dr. Andreas Bachmann recorded movies of the interaction

kinetics using a 50 ms frame rate, and extracted kinetic traces for each chromatin position (**Figure 36B**). Using a thresholding algorithm, he detected the length of individual PRC2 binding events to unmodified, H3K36me3 or H3K27me3 containing chromatin fibers, and using this data he reconstructed lifetime histograms of the PRC2 bound state (**Figure 36C**). For all three chromatin types, biphasic dissociation kinetics were observed. An initial fast phase, corresponding to binding events with sub-second lifetimes, reveals transient PRC2 probing interactions.[331] On the other hand, the longer-lived PRC2 binding events (with seconds lifetimes) result from more specific interactions, most probably associated with catalytically competent PRC2-chromatin complexes. The fraction of long and short binding events (i.e. the amplitudes of the kinetic phases) and the time constants $\tau_{off,1}$ and $\tau_{off,2}$ report on the mechanism of PRC2 chromatin binding (**Figure 36D**).
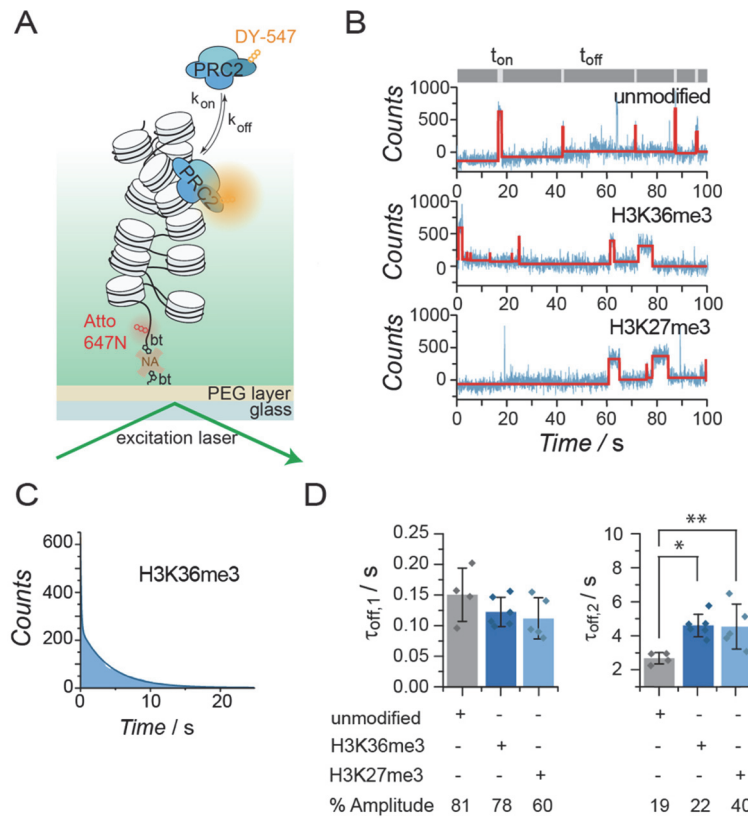
Figure 36: H3K36me3 allosterically inhibits PRC2 activity. **A**) Scheme of the smTIRFM experiment to detect PRC2 chromatin interaction dynamics. bt: biotin. NA: neutravidin. **B**) Representative kinetic traces of PRC2 binding dynamics to unmodified, H3K36me3 or H3K27me3 modified chromatin fibers. An increase in fluorescence above a threshold value (200 counts) indicates transient PRC2 binding. Bound time intervals (t_{on}) are gathered to derive dissociation kinetics. **C**) Cumulative lifetime histogram of t_{on} times (shown for H3K36me3 containing chromatin fibers). The lifetime histograms are fitted with a double exponential function, y = $A_1 \cdot exp(-t/\tau_{off,1})$ + $A_2 \cdot exp(-t/\tau_{off,2})$. The initial fast phase shows short-lived probing interactions, whereas the slow phase is associated with more stably bound PRC2 complexes. **D**) Dissociation time constants $\tau_{off,1}$ and $\tau_{off,2}$ and amplitudes A_1, A_2 for PRC2 binding to unmodified, H3K27me3 and H3K36me3 containing chromatin fibers.

Compared to unmodified chromatin ($\tau_{off,2}$ = 2.7 ± 0.3 s), PRC2 binding is strengthened by the presence of H3K27me3 ($\tau_{off,2}$ = 4.5 ± 1.3 s). In addition, the fraction of long binding events is increased to 40% (compared to 19% in unmodified chromatin). Indeed, PRC2 contains a subunit (EED) which binds to H3K27me3[330] and this interaction is involved in a stimulation of PRC2 activity.[324] However, PRC2

chromatin interactions are mainly governed by DNA interactions,[331,333] explaining a relatively modest increase in residence time. Surprisingly, H3K36me3 also seemed to prolong PRC2 binding times ($\tau_{off,2}$ = 4.6 ± 0.7 s) compared to unmodified fibers (**Figure 36D**), albeit the fraction of long binding events was lower as compared to H3K27me3 containing chromatin. This indicates that PRC2 engages K36me3 modified H3, however, the histone tail is bound in a catalytically non-competent state and, in addition, could further allosterically inhibit PRC2 activity.

2.5 Conclusions

In summary, we have developed a synthetic route to asymmetric nucleosomes containing H3K36me3. The modular approach shown here enables the rapid assembly of libraries of asymmetric nucleosomes. Moreover, we have shown that H3K36me3 reduces PRC2 activity in bivalent nucleosomes *in cis*, whereas this inhibition can partially be overcome by the presence of H3K27me3 in the same nucleosome. However, although modular and general, this strategy suffers from a limitation. In fact, the *lnc*-tag installed at the N-terminus of the protein prevents any other modifications to be established at that position. Yet, the vast majority of eukaryotic proteins are processed by enzymes at their N-terminus, with modifications such as acetylation,[334] methylation,[335] myristoylation,[336] and palmitoylation.[337] These PTMs bring a significant contribution to the complex and diverse proteome and play key roles in the regulation of protein functions and cellular signaling.[338,339] Thus, in the next chapter (**Chapter 3**), I will describe in detail how this methodology can be revised and adapted to the synthesis of asymmetric complexes bearing modified N-termini, such as asymmetric H4 nucleosomes.

Finally, we have demonstrated that the presence of H3K27me3 prolongs PRC2 binding to nucleosomes, and that a weaker effect is also detected for H3K36me3. In both cases, the histone tails make direct contact to PRC2 subunits, resulting either in allosteric activation or repression, and these interactions are reflected in increased binding times. Indeed, besides having access to chemically defined nucleosomes substrates, efficient and straightforward labeling strategies, together with stable and bright fluorophores, are paramount for the biophysical study of chromatin dynamics and binding kinetics. Indeed, in **Chapter 4** I will report on how we addressed this challenge and developed a "doubly orthogonal" labeling strategy that was used to chemically stabilize fluorophores for single-molecule microscopy studies.

Together, we expect that asymmetric nucleosomes, combined with biochemical and single-molecule dynamic methods provide a highly useful toolbox to probe chromatin effector regulation by combinations of PTMs, not limited to PRC2.

2.6 Acknowledgements

Chapter 3: Semisynthesis and Reconstitution of Asymmetrically Modified H4K20me1 Nucleosomes

3.1   Outline

In stem cells, H4 proteins carrying different modifications coexist within single nucleosomes.[102] For functional studies, we report the synthesis of such asymmetric nucleosomes. Asymmetry is achieved by transiently crosslinking H4 by a traceless, protease-removable tag introduced via an isopeptide linkage (iso*lnc*-tag). The iso*lnc*-tag, in contrast to previously developed strategies,[104,309] is applicable to all naturally N-terminal acetylated proteins, such as H4. These nucleosomes are used to study Set8 activity, a key methyltransferase.

3.2   Contributions

This chapter is based on the following publication:

Guidotti, N., Lechner, C. C., & Fierz, B. (**2017**). Controlling the supramolecular assembly of nucleosomes asymmetrically modified on H4. *Chemical Communications*, *53*(74), 10267-10270.

Contributions by other authors are described, where appropriate, along the text and in figure legends. In particular:
- G. N. synthesized the majority of the peptides, performed the ligation and desulfurization reactions, reconstituted the nucleosomes, performed the TEV cleavage on nucleosomes and performed methyltransferase assays.
- L. C. C. synthesized the branched peptide building block and performed TEV cleavage tests on the branched peptide.

3.3   Introduction

Nucleosomes, the basic unit of chromatin, organize 147 bp of DNA wrapped around two of each core histones H3, H4, H2A and H2B.[1] The histone proteins carry combinations of post-translational modifications, which are implicated in regulating crucial chromatin functions.[311] In particular, methylation of lysine residues on H3 and H4 have clearly defined roles in gene activation and repression, with implication for cell differentiation, development and disease.[340] Detailed MS studies found that key methyl-

marks in embryonic stem cells exist in asymmetric nucleosomes, i.e. nucleosomes carrying two differently modified copies of H3 or H4.[102] These PTMs include H4 monomethylated at lysine 20 (H4K20me1), as well as H3K4me3, H3K36me3 and H3K27me3.[92,312,341]

The amounts of H4K20 monomethylation oscillate during the cell cycle: while the methylation mark is normally absent during DNA replication, it reaches its maximal levels during mitosis.[342] The methyltransferase Set8 (also known as PR-Set7 or KMT5A) is the sole enzyme directly responsible for the deposition of H4K20me1 modification and peaks of H4K20me1 are mirrored in Set8 abundance.[74] Set8 is a member of the SET domain family of lysine methyltransferases.[74,343] Its central SET domain is flanked by N- and C-terminal SET helical regions,[344] and a central variable insert, i-SET, divides the domain into two halves and contributes to H4 peptide specific recognition. Set8 enzymatic activity is greater on nucleosome substrates rather than on free histones.[74,343] Indeed, structural and biochemical studies have revealed multivalent interactions between Set8, the H2A/H2B acidic patch, the nucleosomal DNA and the H4 tail.[345] Despite being the only H3K20me writer, a set of other enzymes controls the overall amounts of this modification. Suv4-20h1 and Suv4-20h2 are methyltransferases that catalyze the di- and trimethylation of this residue using H4K20me1 as a substrate, and together with PHF8, a demethylase that removes the methyl mark from H4K20, they balance Set8 activity and provide a tight regulation of H4K20me1 levels along the cell cycle. H4K20 monomethylation roles are multifaced: it is involved in chromatin structure regulation,[346,347] maintenance of the genome integrity,[348,349] proper development,[350,351] plays key role in DNA replication[352] and DNA damage repair,[353,354] and also serves as a binding site for chromatin regulators, such as the chromatin compactor L3MBTL1.[355] Although the evidence for the multiple involvement of H4K20me1 in a palette of different pathways modulating the cell cycle exists, the molecular mechanisms contributing to the H4K20me1 regulation are yet to be fully uncovered. Moreover, the discovery of asymmetric H4K20me1 nucleosomes adds an additional layer of complexity to the intricated interplay existing between establishment, removal and recognition of the H4K20me1 mark. The fact that such nucleosomes are not readily available for detailed *in vitro* mechanistic studies has considerably hindered their investigation.

3.4   Results

Biochemical studies of the functions of asymmetrically modified chromatin require facile access to chemically defined asymmetric nucleosomes. While expressed protein ligation methods readily enable the installation of combinations of marks on nucleosomes,[201,356,357,358] the reconstitution of asymmetric chromatin is not straightforward and is largely based on the attachment of affinity tags, that usually remained on the nucleosomes, followed by multi-step purification schemes.[102,359,105]

Our group has recently developed a chemical method to address this problem and control supramolecular nucleosome assembly.[104,309] The *lnc*-tag at the N-terminus of the H3 protein enabled transient crosslinking (by a heterodisulfide bond) of H3 species during nucleosome reconstitution. Subsequently, the crosslink was reversed and the *lnc*-tag was removed using TEV protease yielding scarless asymmetric H3 nucleosomes (see **Chapter 2.4.1**).

3.4.1 Asymmetric H4 nucleosomes: strategy and design

Based on these studies, we (Dr. Carolin Lechner and myself) wondered if this synthetic method could be adapted to synthesize crosslinked versions of differentially modified H4 (xlncH4).

There is however a complication, which prevented the direct application of the original *lnc*-tag strategy to H4: in cells, H4 is acetylated at its N-terminus.[360] We thus considered to install the *lnc*-tag not at the N-terminus but at an internal position in the H4 tail (**Figure 37**), notably via an isopeptide-bond at a lysine side chain close to the N-terminus (*iso*lnc-tag, **Figure 37B, C**). To recover the native H4 sequence, and thus establish a fully traceless method, the removal of the non-natural *iso*lnc-tag sequence is however paramount. We hypothesized that TEV protease might be able to cleave a glutamine-lysine isopeptide bond, due to the similarity of the lysine side chain to the glycine found at the P1' position in the canonical TEV cleavage sequence. Transient *iso*lnc-tag based crosslinking of differentially modified H4 molecules, followed by TEV digestion of the isopeptide bond would thus result in a synthetic approach compatible with N-terminal protein modifications (**Figure 37A**).
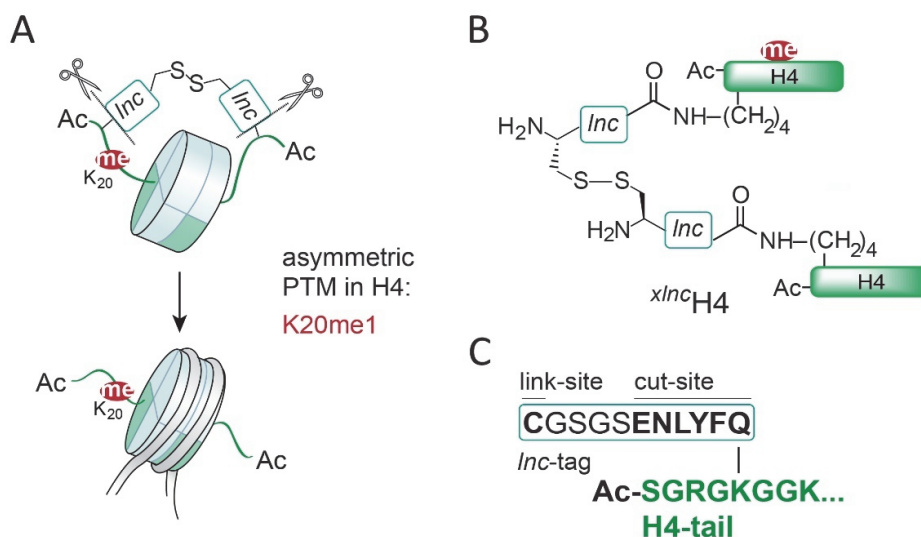
Figure 37: Synthetic strategies for nucleosomes asymmetrically modified on H4. **A)** Synthetic strategy for H4, employing an isopeptide crosslink strategy to preserve the acetyl group at the H4 amino terminus. **B)** xlncH4, with installed *lnc*-tag at the side chain of K5. **C)** Sequence of the *iso*lnc-tag at the H4-tail.

3.4.2 Isopeptide bond and TEV cleavage

To test our hypothesis that TEV protease indeed can digest an isopeptide bond, Dr. Carolin Lechner synthesized the branched peptide **1**, corresponding to H4 residues 1 – 14 (**Figure 38A**), using a Fmoc protection scheme on a hydrazine resin. The lysine at position 5 was initially incorporated carrying the orthogonal Alloc protecting group. After acetylation of the N-terminus, Alloc was removed using Pd(PPh$_3$)$_4$ in the presence of PhSiH$_3$ as a scavenger, making the lysine side chain amino group available for *^{iso}lnc*-tag synthesis. The *^{iso}lnc*-tag included an Acm-protected cysteine as a final residue, the key amino acid required for the formation of the critical crosslink in our synthesis scheme (**Figure 38A**). She then incubated the purified peptide **1** with TEV protease and analyzed the isopeptide cleavage reaction over extended time (**Figure 38B**). To our delight, TEV protease indeed catalyzed complete cleavage of the Q-K isopeptide bond and released the native H4 peptide **1'** as judged by RP-HPLC (**Figure 38B**) and MS analysis (**Figure 38C, D**).
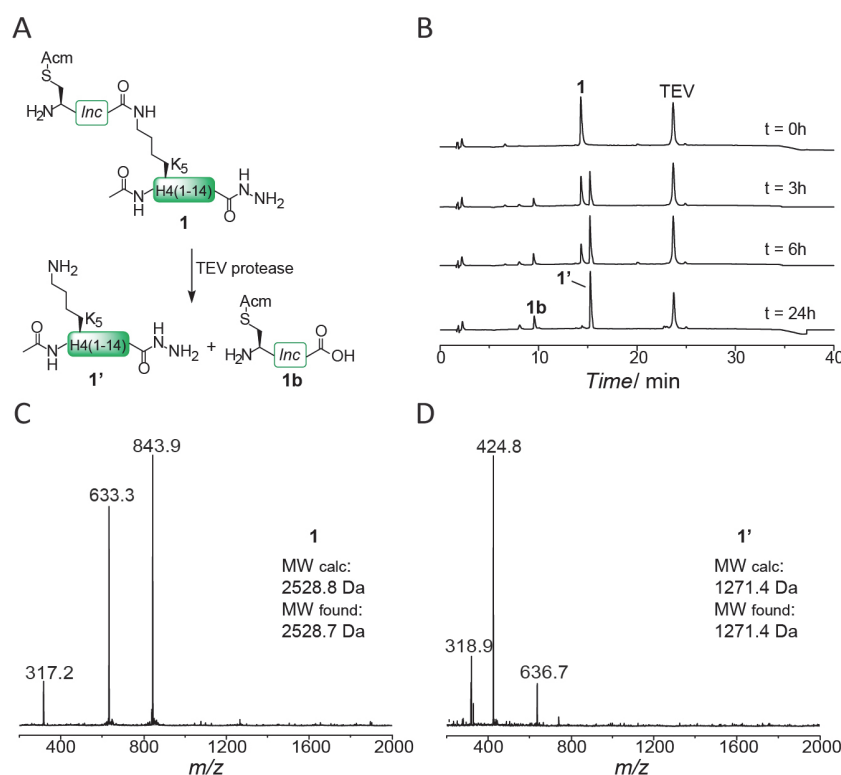
Figure 38: The *^{iso}lnc*-tag is cleaved by TEV protease. A) Branched peptide **1** is cleaved by TEV protease to yield the H4 tail fragment **1'** and the free *lnc*-tag **1b**. **B)** Analysis of the cleavage reaction by RP-HPLC at the indicated time points. **C)** ESI-MS analysis of peptide **1** (MW calculated: 2528.8 Da, MW found: 2528.7 Da) **B)** ESI-MS analysis of peptide **1'** (MW calculated: 1271.4 Da, MW found: 1271.4 Da). Data by Dr. Carolin Lechner.

3.4.3 Synthesis of full-length *^{lnc}*H4

Encouraged by this proof-of-concept experiment we developed, based on previous routes for H4 synthesis,[361,362,363] a general and modular synthetic access to ^{xlnc}H4, asymmetrically mono-methylated at

K20 ([xlnc]H4K20me1 (**8**), **Figure 39**). To ensure a modular approach allowing the incorporation of different histone PTM patterns, as well as a traceless synthesis, we decided to divide H4 into 3 segments at alanine sites: ac-H4(1-14)K5([iso]lnc[Acm])-NHNH$_2$ (**1**), H4(15-37)A15C-NHNH$_2$ (at K20 either unmodified (**2**) or monomethylated (**2a**)) and H4(38-102)A38C (**4**). This allowed a modular assembly of differentially modified H4 proteins, as fragments **1** and **4** remained invariant and PTMs were solely introduced in the middle fragments **2**.
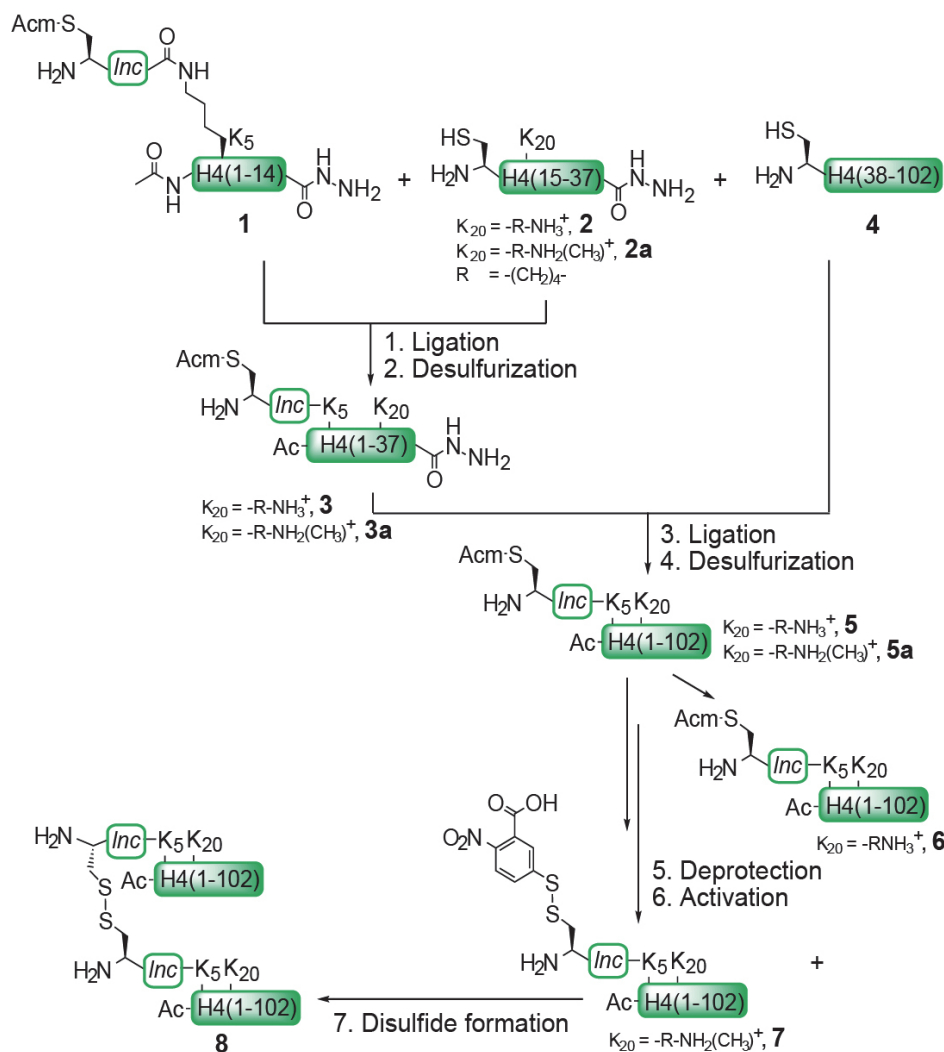
Figure 39: Synthesis of [xlnc]H4K20me1 for the reconstitution of asymmetrically modified nucleosomes. The strategy employs two sequential one-pot ligation-desulfurization steps, followed by the deprotection-activation of the [iso]lnc-tag and formation of the asymmetric disulfide **8**.

Peptides **2** and **2a** were synthesized by Fmoc protocols on a hydrazine resin, employing Fmoc-Lys(Boc,Me)-OH at position 20 for **2a** (**Figure 40A, B, D, E**). On the other hand, H4(38-102)A38C (**4**) was recombinantly expressed and purified (**Figure 40C, F**).

We then proceeded, in a first ligation reaction, to connect peptide **1** to either **2** or **2a** to produce the *isolnc*-tag containing N-terminal tail peptides **3** and **3a**. As the synthesis of *xlnc*H4K20me1 involves multiple steps, limiting purification steps is key. We thus aimed for employing ligation conditions which allow one-pot free-radical desulfurization.[285,364] Initial attempts with the native NCL catalyst TFET[286] resulted in large amounts of thioester hydrolysis of the branched peptide **1**. I thus opted for MTG as a thiol additive,[287] which resulted in slower ligation kinetics but showed drastically reduced hydrolysis product formation, while still allowing one-pot desulfurization. Implementing this strategy, I oxidized the C-terminal hydrazide in **1** *in situ* with $NaNO_2$ in 6 M GmdCl and phosphate buffer, pH 3, followed by conversion into a thioester by MTG addition.[287,302] Addition of either **2** or **2a** at neutral pH resulted in quantitative ligation within 2-4 hours with minimal amount of thioester hydrolysis. Importantly, only one equivalent of $NaNO_2$ was used in the oxidation reaction step to avoid conversion of the hydrazide on **2** or **2a**.
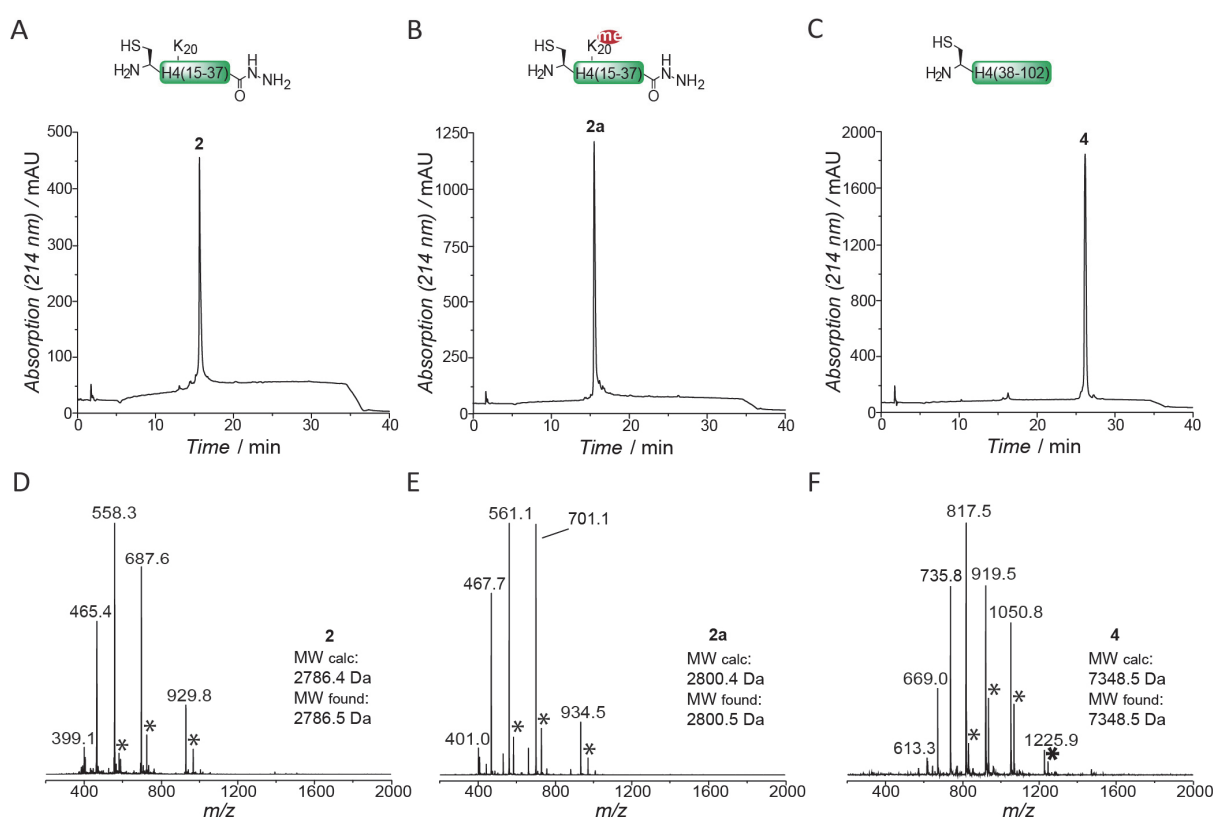
Figure 40: **Analysis of purified peptides 2, 2a and purified truncated H4 protein 4**. **A-B-C)** Analytical RP-HPLC chromatogram of purified peptide **2** (**A**), **2a** (**B**) and truncated H4 (**4**). **D)** ESI-MS analysis of purified peptide **2** (MW calculated: 2786.4 Da, MW found: 2786.5 Da). **E)** ESI-MS analysis of purified peptide **2a** (MW calculated: 2800.4 Da, MW found: 2800.5 Da). **F)** ESI-MS analysis of purified truncated H4 **4** (MW calculated: 7348.5 Da, MW found: 7348.5 Da). Asterisks correspond to TFA adducts.

With no intermediary purification step, free-radical desulfurization of the cysteine residue at the ligation site to the native alanine was performed employing TCEP and the radical initiator VA-044 at 42 °C to yield the peptides **3** and **3a** over 7h (isolated yield: 46%, **Figure 41**).
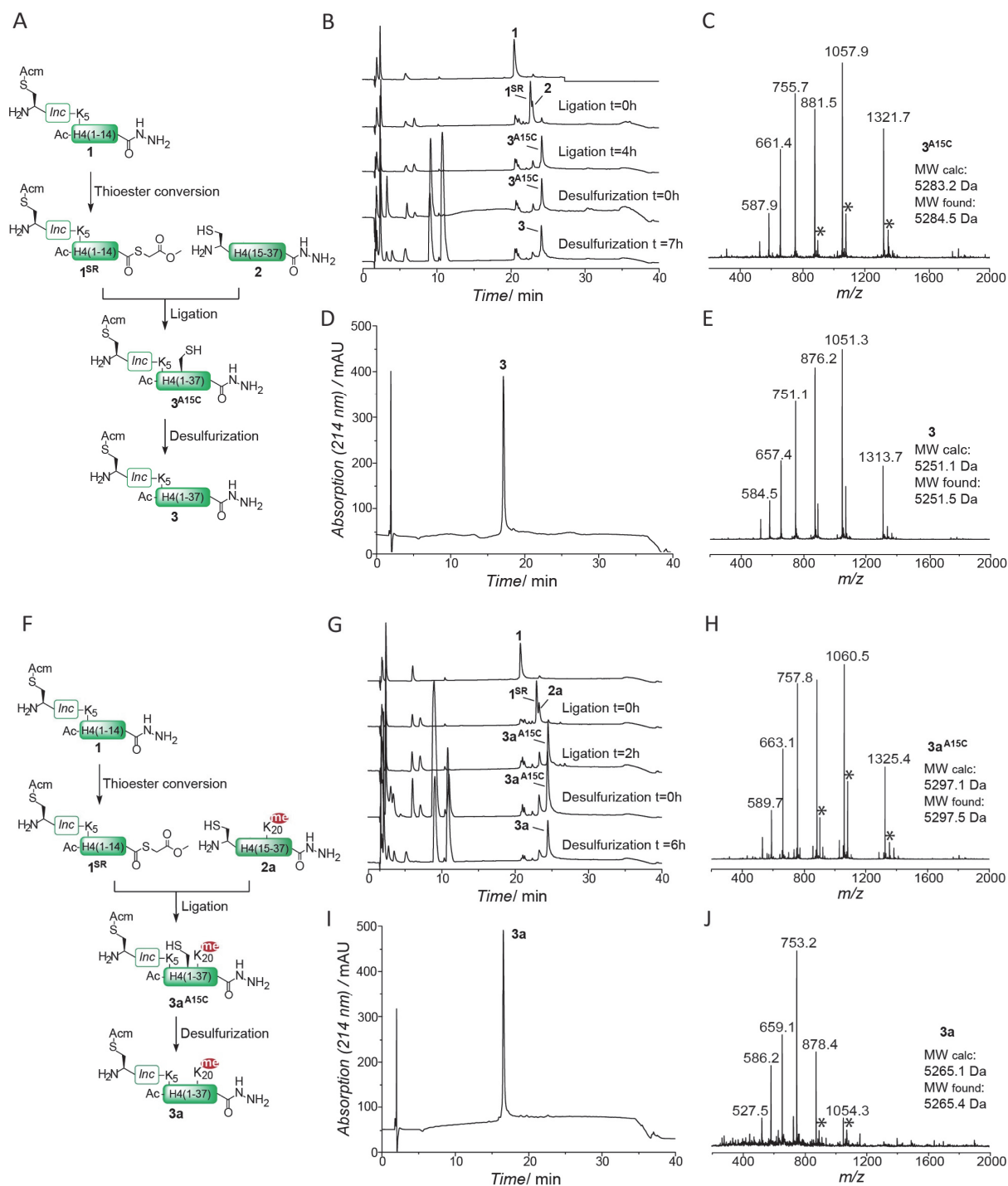
Figure 41: **Synthesis of fragments 3 and 3a by one-pot ligation and desulfurization. A**) Scheme of the reaction steps yielding product **3**. **B**) Analytical RP-HPLC chromatograms of the reaction steps. **C**) ESI-MS analysis of product **3**[A15C] (MW calculated: 5283.2 Da, MW found: 5284.5 Da). **D**) Analytical RP-HPLC chromatogram of purified desulfurized product **3**. **E**) ESI-MS analysis of product **3** (MW calculated: 5251.1 Da, MW found: 5251.5 Da). **F**) Scheme of the reaction steps yielding product **3a**. **G**) Analytical RP-HPLC chromatograms of the reaction steps. **H**) ESI-MS analysis of product **3a**[A15C] (MW calculated: 5297.1 Da, MW found: 5297.5 Da). **I**) Analytical RP-HPLC chromatogram of purified desulfurized H4 fragment **3a**. **H**) ESI-MS analysis of peak **3a**; Calculated mass for purified desulfurized H4 fragment **3a** (MW calculated: 5265.1 Da, MW found: 5265.4 Da). Asterisks correspond to TFA adducts.
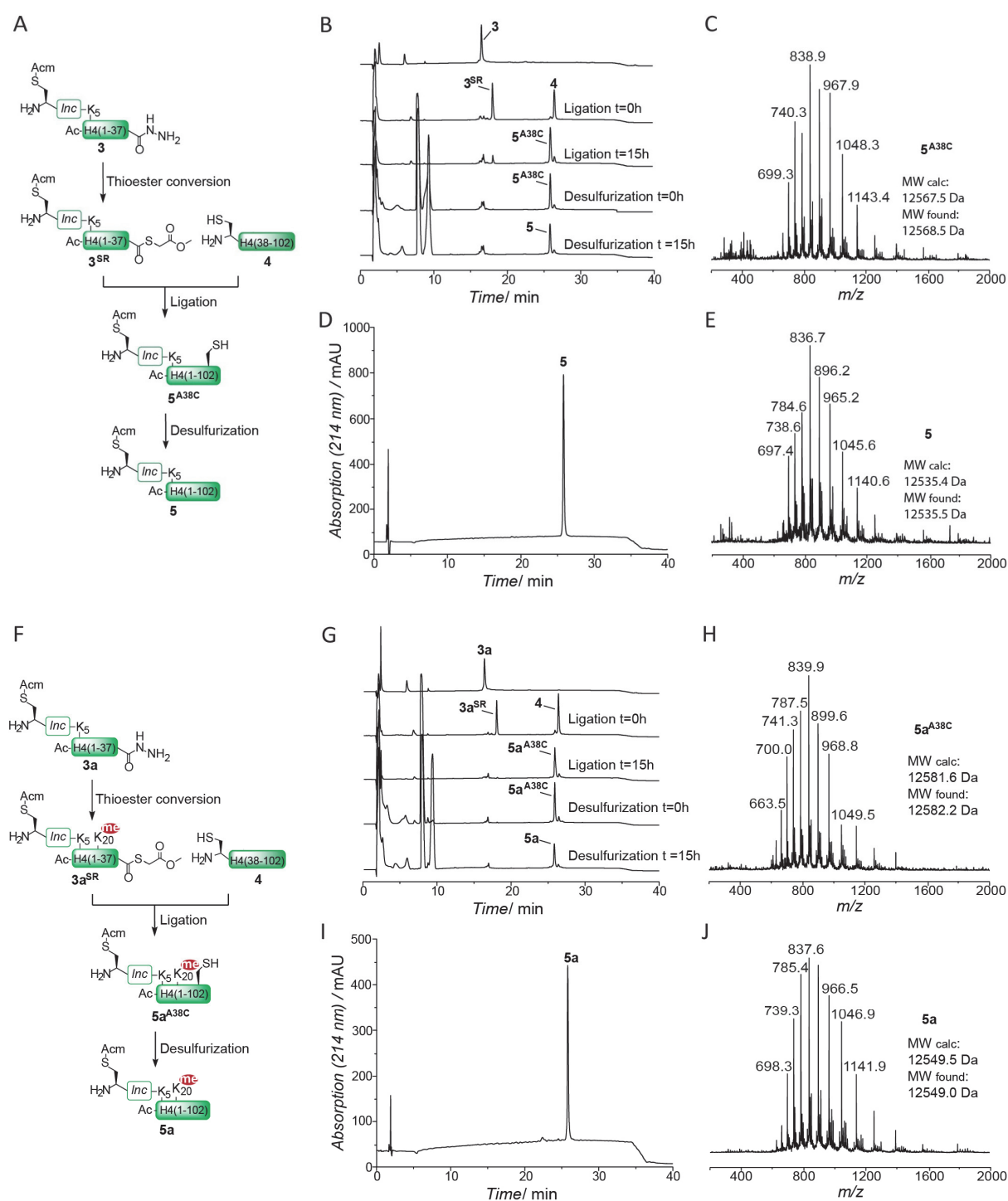
Figure 42: **Synthesis of H4 protein 5 and 5a by one-pot ligation and desulfurization**. **A**) Scheme of the reaction steps yielding product **5**. **B**) Analytical RP-HPLC chromatograms of the reaction steps. **C**) ESI-MS analysis of **5^A38C** (MW calculated: 12567.5 Da, MW found: 12568.5 Da). **D**) Analytical RP-HPLC chromatogram of purified desulfurized H4 protein **5**. **E**) ESI-MS of product **5** (MW calculated: 12535.4 Da, MW found: 12535.5 Da). **F**) Scheme of the reaction steps yielding product **5a**. **G**) Analytical RP-HPLC chromatograms of the reaction steps. **H**) ESI-MS analysis of product **5a^A38C** (MW calculated: 12581.6 Da, MW found: 12582.2Da). **I**) Analytical RP-HPLC chromatogram of purified desulfurized H4 protein **5a**. **J**) ESI-MS analysis of product **5a** (MW calculated: 12549.5 Da, MW found: 12549.0 Da). Asterisks correspond to TFA adducts.

Both **3** and **3a** were then ligated in a second reaction to H4(38-102)A38C (**4**), employing MTG as ligation catalyst, followed by one-pot desulfurization. During the desulfurization step, I observed significant precipitation of the ligation products **5**^{A38C} and **5a**^{A38C}. After testing different reaction conditions, I found that tenfold dilution of the reaction mixture after the ligation step and a reduced temperature (30 °C) for the desulfurization reaction ameliorated the precipitation problems and yielded **5** and **5a** with 35-40% isolated yield (**Figure 42**).

To liberate the cysteine residue required for crosslinking at the N-terminus of the *isolnc*-tag in **5** I removed the Acm group using AgOAc to yield **6** (isolated yield 63%, **Figure 43A, B, C**), and in the case of **5a**, activated the cysteine directly after deprotection with Ellman's reagent in the same-pot, yielding **7** (isolated yield 52%, **Figure 43D, E, F**). The fragments **6** and **7** were stored as lyophilized proteins at -20 °C. The advantage of our modular approach is that, at this time, differently modified histones can be combined to form various asymmetric heterodisulfide histone dimers.
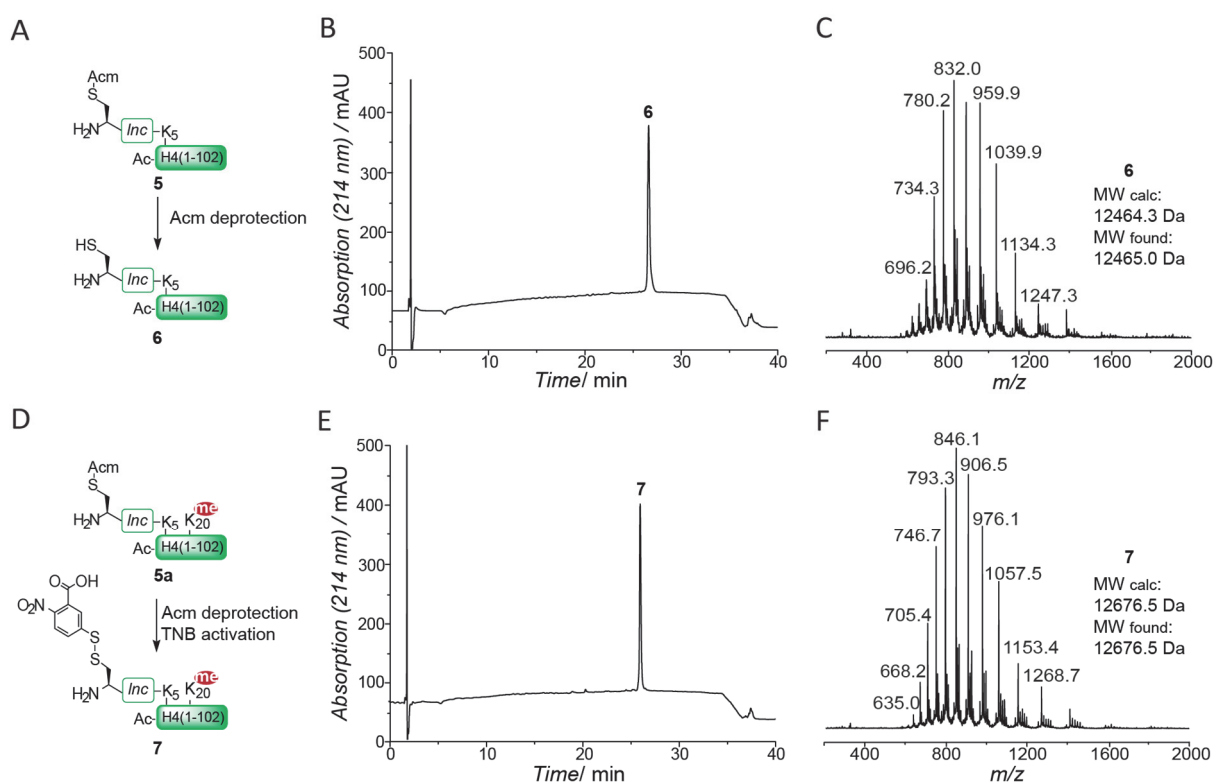
Figure 43: **Synthesis of H4 proteins 6 and 7 by Acm deprotection and TNB activation**. **A**) Scheme of the reaction yielding product **6**. **B**) Analytical RP-HPLC chromatogram of purified **6**. **C**) ESI-MS analysis of product **6** (MW calculated: 12464.3 Da, MW found: 12465.0 Da). **D**) Scheme of the reaction yielding product **7**. **E**) Analytical RP-HPLC chromatogram of purified **7**. **F**) ESI-MS analysis of product **7** (MW calculated: 12676.5 Da, MW found: 12676.5 Da).

3.4.4 Formation of ^{xlnc}H4K20me1 heterodisulfide dimers

To prepare asymmetrically modified H4 nucleosomes, I then continued to form the heterodisulfide dimers **8**, containing one H4 protein with and one without monomethylated K20, connected via the ^{iso}lnc-tag. To this end, I combined **6** and **7** in denaturing buffer (6M GmdCl) at pH 6, with **7** in 1.1-fold excess. The reaction was further allowed to proceed for 30 s and immediately quenched to low pH followed by HPLC purification, to avoid disulfide exchange (**Figure 44**). The asymmetric disulfide dimers **8** (^{xlnc}H4K20me1) were obtained in 45% yield and stored as lyophilized powder for nucleosome assembly. I thus established an efficient synthesis for asymmetrically modified H4, employing two one-pot ligation-desulfurization steps, followed by a one-pot deprotection/activation step of the ^{iso}lnc-tag and a final disulfide dimers formation reaction.
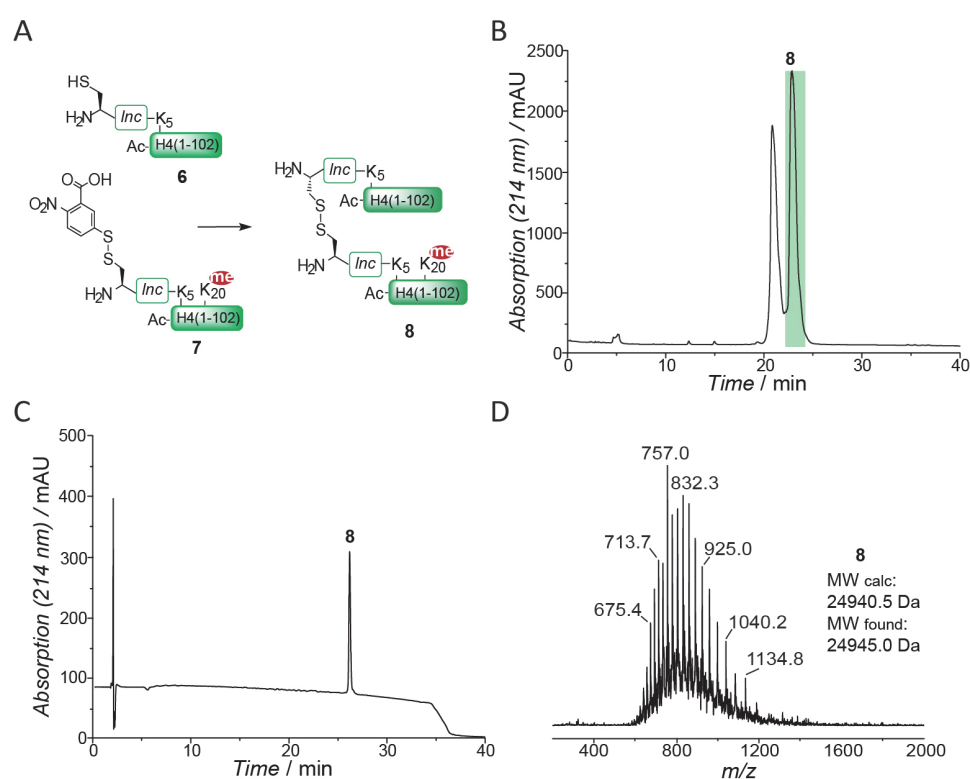
Figure 44: Synthesis of asymmetric ^{xlnc}H4K20me1 heterodisulfide dimers 8. A) Scheme of the reaction yielding heterodisulfide dimers **8**. **B**) Semipreparative RP-HPLC chromatogram of the formation of the asymmetric disulfide ^{xlnc}H4K20me1 **8**. The green rectangle corresponds to the pooled fraction. **C**) Analytical RP-HPLC chromatogram of the final purified product **8**. **D**) ESI-MS analysis of product **8** (MW calculated: 24940.5 Da, MW found: 24945.0 Da).

3.4.5 Synthesis of H4K20me1 histones for symmetric H4K20me1 nucleosomes

In order to synthesize the control symmetric nucleosomes labeled with K20me1 at both H4, peptide hydrazide **1'**, corresponding to the N-terminal acetylated H4(1-14) tail and devoid of the *lnc-tag*, was synthesized by Fmoc-SPPS and ligated to peptide **2a**. Following purification, a second ligation of the H4

fragment **3a'^A15C** to truncated H4 **4** and subsequent desulfurization of both junction cysteines afforded the full length H4K20me1 histone protein **5a'** (**Figure 45**).
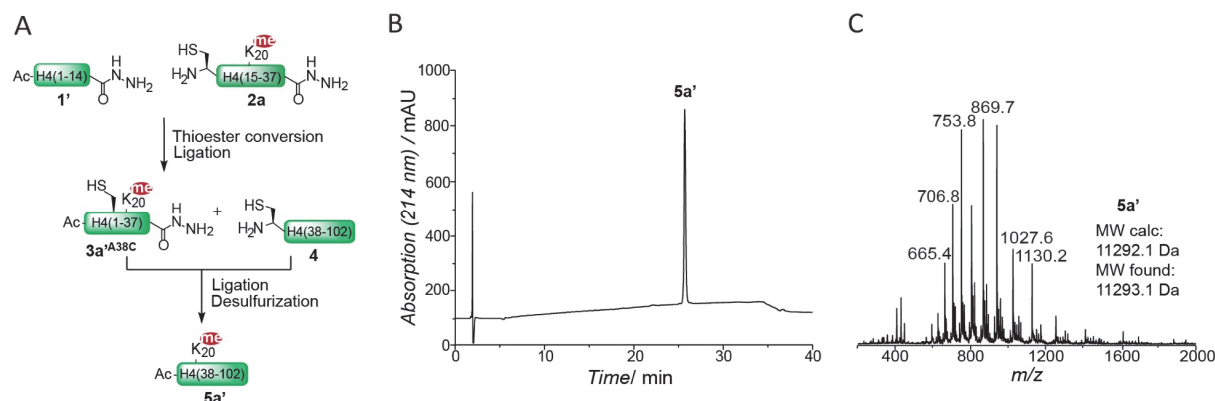
Figure 45: **Synthesis of protein 5a'**. **A**) Scheme of the reaction steps yielding product **5a'**. **B**) Analytical RP-HPLC chromatogram of purified desulfurized H4 protein **5a'**. **C**) ESI-MS analysis of **5a'** (MW calculated: 11292.1 Da, MW found: 11293.1 Da).

3.4.6   Refolding of histone octamers and reconstitution of nucleosomes

Having [xlnc]H4K20me1 dimers **8** in hand, I could assemble asymmetrically modified nucleosomes. I thus combined 1 eq. **8** with 2 eq. of recombinant human histones H3 and 2.2 eq. of H2A and H2B, under denaturing conditions, and refolded histone octamers by dialysis into native buffer, containing 2M NaCl. The reconstituted histone octamers were subsequently purified by gel filtration chromatography (**Figure 46**). Importantly, the crosslinked [xlnc]H4 dimer **8** did not impair octamer formation as judged by the symmetrical elution peak from the gel filtration column, at an elution volume corresponding to the expected molecular weight for a histone octamer (**Figure 46A**). SDS-PAGE of the elution peak verified the equimolar presence of all histones (**Figure 46B**). The addition of a reduction agent (dithiothreitol, DTT) readily reduced the disulfide bond, recovering monomeric [lnc]H4 (**Figure 46B**).
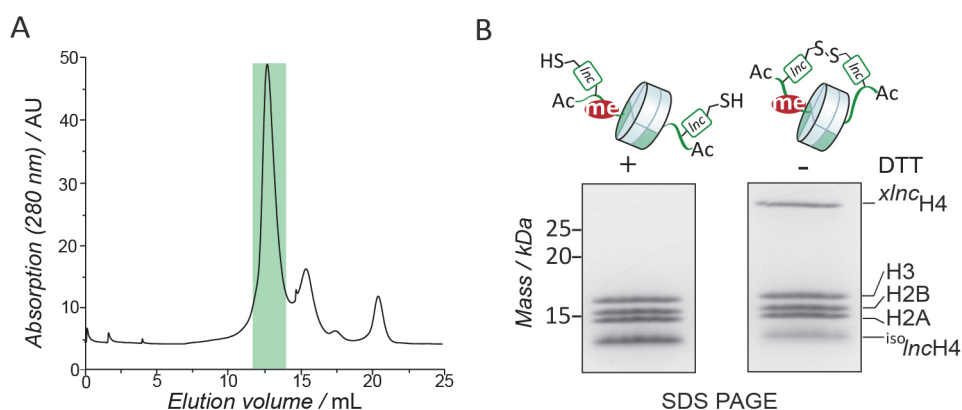
Figure 46: Heterodisulfide dimers purification and analysis. A) Gel filtration analysis of histone octamers containing asH4K20me1. **B)** SDS-PAGE analysis of [xlnc]H4K20me1-containing histone octamers with and without DTT.

63

Concomitantly, symmetric H4K20me1 and unmodified octamers were produced following the same procedure. I then employed the purified octamers to reconstitute nucleosomes (**Figure 47**).

Both the oxidized form ˣˡⁿᶜH4K20me1 as well as the reduced ˡⁿᶜH4K20me1 octamers resulted in defined nucleosomes, demonstrating that the crosslinked H4 tails did not interfere with DNA wrapping (**Figure 47A**). To finalize our synthesis, I then treated the reconstituted nucleosomes with TEV protease, which proceeded to remove the *isolnc*-tag over the course of 16 h, yielding fully native nucleosomes containing asymmetric H4K20me1 (**Figure 47B**). In summary, we established a methodology to control the incorporation of differentially modified H4 into single nucleosomes.
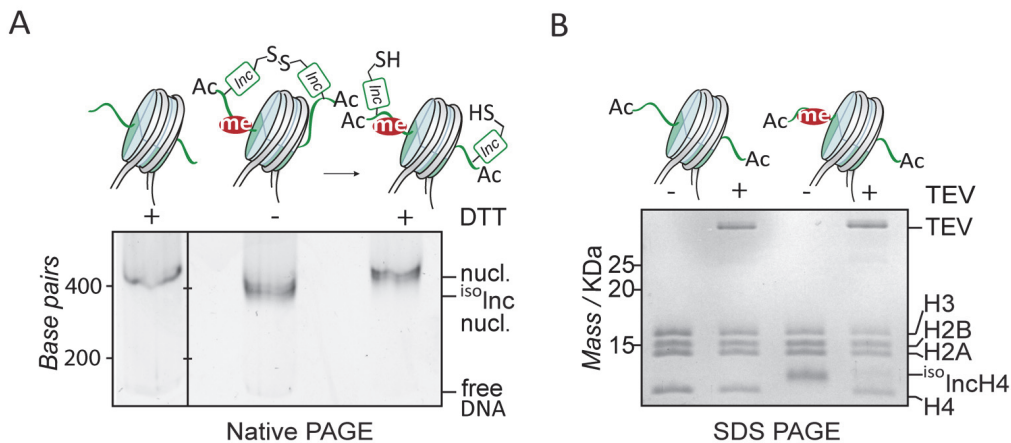
Figure 47: Asymmetric nucleosomes reconstitution, disulfide bond reduction and TEV cleavage of the *isolnc*-tag. A) Native-PAGE analysis of ˣˡⁿᶜH4 nucleosomes with and without reducing agent. **B**) SDS-PAGE analysis of the *isolnc*-tag removal by TEV protease.

3.4.7 Set8 activity is not influenced by pre-existing H4K20me1 modifications

Finally, we employed both unmodified and asymmetric H4K20me1 nucleosomes to study the regulation and enzymatic mechanisms of Set8 in establishing an asymmetric chromatin state. To this end I performed methyltransferase assays using recombinant Set8 together with ^{3}H-S-adenosyl methionine as a cofactor, employing fluorography and scintillation counting as readout. In endpoint experiments (**Figure 48A**) I observed that asymmetric H4K20me1 nucleosomes resulted in about two-fold reduced ^{3}H incorporation. In contrast, synthetic symmetrically H4K20me1 modified nucleosomes (symH4K20me1) were no substrate, demonstrating the specificity of Set8 for H4K20 monomethylation (**Figure 48A**). Kinetic experiments further showed that the degree of reduction of Set8 activity on asymmetric H4K20me1 nucleosome remained constant over the whole time range (**Figure 48B**).

Our results indicate that a pre-existing H4K20me1 modification within a nucleosome is not influencing the enzymatic activity of Set8 for the other H4 tail (**Figure 48C**). This is in contrast to other methyltransferases that are activated by their own enzymatic product, including PRC2[102,104,322] and SUV39H1.[144] Together with structural models of the Set8-nucleosome complex[345] we conclude that Set8 accesses both faces of the

nucleosome, and thus both H4 tails, independently. Thus, Set8 has an intrinsic propensity to produce asH4K20me1 nucleosomes, as observed in ES cells.
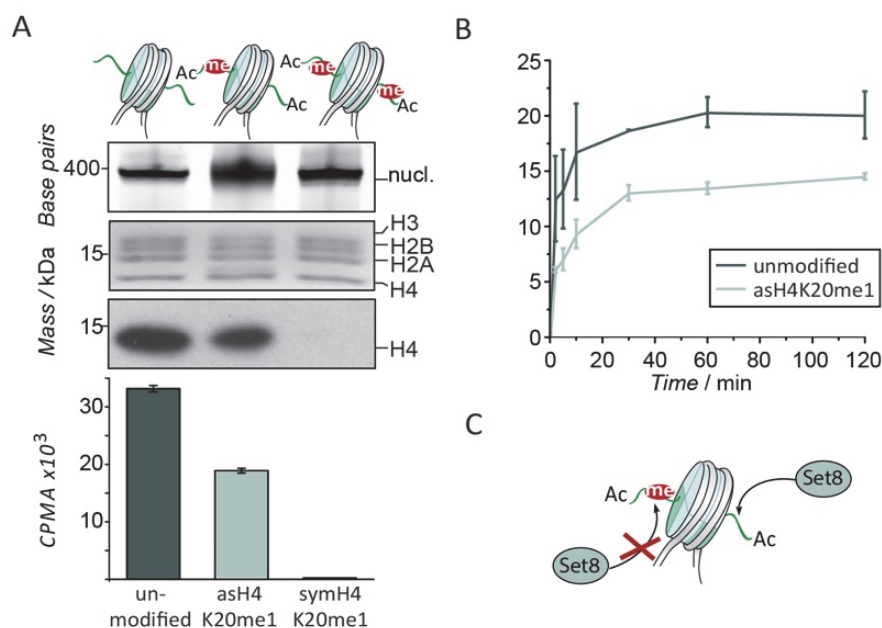
Figure 48: Methyltransferase assays demonstrate reduced activity of Set8 towards asymmetric H4K20me1 nucleosomes. A) End point experiments at t= 2h. From left to right: unmodified, asymmetric H4K20me1 and symmetric H4K20me1 nucleosomes. From top to bottom: native-PAGE, SDS-PAGE, fluorography, scintillation counting. **B)** Kinetic analysis of Set8 methyltransferase activity towards asymmetric H4K20me1 nucleosomes compared to unmodified nucleosomes. CPMA = counts per minute for alpha radiation. Error bars represent standard deviation. n=3. **C)** Model of Set8 nucleosome methylation.

3.5   Conclusions

In summary, we have developed chemical access to nucleosomes carrying asymmetric modifications on H4. Together with our previously reported syntheses,[104,309] we have now assembled a library of asymmetric nucleosomes comprising a large fraction of asymmetric PTMs found in stem cells,[102] which sets the stage for mechanistic studies of stem cell specific PTM regulation. Our method further enables free modification of the histone N-terminus, due to the use of the *isoInc*-tag. This is of particular importance as histone H4 (similarly to other histones, such as H2B) is naturally acetylated at the N-terminus. In addition, the design of a three-segment ligation ensures a highly modular strategy and enables the assembly of combinatorially modified asymmetric H4 nucleosomes in a straightforward manner. Moreover, we note that TEV-protease cleavable peptides installed at lysine side chains provide means to introduce a wide variety of tags at internal sites into proteins that then can be removed in a traceless manner. Together, we anticipate that our synthetic approach will have wide application for the study of chromatin reader interactions with different H4 PTM patterns, including acetyl-readers such as BET-bromodomain proteins[365], or methyl readers such as 53BP1.[65] Finally, our synthesis is not limited to histone proteins, but can be generalized to control the supramolecular assembly of general protein complexes.

As anticipated in the previous chapter, the next challenge in studying the biochemical and biophysical functions of chromatin - chromatin effectors interactions is having access to clean, efficient and selective protein labeling methods. In addition, the stabilization of dyes towards photobleaching is key for single-molecule investigation of these binding events. In the following chapter, I report on how we contributed to this with the development of a "doubly orthogonal" labeling method, adding an interesting and innovative strategy to the cysteine labeling toolbox.

3.6 Acknowledgements

Chapter 4: Development and Application of a Doubly Orthogonal Labeling Strategy for Protein Modification

4.1 Outline

The biophysical study of chromatin dynamics and binding interactions with chromatin effectors requires efficient and specific protein labeling methods. New bioconjugation strategies that provide further versatility in modifying proteins are therefore needed. Herein, I report a dual-orthogonal cysteine bioconjugation methodology for the introduction of hypervalent iodine compounds onto biomolecules. Ethynylbenziodoxolones (EBXs) engage thiols in small organic molecules and cysteine-containing peptides and proteins in a fast and selective addition onto the alkynyl triple bond, resulting in stable vinylbenziodoxolone hypervalent iodine conjugates. The use of an azide-bearing EBX reagent enables a "doubly orthogonal" functionalization of the bioconjugate via strain-release driven cycloaddition and Suzuki-Miyaura cross-coupling of the vinyl hypervalent iodine bond.

Indeed, the methodology was successfully applied on relevant and complex biomolecules such as histone proteins and nucleosomes. The potential of this doubly-reactive bioconjugate was illustrated with the introduction of a triplet-state quencher close to a fluorophore, extending its lifetime through suppression of photobleaching, as demonstrated on single molecule experiments. This work is therefore expected to find broad applications for peptide, proteins and chromatin fibers functionalization.

4.2 Contributions

This chapter is based on the following publication:

Tessier, R., Ceballos, J.,* Guidotti, N.,* Simonet-Davin, R., Fierz, B. & Waser., J., (**2019**). "Doubly orthogonal labeling of peptides and proteins. *Chem*, *5*(8), 2243-2263.
* These authors contributed equally to the work.

Contributions by other authors are described, where appropriate, along the text and in figure legends. In particular:

- T. R. discovered and designed the reaction, synthesized the reagents and substrates, performed the optimization studies, the scope of the reaction and modification studies on short peptides.
- C. J. synthesized the reagents and substrates and performed the modification studies on short peptides.
- G. N. led several practical courses with students that contributed to the development of the reaction on glutathione and larger peptides. G. N. designed and performed the reactions on larger peptides, modified

ubiquitin, nucleosomes and Substance P, performed the cell experiments and single-molecule photobleaching studies.

- S-D. R made several experiments important for the discovery and optimization of the reaction on glutathione.

4.3 Introduction

Efficient, selective and flexible labeling techniques are essential for the study and manipulation of proteins. Novel bioconjugation techniques are still in high demand, but their development is complicated by the need for fast kinetics, high efficiency and excellent selectivity under mild conditions. Nevertheless, several labeling methods have emerged, exploiting both natural and unnatural amino acids.

Because of its key role in the structure and catalytic activity of proteins, cysteine has been one of the most studied natural amino acids in chemical biology. Its low abundance in proteins and its intrinsic high nucleophilicity makes it an ideal target for chemoselective modification techniques often based on site-directed mutagenesis.[366,367,368] Accordingly, significant efforts have been made on the development of reagents enabling cysteine labeling with high selectivity and efficiency. Among those, iodoacetamides, maleimide derivatives, the thiol-ene and -yne reactions,[207,369,210,370] the oxidative elimination of cysteine to dehydroalanine followed by Michael addition,[366,371] metal-free arylations,[213,372] or metal-assisted functionalizations[214,94,373,374] are the most frequently employed strategies. Although their reactivity has been extensively studied, significant limitations persist, such as lack of chemoselectivity, product instability or uncontrolled reactivity (see **Chapter 1.3.2**). In addition, many of these methods are based on chemical compounds requiring challenging preparations, careful storage and/or stringent use of inert atmosphere.

In this context, hypervalent iodine reagents are outstanding compounds as they combine high reactivity with sufficient stability and low toxicity.[375] Iodine is a large and highly polarizable atom with the lowest electronegativity of its group. The most common hypervalent iodine compounds are either λ^3-iodanes (trivalent iodine derivatives with oxidation state +III) or λ^5-iodanes (oxidation state +V). The overlap of the 5p orbital of the iodine and the orbitals of the two ligands leads to the formation of a weak, long and highly polarizable linear three-center-four-electron bond (L-I-L).[376] The presence of this hypervalent bond in λ^3-iodanes is responsible for the high electrophilicity and reactivity of such compounds. Indeed, alkynyl-hypervalent iodine compounds have revealed to be excellent reagents for electrophilic alkynylation reactions, where the classical nucleophilic nature of the alkyne ligand is inverted (umpolung), allowing the electrophilic introduction of alkynes under mild conditions.

The Waser lab previously reported a highly efficient and chemoselective alkynylation of thiols using hypervalent iodine-based Ethynylbenziodoxolone reagents (EBXs) in organic solvents,[377,378,379] which was also successful for amino acids and dipeptides (**Figure 49, top**).[219,220] In addition, in collaboration with the

Adibekian group, they successfully applied JW-RF-010 (**1**), an azide-bearing EBX, in the labeling of native cysteines *in vitro* and *in vivo* to afford alkynes as major products (**Figure 49, bottom**).[221] JW-RF-010 (**1**) displayed exceptional stability and chemoselectivity, outperforming iodoacetamide, the gold standard in cysteine targeting. However, only hyper-reactive cysteines were efficiently functionalized in aqueous media, making the method not general. In addition, the reaction between EBX and proteins was not thoroughly investigated, as JW-RF-010 (**1**) was applied on cell lysate for proteomic studies only. Indeed, I set out to characterize in detail and develop applications of the labeling strategy on peptides and proteins.

Figure 49: Previous work on cysteine alkynylation with EBX reagents.

Common to all reported approaches for biomolecule functionalization with hypervalent iodine reagents is the use of the inherent reactivity of the hypervalent bond to perform the bioconjugation step. In contrast, I present herein a highly efficient, fast, chemoselective and clean labeling of cysteine-containing peptides and proteins giving stable adducts of EBX and the thiols without cleavage of the hypervalent bond (**Figure 50**). This methodology can be applied to all cysteine- and other thiol-bearing compounds, including modified ubiquitin and histone octamers. The unprecedented peptide- and protein bound vinylbenziodoxolone reagents are stable, yet their inherent reactivity[380,381,382] opened the way for bioconjugations orthogonal to natural functional groups existing in biomolecules. Here, I report a palladium catalyzed Suzuki-Miyaura cross-coupling reaction[383,384,385,386] selective at the vinyl hypervalent iodine bond (**Figure 50**).

Efficient, robust, fast and under physiological conditions, this methodology allows swift peptide and protein modifications, without the need for incorporation of unnatural amino acids. In addition, we have combined this powerful methodology with metal-free Strain-Promoted Alkyne-Azide Cycloaddition

(SPAAC),[387,388] which allowed functionalization of the adduct without cleaving the hypervalent bond, leading to "doubly orthogonal" modifications. I demonstrated the potential of the method by introducing both a fluorophore and a triplet state quencher on neuropeptide substance P, enabling single molecule fluorescence experiments with increased sensitivity.
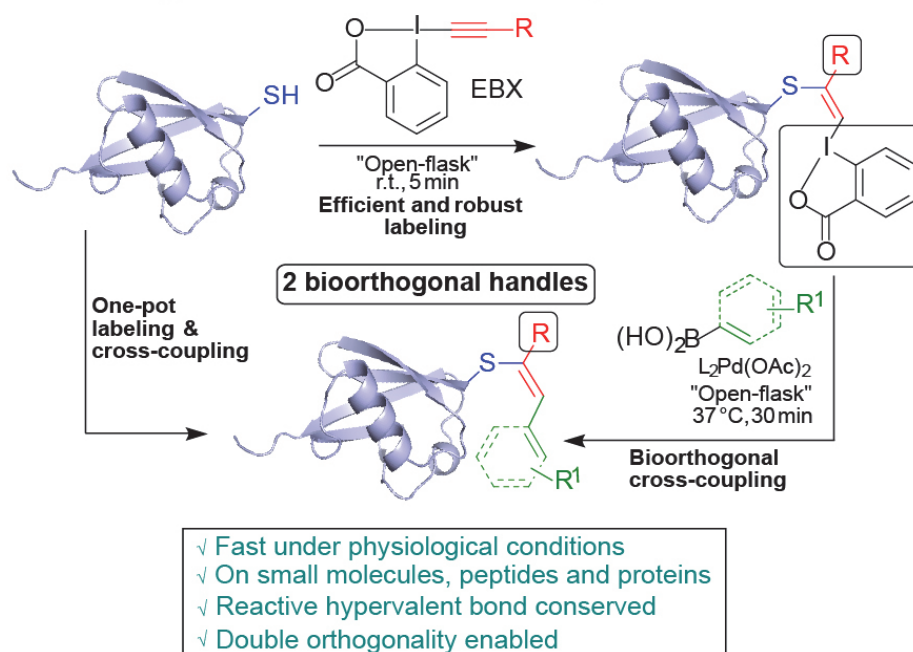
Figure 50: Thiol labeling approach presented in this work. The hypervalent iodine bond is conserved and it enables doubly orthogonal functionalization. Ubiquitin crystal structure is shown in purple (PDB: 1UBQ).

4.4 Results

4.4.1 Discovery, optimization and scope of the hypervalent iodine transfer reaction

The discovery and investigation of the reaction conditions on glutathione and small peptides was performed by T. R. , C. J. and S-D. R. and will be briefly summarized in the following paragraph.

The tripeptide glutathione (GSH) was chosen as a first case-study thiol-containing molecule as it is a highly abundant molecule in cells that plays key roles as disulfide-bond reducing agent. When GSH was treated with **1** in Tris buffer, pH 8.2, the vinylbenziodoxolone (VBX) adduct was detected as the only product, instead of the expected alkynyl sulfide compound (**Figure 51**). In fact, computational studies on the mechanism of the reaction revealed that in organic solvents the vinylic carbanion intermediate, derived from concerted β-addition of the sulfide to the EBX reagent, preferentially undergoes alpha elimination

followed by 1,2-sulfur shift to give an alkynyl sulfide.[389] However, in aqueous environment, the vinyl carbanion is directly protonated, yielding the VBX (**Figure 51**).[390]
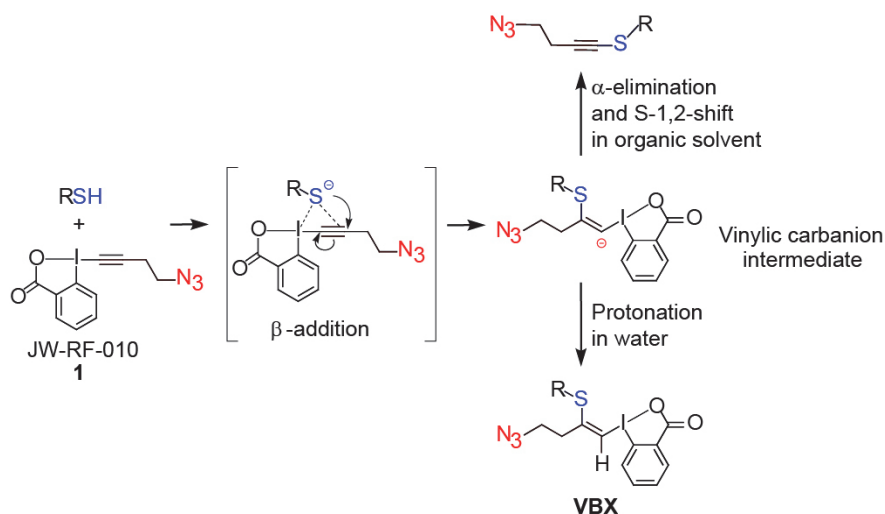
Figure 51: **Speculative reaction mechanism for the formation of vinylbenziodoxolone (VBX).**

The excellent regio- and stereoselectivity,[389] the nonrequirement for additional reagents or catalysts and the absence of byproducts formation made this transformation a potential precious tool for protein bioconjugation. Indeed, further investigations highlighted the high stability of the adduct over a wide range of pH and temperature conditions, as well as its great tolerance towards excess external thiol nucleophiles and reducing agents, such as TCEP. The pH was found to be a crucial parameter for the reaction. High yields (87%) and fast kinetics (comparable to iodoacetamide) were observed at pH 8.2 and, although they slightly decreased at physiological pH, 74% conversion was obtained after 1h incubation at pH 7.4. Importantly, reaction between GSH and a library of hypervalent iodine reagents bearing different functional groups, such as azides, alkynes, alkenes, halogens and alcohols substituents, rapidly afforded the VBX product in excellent yields. Further, a library of cysteine-containing tetrapeptides allowed the investigation of the scope of the thiol reagents. Most amino acid side chains were well tolerated, including Asp, Asn, Met, Ser, Tyr, Trp, Pro, the sterically hindered Phe, Val and Ile and the nucleophilic His and Lys. Only Arg residues showed lower yields due to the formation of side products. All together, these promising results prompted us to apply this labeling technique to more complex biomolecules, such as longer peptides and folded proteins.

4.4.2 Labeling of longer peptides with EBX

Two case study longer peptides were used to test the labeling efficiency (specificity, ease of use, etc.) of our bioconjugation method. The first corresponds to the Human Serum Albumin (HSA) fragment Leu_{55}-His_{63} (**2**), bearing a native cysteine at position Cys_{58}, while the second one is the so-called Substance P (SP, **4**), a neuropeptide and high-affinity ligand of the neurokinin 1 receptor (NK1-receptor)[391] carrying an additional cysteine residue close to the N-terminus. Treatment of excess HSA fragment under native buffer

conditions with EBX reagent afforded 88% conversion after 1h incubation at RT (**Figure 52A, B, C**). Of note, due to its chromatographic properties, the EBX reagent was kept as limiting reagent to allow easy discrimination of reagents and product, since similar retention times were observed for both EBX reagent **1** and HSA-EBX adduct **3**. On the other hand, slight excess EBX was reacted with SP **4** in 50 mM phosphate buffer pH 8.2 to obtain SP-EBX product **5** after 1h at RT with 81% yield. As confirmed by RP-HPLC analysis, the reaction proceeded very cleanly with the oxidized disulfide SP dimer being the only minor side product (**Figure 52D, E, F**).
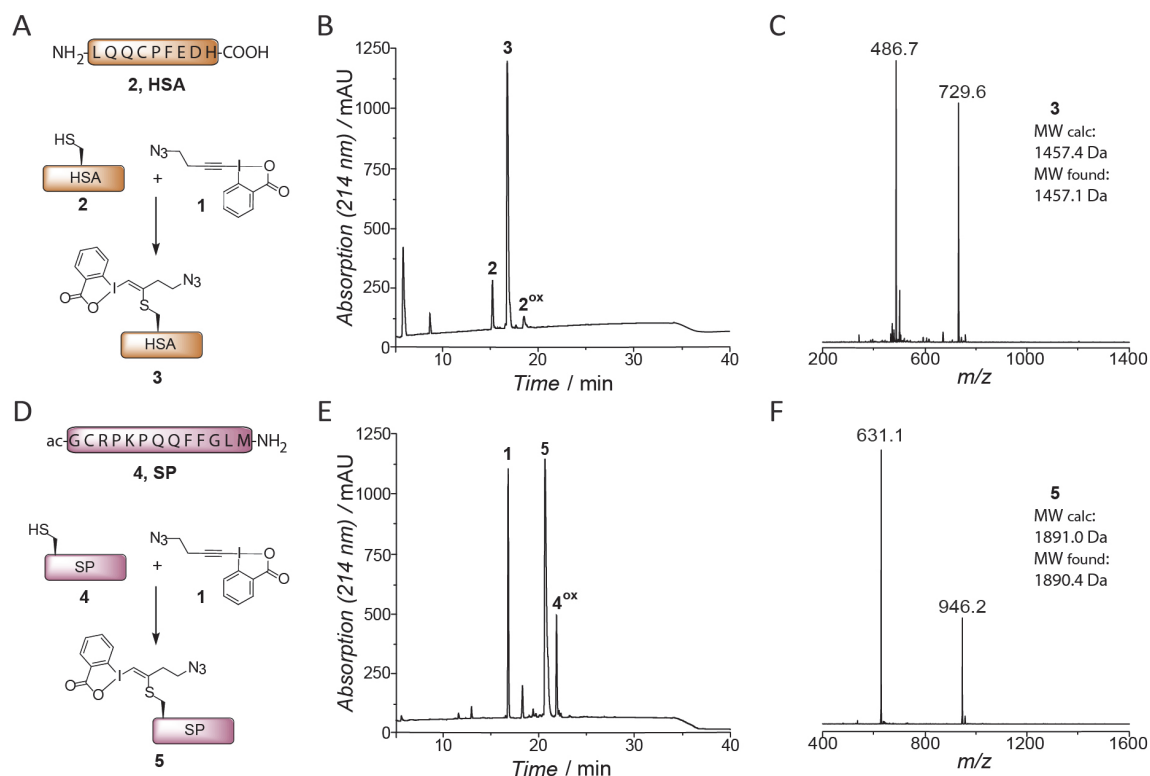
Figure 52: Labeling of peptides with reagent 1. A) Scheme of the reaction between **1** and peptide HSA **2**. **B**) Analytical RP-HPLC chromatogram of the reaction mixture after 1h incubation at RT. **C**) ESI-MS analysis of product **3** (MW calculated: 1457.4 Da, MW found: 1457.1 Da). **D**) Scheme of the reaction between **1** and peptide SP **4**. **E**) Analytical RP-HPLC chromatogram of the reaction mixture after 1h incubation at RT. **F**) ESI-MS analysis of product **5** (MW calculated: 1891.0 Da, MW found: 1890.4 Da).

4.4.3 Labeling of Cys-ubiquitin with EBX

We then decided to employ our method to modify cysteines in a protein context. As a model system I chose the protein ubiquitin, carrying an N-terminal hexahistidine-tag, followed by a single cysteine residue (His6-Cys-ubiquitin, **6**). To our delight, treatment of **6** with only 2 equivalents of JW-RF-010 (**1**) under native conditions afforded a quantitative transformation into VBX **7** in less than an hour (**Figure 53A, B, C**). To the best of my knowledge, this constitutes the first example of a protein-bound hypervalent iodine reagent. This efficient conjugation was also compatible with a protein denaturing buffer (**Figure 53D, E, F**).
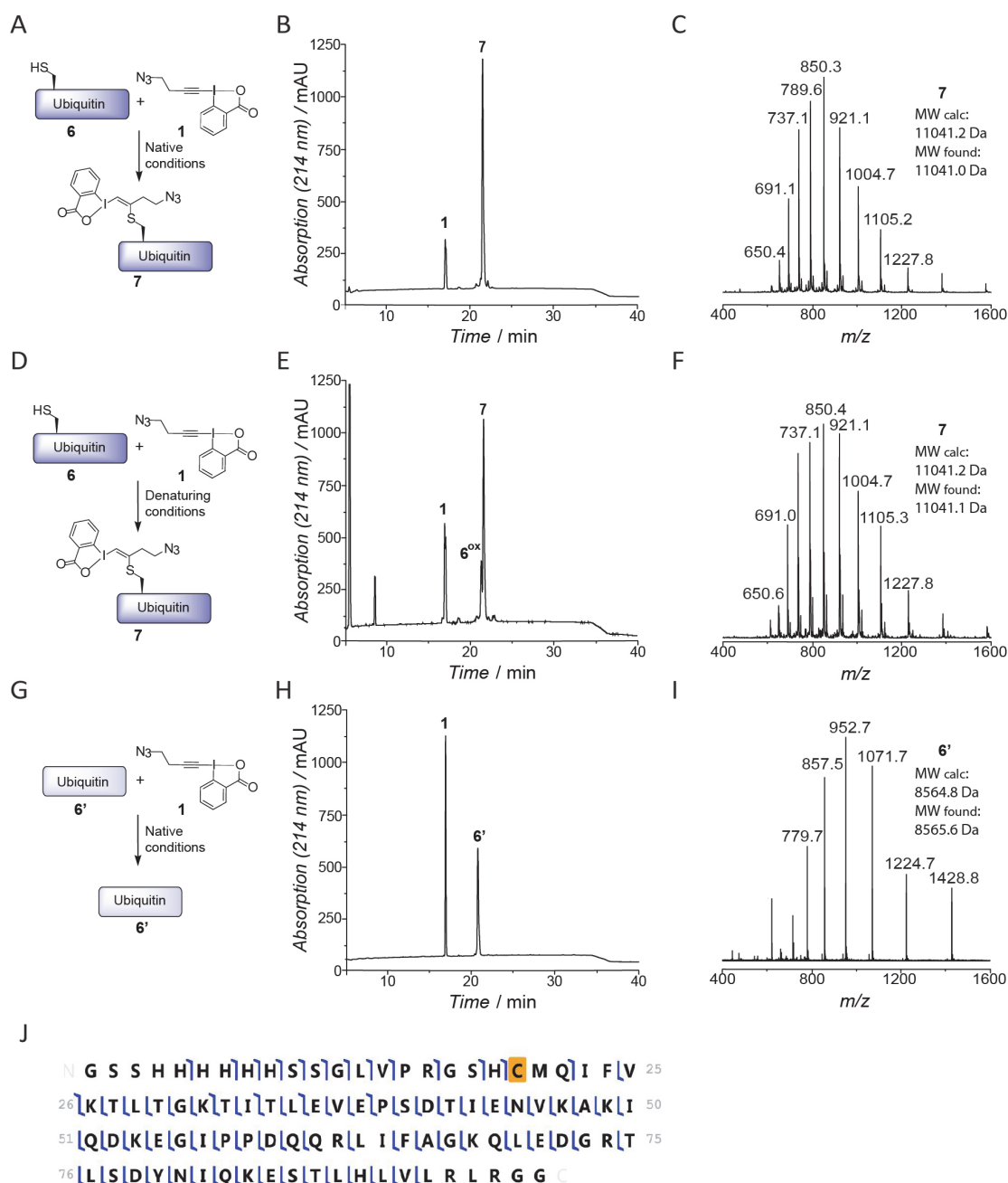
Figure 53: Labeling of ubiquitin with reagent 1. A) Scheme of the reaction between **1** and ubiquitin **6** under native conditions. **B)** Analytical RP-HPLC chromatogram of the reaction mixture after 1h incubation at RT. **C)** ESI-MS analysis of product **7** (MW calculated: 11041.2 Da, MW found: 11041.0 Da). **D)** Scheme of the reaction between **1** and ubiquitin **6** under denaturing conditions. **E)** Analytical RP-HPLC chromatogram of the reaction mixture after 1h incubation at RT. **F)** ESI-MS analysis of product **7** (MW calculated: 11041.2 Da, MW found: 11041.1 Da). **G)** Scheme of the reaction between **1** and the control ubiquitin **6'**. **H)** Analytical RP-HPLC chromatogram of the reaction mixture after 2h incubation at RT. **I)** ESI-MS analysis of ubiquitin **6'** after 2h incubation at RT with **1** (MW calculated: 8564.8 Da, MW found: 8565.5 Da). **J)** Fragmentation map of product **7**. Obtained sequence coverage was 80% with p-score of $7e^{-24}$. Assigned b- and y-fragmentation ions (in blue) confirm the cysteine-specific modification. The presence of key fragments b_{19}, b_{22}, y_{72} and pair b_{25}/y_{71} was validated manually.

Of note, side reactions due to arginine were not detected in ubiquitin, which contains five arginine residues. We thus conclude that such side reactions, as observed on tetrapeptides, only occur when arginines are

73

immediately next to the cysteine. Importantly, no reaction was observed with native (cysteine-free) ubiquitin **6'** (**Figure 53G, H, I**) further demonstrating the cysteine-specificity of the reaction in a protein context. Moreover, the correct modification site in ubiquitin was confirmed by top-down mass spectrometry (**Figure 53J**).

4.4.4 Application for fluorescent peptide labeling

Next, I investigated the potential of the obtained bioconjugates for further functionalization. At first, I examined if standard methods for azide modification could be used in the presence of a reactive hypervalent iodine center.

In order to avoid the use of metal catalysts, we chose the well-known copper-free Strain-Promoted Alkyne-Azide Cycloaddition (SPAAC).[388,387] In principle, two strategies could be adopted: 1) labeling of the peptide/protein with JW-RF-010 (**1**), followed by cycloaddition with a functionalized alkyne (**Figure 54A,** "post-click") or 2) prior reaction of JW-RF-010 (**1**) with the alkyne, followed by reaction with the peptide/protein (**Figure 54B,** "pre-click").
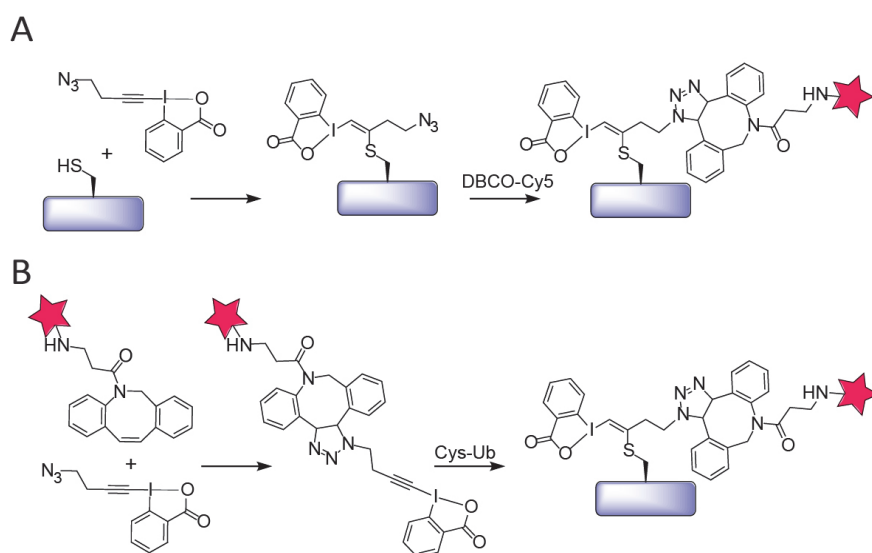
Figure 54: Post-click vs. pre-click labeling strategies. A) Post-click: one-pot cysteine labeling with EBX reagent **1** followed by click reaction with DBCO-Cy5. **B**) Pre-click: one-pot click reaction between EBX reagent **1** and DBCO-Cy5 followed by cysteine labeling.

The former appears easier, as the reactivity of the hypervalent iodine in the VBX product is lower compared to the EBX reagent, whereas the latter is attractive for applications, as it would allow preparation of multi-functionalized labeling reagents *in-situ* starting from **1** and commercially available alkynes. Initial proof-of-concept experiments on GSH with EBX and various dibenzocyclooctynes (DBCO) showed that both post-click and pre-click strategies were successful (performed by T.R.), thus demonstrating that azide-cyclooctyne cycloaddition is fully orthogonal to hypervalent iodine reactivity.
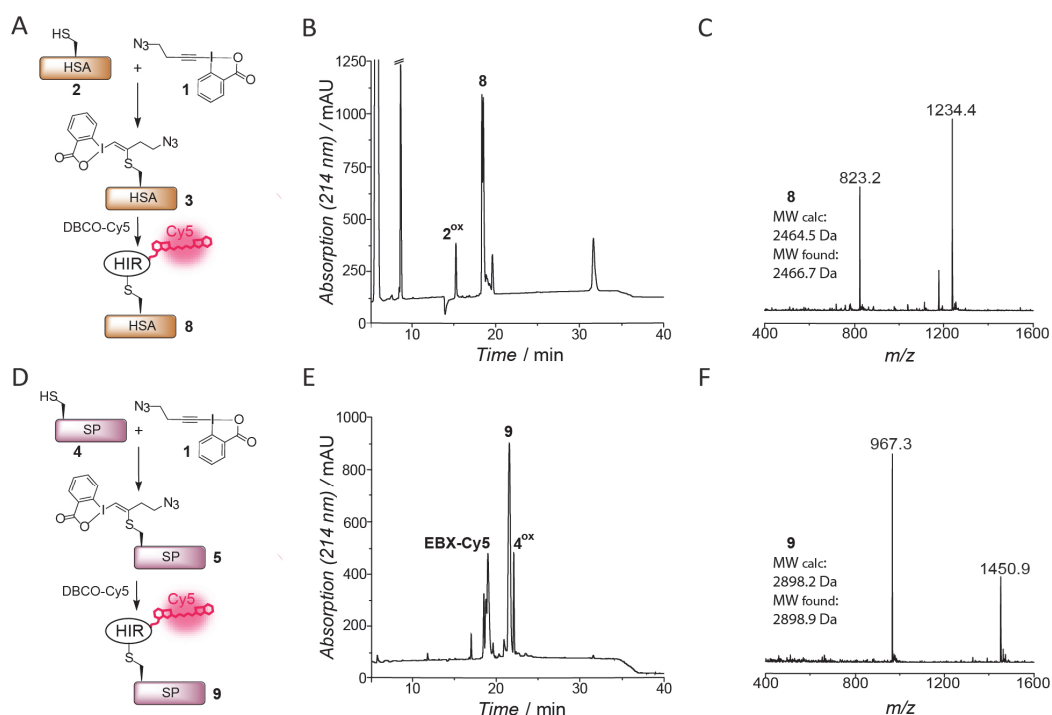
Figure 55: Coupling of EBX labeling and click reaction on peptides (post-click). A) Scheme of the reaction on peptide HSA **2. B**) Analytical RP-HPLC chromatogram of the reaction mixture after reaction between **1** and peptide **2**, followed by click reaction with DBCO-Cy5. **C**) ESI-MS analysis of product **8** (MW calculated: 2464.5 Da, MW found: 2466.7 Da). **D**) Scheme of the reaction on peptide SP **4. E**) Analytical RP-HPLC chromatogram of the reaction mixture after reaction between **1** and peptide **4**, followed by click reaction with DBCO-Cy5. **F**) ESI-MS analysis of product **9** (MW calculated: 2898.2 Da, MW found: 2898.9 Da). HIR: hypervalent iodine reagent.
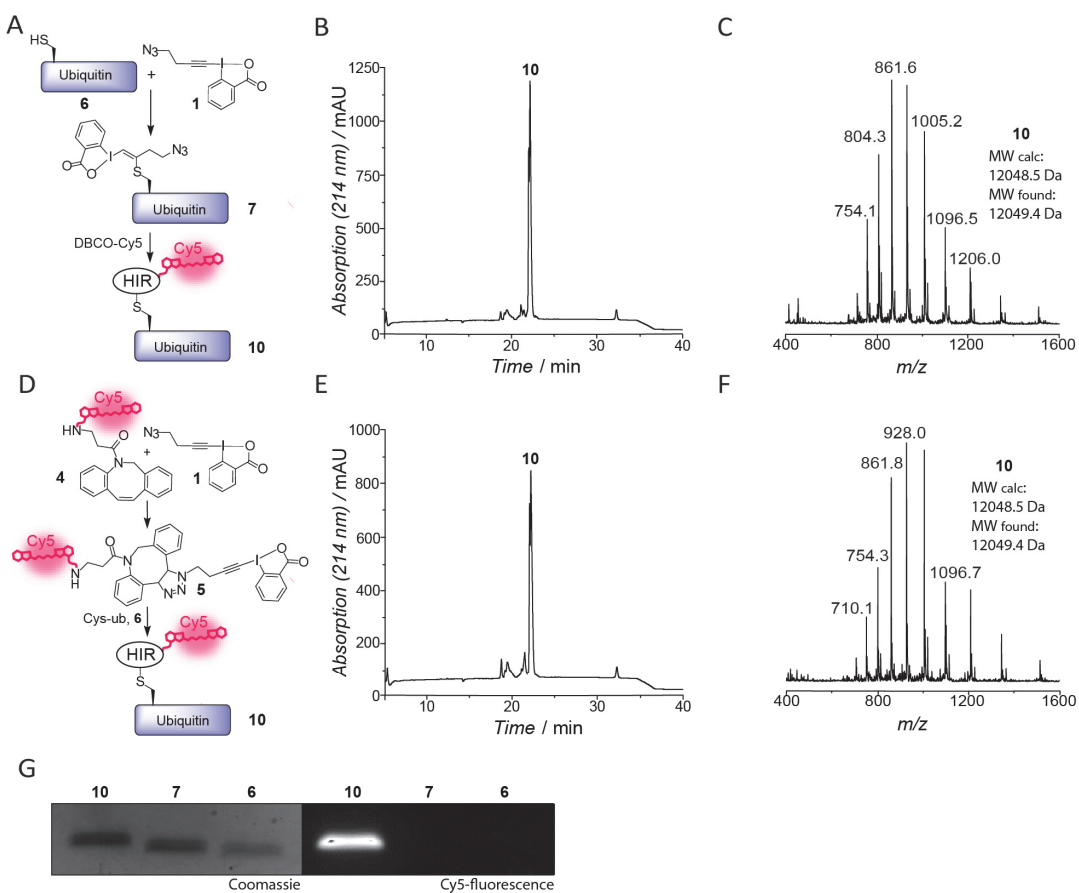
Figure 56 (previous page): Coupling of EBX labeling and click reaction on ubiquitin. A) Scheme of the post-click reaction on ubiquitin **6**. **B**) Analytical RP-HPLC chromatogram of the reaction mixture after post-click reaction between **1**, ubiquitin **6** and DBCO-Cy5. **C**) ESI-MS analysis of product **10** (MW calculated: 12048.5 Da, MW found: 12049.4 Da). **D**) Scheme of the pre-click reaction on ubiquitin **6**. **E**) Analytical RP-HPLC chromatogram of the reaction mixture after pre-click reaction between **1**, DBCO-Cy5 and ubiquitin **6**. **F**) ESI-MS analysis of product **10** (MW calculated: 12048.5 Da, MW found: 12049.4 Da). **G**) SDS-PAGE analysis of post-click reaction on ubiquitin **6**. Left: Coomassie staining, right: Cy5 fluorescence. HIR: hypervalent iodine reagent.

Indeed, one-pot reaction between HSA **2** or SP **4** and EBX **1**, followed by click with DBCO bearing the cyanine dye Cy5 (DBCO-Cy5) was efficiently employed to fluorescently label both peptides with 99% and 82% yield, respectively (**Figure 55**). Further, both approaches of labeling also performed exquisitely well on His-Cys-ubiquitin **6** (post-click: 99%, pre-click: 95% yield, **Figure 56**).

4.4.5 Fluorescent labeling of native histone octamers and nucleosomes

To showcase the compatibility of our methodology with more complex proteins than ubiquitin and determine if the resulting conjugates are well tolerated, I generated fluorescently labeled nucleosomes.
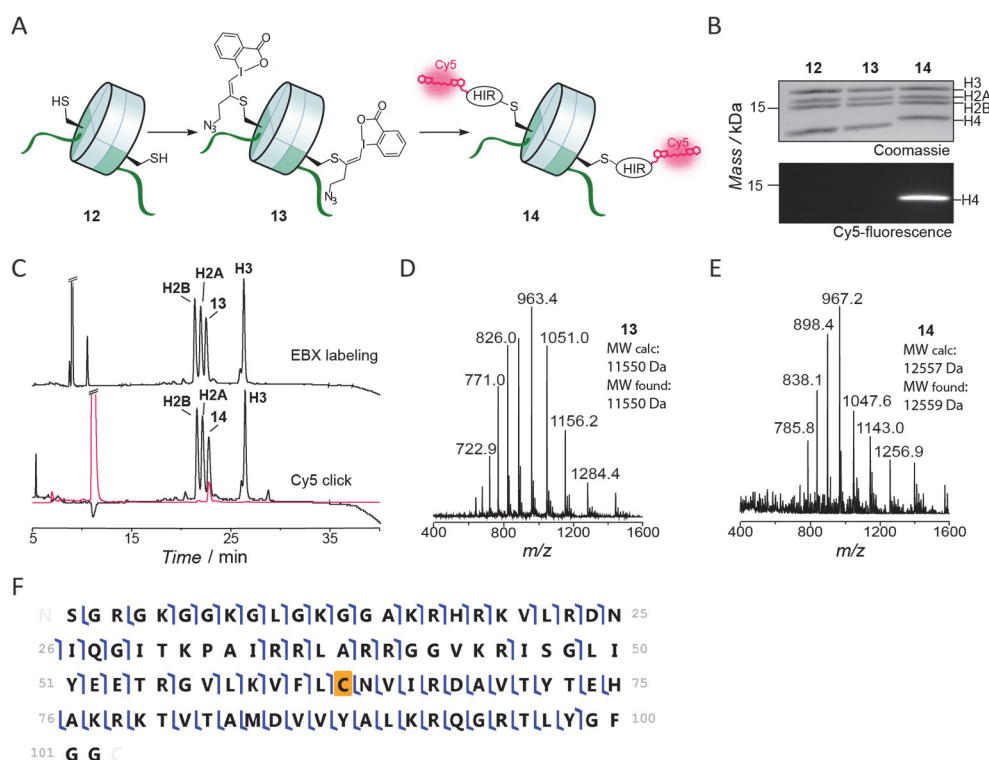
Figure 57: One-pot labeling of histone octamers with EBX reagent 1 and DBCO-Cy5. A) Scheme of the reaction yielding Cy5-labeled histone octamers **14**, including labeling of H4C63 octamers **12** with reagent **1** and click reaction with DBCO-Cy5. HIR: hypervalent iodine reagent. **B**) SDS-PAGE analysis of histone octamers followed by Coomassie staining or imaging of Cy5 fluorescence. **C**) Analytical RP-HPLC chromatogram of the reaction. Top: after 1h incubation at RT of **1** and **12**. Bottom: after 15min incubation at RT of crude **12** and DBCO-Cy5. Black: absorption at 214nm, red: absorption at 647nm. Method: 30-70%B in 30min. **D**) ESI-MS analysis of EBX-octamers **13** (MW calculated: 11550 Da, MW found: 11550 Da). **E**) ESI-MS analysis of Cy5-octamers **14** (MW calculated: 12557 Da, MW found: 12559 Da). HIR: hypervalent iodine reagent. **F**) Fragmentation map of product **13**. Obtained sequence coverage was 75% with p-score of $6e^{-16}$. Assigned b- and y-fragmentation ions (in blue) confirm the cysteine-specific modification. The presence of key fragments b_{62}, y_{39}, b_{98}, y_{99} and y_{101} was validated manually.

For labeling, glutamate 63 in H4 was mutated to cysteine (E63C) and the protein (**11**) was expressed and purified. I then assembled histone octamer complexes, using the remaining other human histone proteins (H2A, H2B and H3C110A, **Figure 57A, B**). The resulting protein complex **12** contained two cysteine residues, which efficiently reacted with **1** within 60 minutes under aqueous conditions to yield **13** (**Figure 57A, B, C, D**). I then successfully labeled **13** with Cy5-DBCO via azide-cyclooctyne cycloaddition to afford product **14** (**Figure 57A, B, C, E**). SDS-PAGE analysis of the product **14** revealed the high specificity of the labeling procedure, as only histone H4 exhibited a fluorescence signal (**Figure 57B**). The selectivity of the reaction was also corroborated by top-down MS-MS (**Figure 57F**).

To demonstrate that the protein functionality was not affected by the modification I continued to reconstitute full nucleosomes, which will only form if the histone octamer structure is not disrupted. I thus combined **13** with a 170 bp-long segment of the 601 nucleosome positioning sequence[392] and reconstituted nucleosomes via dialysis from high (2M NaCl) to low (10 mM KCl) salt (**Figure 58**). To our delight, nucleosomes were readily formed as judged by native PAGE (**Figure 58B**), yielding both unmodified or EBX-modified (**15**) nucleosomes with equal efficiency. Finally, nucleosome assembly also proceeded smoothly with Cy5 labeled octamers **14** to yield fluorescent nucleosomes **16** (**Figure 58C**). Together, these experiments demonstrate that EBX modification at exposed cysteines is compatible with protein function and a viable strategy for highly efficient protein labeling.
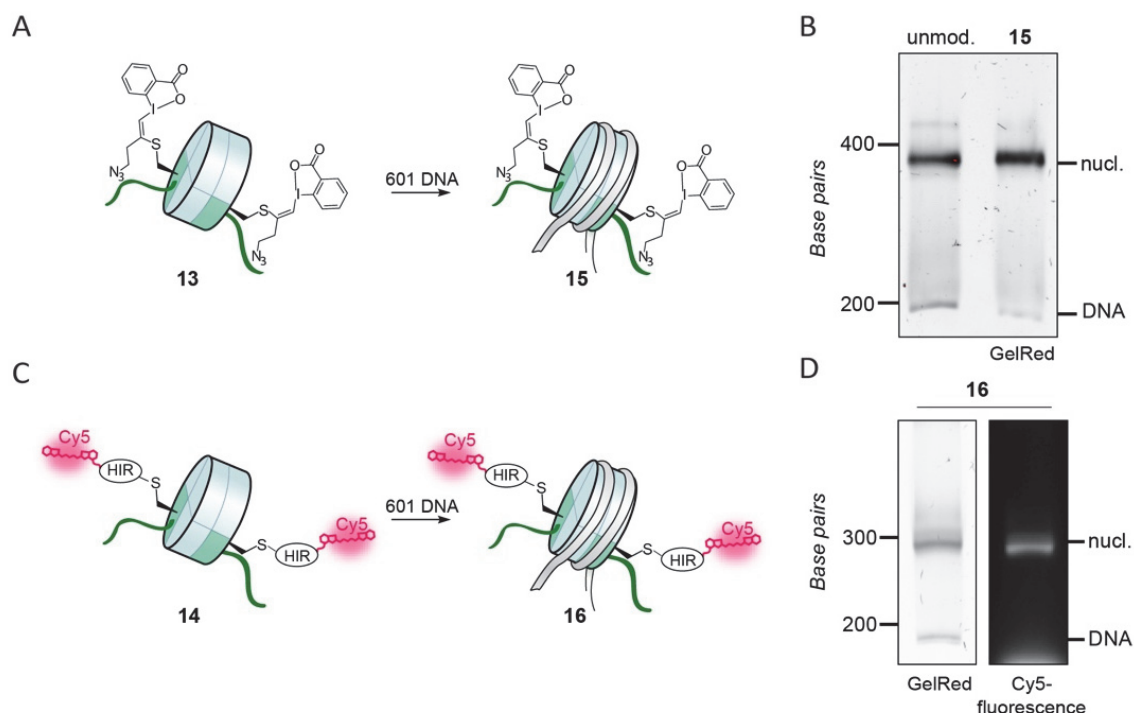
Figure 58: Nucleosome reconstitution. A) Reconstitution of EBX-labeled nucleosomes **15** starting from EBX-labeled octamers **13**. **B**) Native-PAGE analysis of reconstituted unmodified and EBX-labeled nucleosomes **15** (GelRed stain). **C**) Reconstitution of Cy5-labeled nucleosomes **16** starting from Cy5-labeled octamers **14**. HIR: hypervalent iodine reagent. **D**) Native-PAGE analysis of reconstituted Cy5-labeled nucleosomes **16**. Left: GelRed stain, right: Cy5 fluorescence.

4.4.6 Development of "doubly orthogonal" functionalization

The investigation and optimization of the cross-coupling reaction on VBX-GSH was performed by T.R. and will be briefly summarized in the following paragraph.

After having demonstrated that selective reaction of the azide was possible, the functionalization of the hypervalent bond was investigated. Encouraged by recent progress in the use of palladium catalysis for the functionalization of peptides and proteins,[383.384,385,386] a Suzuki-Miyaura cross-coupling was selected as promising candidate. While avoiding transition metals in biomolecule functionalization is preferred, it should be noted that the palladium content can be easily reduced down to 1.0 ppm through scavenging and size-exclusion chromatography.[385] Furthermore, palladium catalysis has been even used in living cells, showing low toxicity.[393,394] After an extensive study on various palladium catalytic systems, air- and moisture-stable bis-lithium-2-(dimethylamino)-4,6-dihydroxylate-pyrimidine palladium diacetate complex **17**[393] was found to be the most suitable catalyst-ligand system for our model reaction between VBX-GSH and boronic acid **18** (**Figure 59**). The reaction proceeded efficiently in an open flask, in non-degassed phosphate buffer at 37 °C (70% yield) and no degradation of the azide group was observed. The reaction also showed excellent tolerance towards different boronic acid substrates, as both electron-rich and electron-deficient phenyl boronic acids were successfully coupled to VBX-GSH.
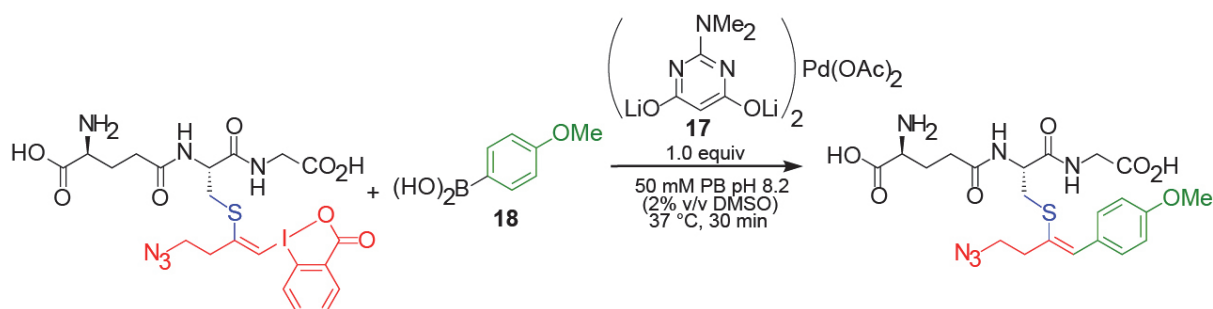
Figure 59: Cross-coupling reaction between VBX-GSH and boronic acid 18, catalyzed by Pd complex 17.

4.4.7 Demonstration of "doubly orthogonal" labeling of ubiquitin

In collaboration with T.R. I then applied the "doubly orthogonal" labeling strategy to the more complex case of modified ubiquitin **7**. To our delight, the cross-coupling worked in a one-pot labeling/cross-coupling approach to give **19** (**Figure 60A, B, C**). In addition, both reactive handles were successfully employed in a labeling/SPAAC/Suzuki-Miyaura one-pot process, on a proof-of-principle scale, to afford doubly-functionalized Cys-labeled ubiquitin **20** (**Figure 60D, E, F**).
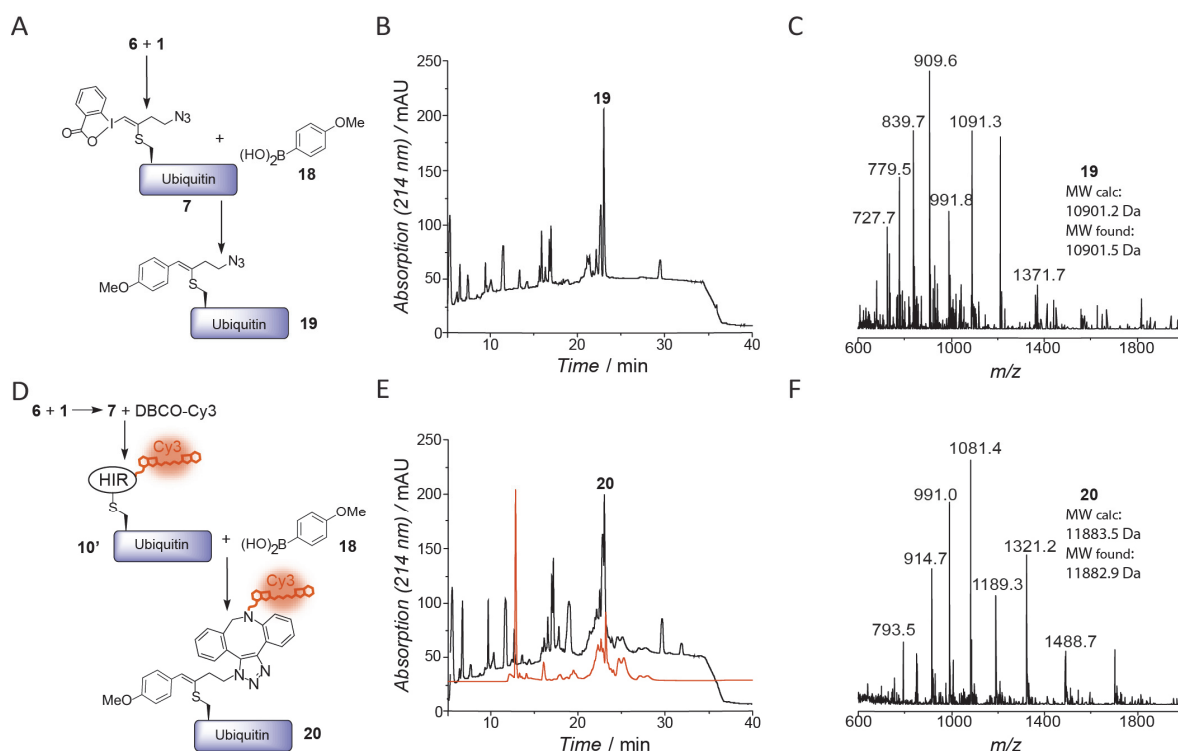
Figure 60: Cross-coupling on ubiquitin. A) Scheme of the one-pot labeling of ubiquitin **6** with EBX **1** and cross-coupling reaction with reagent **18** to yield **19**. **B)** Analytical RP-HPLC chromatogram of the cross-coupling reaction on VBX ubiquitin **7** and boronic acid **18. C)** ESI-MS of product **19** (MW calculated: 10901.2 Da, MW found: 10901.5 Da). **D)** Scheme of the one-pot labeling of ubiquitin **6** with EBX **1**, followed by SPAAC with DBCO-Cy3 and cross-coupling with boronic acid **18** to yield **20**. HIR: hypervalent iodine reagent. **E)** Analytical RP-HPLC chromatogram of the cross-coupling reaction on VBX-Cy3 ubiquitin **10'** and boronic acid **18.** Black: 214nm absorption, Orange: 555 nm absorption. **F)** ESI-MS of product **20** (MW calculated: 11883.5 Da, MW found: 11882.9 Da).

4.4.8 Application of "doubly orthogonal" functionalization to stabilize fluorescent dyes

I then exploited our reagents for fluorescent labeling of receptors on living cells. To this end, the fluorescently labeled SP peptide **9** was incubated with a previously generated human embryonic kidney (HEK) cell line expressing GFP-tagged NK1-receptor at the cell surface (**Figure 61A**).[391,395] Substrate **9** readily labeled the cells and exhibited a distinct colocalization with the GFP-tagged NK1-R (**Figure 61B**), demonstrating the stability of our bioconjugate for experiments on live cells.
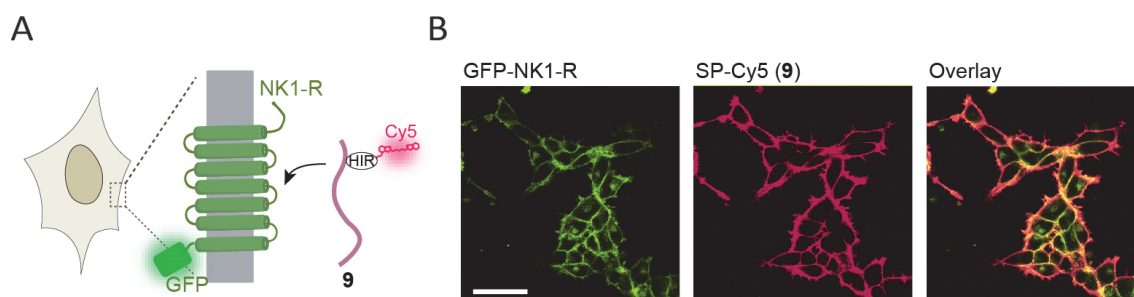
Figure 61: Visualization of membrane receptors in live cells. A) Scheme of the NK1-R-GFP-fusion-expressing cells and the cell-labeling experiment. HIR: hypervalent iodine reagent. **B)** Confocal imaging of cells expressing GFP-NK1-receptors labeled with peptide SP-Cy5 **9**. Scale bar, 50 µm.

Fluorescent dyes often suffer from poor photophysics and photochemistry, resulting in dye bleaching. We envisioned that our dual-modification scheme might enable the attachment of photoprotection compounds which positively affect the photophysical and photochemical properties of nearby fluorophores.
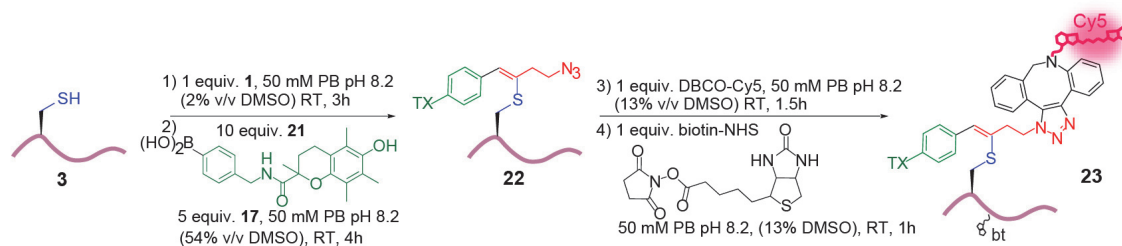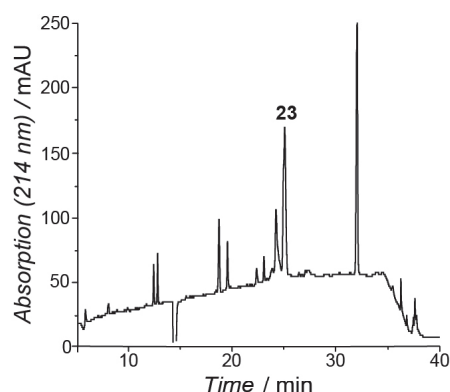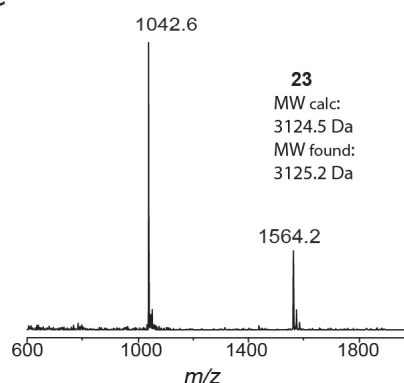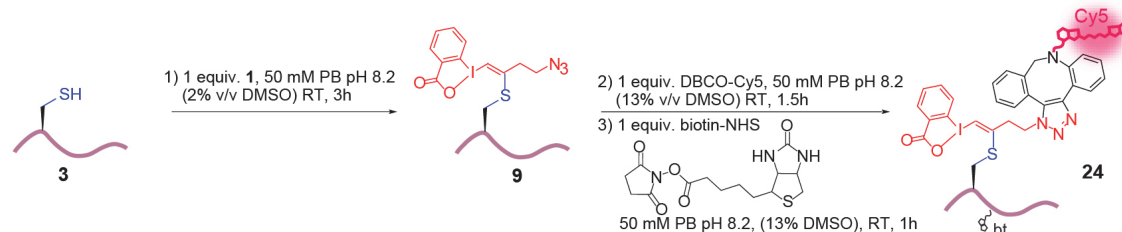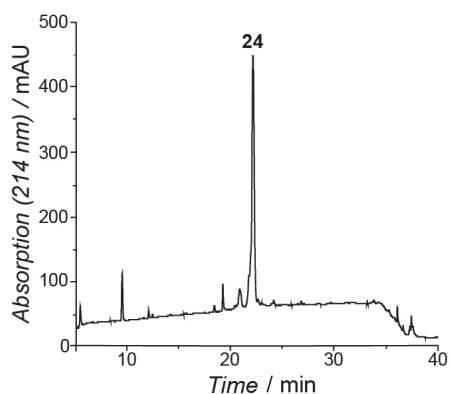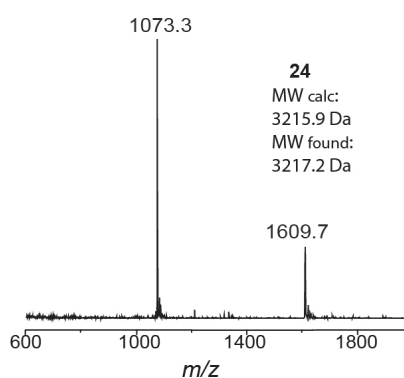
Figure 62: Synthesis of peptides 23 and 24. A) Scheme of the synthetic strategy for doubly modified biotinylated SP-Cy5-TX peptide **23**. **B**) Analytical RP-HPLC chromatogram of the reaction mixture after biotinylation, yielding **23. C**) ESI-MS of product **23** (MW calculated: 3124.5 Da, MW found: 3125.2 Da). **D**) Scheme of the synthetic strategy for the control biotinylated SP-Cy5 peptide **24**. **E**) Analytical RP-HPLC chromatogram of the reaction mixture after biotinylation, yielding **24. F**) ESI-MS of product **24** (MW calculated: 3215.9 Da, MW found: 3217.2 Da).

We thus decided to exploit the "doubly orthogonal" functionalization scheme to increase the photostability of cyanine dye Cy5, which should increase dye brightness, decrease bleaching kinetics and allow for longer tracks in single-particle tracking applications. To this end, we placed the triplet-state quencher (TSQ) 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) in close proximity to the Cy5 fluorophore.[396] Such positioned TSQs reduce blinking rates, photobleaching rates and dark state lifetimes.[397] A boronic-acid adduct of Trolox (**21**) was thus synthesized (by C. J.). Substance P (**3**) was labeled with **1** and coupled to **21** via Suzuki-Miyaura cross-coupling (C. J.). The obtained product **22** was further labeled using DBCO-Cy5 as described above. Finally, I attached a biotin-moiety to Lys5 for subsequent surface-immobilization, yielding the final peptide **23** (**Figure 62A, B, C**). Similarly, I synthesized a control peptide **24** labeled with Cy5 and biotin, but lacking the Trolox moiety (**Figure 62D, E, F**).

I then surface-immobilized **23** (containing a Trolox moiety) and **24** (without Trolox), and observed the stability of individual Cy5 molecules by smTIRFM (**Figure 63**). Individual fluorescent molecules were imaged over time, under conditions of oxygen exclusion. After 50 seconds imaging, most Cy5 dyes were bleached in the absence of coupled Trolox in **24**, whereas for the Trolox-containing molecule **23**, a large portion of Cy5 were still fluorescent (**Figure 63B, C, D**). When measuring bleaching time constants for both conditions (**Figure 63D**) a 3-fold increase in the dye's bleaching time constant was detected due to the proximity to the Trolox moiety (from 17.9 ± 8.8 s for **24** to 55.9 ± 18.6 s for **23**).
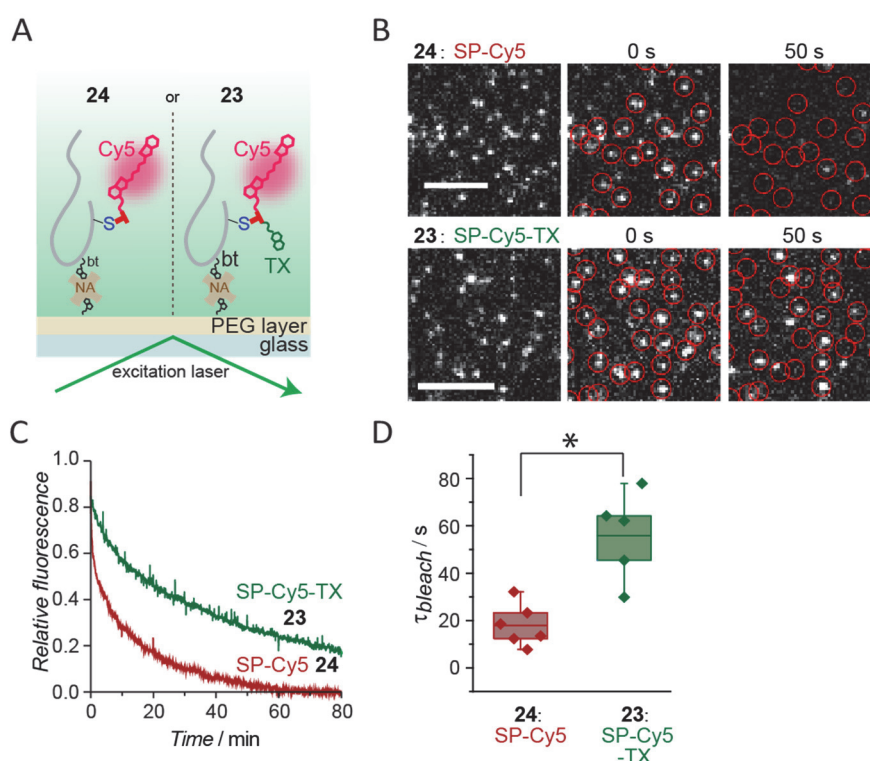
Figure 63: Stabilization of fluorophores by proximal Trolox. A) Scheme of the single-molecule TIRF experiment with immobilized peptide **23** or **24**. bt: biotin. NA: neutravidin. TX: Trolox. **B**) Single-molecule images of immobilized **23** and **24** at indicated time points. Scale bar, 5 μm. **C**) Averaged fluorescence photobleaching kinetics for **23** and **24**. **D**) Fluorescence bleaching time constants from n = 6 experiments for **24** and n = 5 experiments for **23**. *p=0.007.

Together, these results demonstrate that a "doubly orthogonal" functionalization strategy can be exploited in a modular fashion to label proteins with fluorophores stabilized by a covalently coupled TSQ.

4.5   Conclusions

In summary, we have reported a general cysteine labeling protocol for the installation of an unprecedented hypervalent iodine structure on both peptides and proteins. The obtained bioconjugate contains two reactive groups, an azide and a hypervalent iodine, which are orthogonal in reactivity to each other and to existing natural functional groups in peptides and proteins.

The cysteine labeling protocol proceeds with high efficiency, chemoselectivity and functional group tolerance under native conditions. In contrast to previous methods based on the use of hypervalent iodine reagents, the reactive carbon-iodine bond does not react and is transferred intact to the biomolecule. A wide range of peptidic-hypervalent iodine conjugates as well as the first protein-bound hypervalent iodine compound were efficiently generated. Importantly, the obtained conjugates are compatible with protein structure and function as demonstrated by the assembly of nucleosome particles.

The obtained bioconjugates allow "doubly orthogonal" functionalization: an azide group can be selectively functionalized by cycloaddition with cyclooctynes, without affecting the hypervalent bond. Alternatively, the hypervalent iodine structure could be successfully engaged in an aqueous palladium catalyzed cross-coupling with boronic acids, without losing the azido group. This provides means for interesting and highly useful dual functionalization. Here, we demonstrated the usefulness of the approach by improving the photophysics of cyanine dyes by the attachment of a triple-state quencher using the dual-reactive handle. We thus provide a modular method to stabilize organic dyes in biomolecules.

Taken together, our approach allows fast and selective peptide and protein modification. We are convinced that our work has just started to unravel the potential of hypervalent iodine reagents in biomolecule labeling, and the stage is now set to develop a day-to-day use in chemical biology and potential medicinal applications.

4.6   Acknowledgements

Chapter 5: Multivalent Peptide Probes for Chromatin States Investigation in Live Cells

5.1 Outline

Recent studies revealed that the formation of heterochromatin domains involves liquid-liquid phase separation, mediated by multivalent HP1α-chromatin interactions. Motivated by these new discoveries, we wondered whether heterochromatin was the only chromatin state able to phase-separate, or if we can directly detect further histone PTM dependent sub-nuclear domains. We thus designed a modular strategy for the synthesis of multivalent peptide probes capable of detecting chromatin phase-separating states in living cells. Initially designed and optimized for H3K9me3-dependent heterochromatin, we expect that our novel reagents can be easily tuned to the detection of other chromatin states, thus enabling the visualization of different chromatin compartments in single cells and providing a unique dynamic view of the 3-dimensional chromatin state organization.

5.2 Introduction

Combinations of histone post-translational modifications define complex chromatin states.[32] PTMs are bound by effector proteins which organize the chromatin fiber into three dimensional, sub-nuclear compartments. Heterochromatin silencing relies on the binding of HP1α to PTMs of histone tails, such as H3K9me3,[52,53] that in turn triggers chromatin condensation and restricts access for the transcription machinery. Recent studies suggested that the formation of such compartments, e.g. heterochromatin domains, is driven by liquid-liquid phase separation and is mediated by multivalent readers - chromatin interactions.[157,158] The resulting membraneless and distinct heterochromatin foci within the nucleus are characterized by rapid diffusion of proteins inside those domains (see **Chapter 1.2**). It is however not clear if this is a general organizational principle. Moreover, general methods to detect and image LLPS-dependent chromatin organization are lacking. We thus set out to design a modular strategy for the synthesis of multivalent peptide based LLPS probes which are designed to invade LLPS chromatin domains in living cells and enable the detection and dynamic localization of chromatin states *in situ.*

5.2.1 Probe design

Protein-protein interactions or protein-ligand binding are often described as "lock and key" interactions, where the two interacting partners exhibit complementary shapes and geometries that perfectly fit.[398] This simplistic model can be translated to computational molecular docking techniques and used to design lead

compounds and small molecule inhibitors as drug candidates.[399] On the other hand, the same principle can be applied to the development of protein domains that specifically recognize small molecules or protein PTMs, such as antibodies.[400] Recently, in our laboratory, a multivalent protein probe engineered to detect bivalent domains in stem-cells was developed.[97] However, the engineering and development of such protein domains can be tedious and suffers from poor modularity. Indeed, we envisioned a simple and modular strategy to synthesize peptide-based probes that accumulate in the nucleus of live cells at specific chromatin compartments. As mentioned in **Chapter 1.1.2**, different chromatin states are enriched with specific combinations of histone PTMs. Instead of directly targeting the PTM of interest, as a protein domain would do, our probe carries itself multiple copies of the PTM. The accumulation of the probe at the regions of interest is ensured by its binding to endogenous multivalent effector proteins (such as HP1α at heterochromatin) that sequester the probe to sites with high density of this specific endogenous PTM (**Figure 64**). In other words, the PTM present on the probe acts as an "entry ticket" that allows its ingress into a defined phase separated compartment (see **Chapter 1.2.3**, **Figure 15B left**). The sub-nuclear distribution of the probes is then imaged using confocal fluorescence microscopy. In addition, by coupling small molecule probes to our peptide-based delivery agent, such as viscosity or polarity sensors, the properties of the LLPS environment could be revealed.
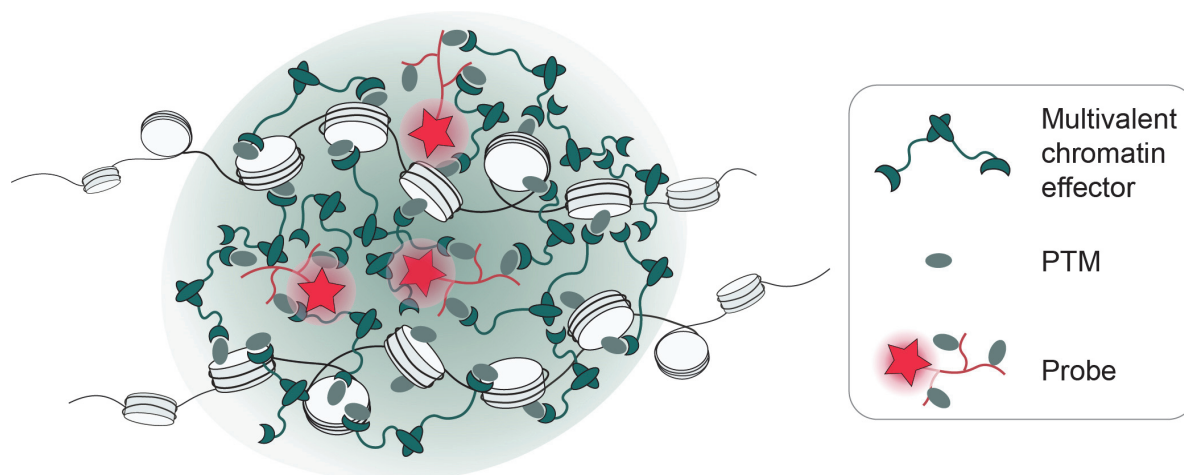
Figure 64: Scheme of the working principle behind the detection of liquid-liquid phase separated chromatin states by our multivalent fluorescent peptide probes. The peptide probe carries multiple times a specific PTM and its binding to endogenous multivalent chromatin effectors causes its accumulation at chromatin foci enriched with the PTM of interest. The dynamic localization of these foci is visualized by confocal microscopy.

Potential advantages that boosted the development of such a technology include: 1) the ease of use, as short incubation instead of transfection allows the probe to reach the cell nucleus, enabling chromatin visualization, 2) the high modularity of the synthetic strategy, that can easily tune the probe specificity to different chromatin states, 3) the ability to stain multiple phase separated chromatin states simultaneously in living cells and follow them over time with a simple reagent, 4) the compatibility with live cell manipulation, as cells fixation would prevent the study of the highly dynamic LLPS chromatin compartments and 6) the ability to target sites with poorly known binding effectors.

Our probes consist of a fluorescently labeled peptide scaffold that allows the efficient coupling of histone fragments, containing combinatorial PTMs characterizing a specific chromatin state. I employed a combination of maleimide-cysteine labeling (to fluorescently label the peptide),[206] disulfide-directed modification (to couple a cell penetrating peptide)[247] and Huisgen azide-alkyne 1,3-dipolar cycloaddition (click chemistry)[401] to rapidly derivatize the peptide scaffold with modified histone peptides.
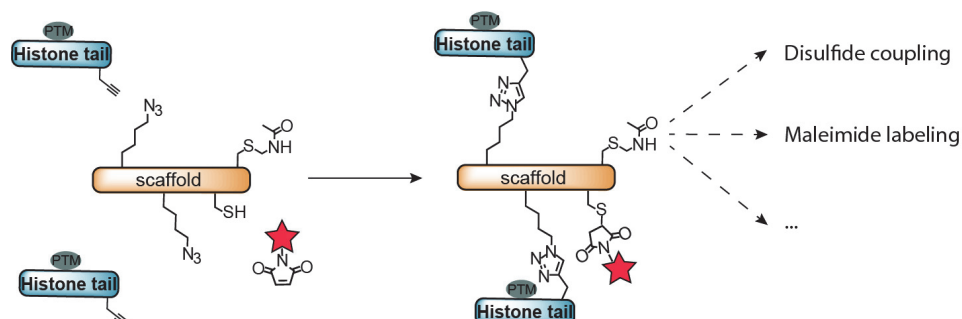
Figure 65: General strategy for the synthesis of the multivalent probes. The modular strategy designed for the synthesis of the multivalent probes relies on the one-pot reaction between a scaffold peptide, the histone-tail peptides carrying the PTM of interest and a maleimide-fluorophore. Orange: scaffold peptide. Blue: modified histone tails. Red star: fluorescent dye.

5.3 Results

As a first case-study, I selected the interaction between HP1α and the H3K9me3 modification, which form well defined phase-separated heterochromatin compartments in mammalian cells.[310] I reasoned that a multivalent probe, functionalized with intrinsically disordered regions such as the modified histone tails would allow to invade specific LLPS chromatin compartments in the nucleus of live cells. For easier understanding, the following table summarizes the characteristics of the different probes synthesized in this work (**Table 1**):

#	Name	Modification	Scaffold feature	Histone length	Multivalency	Fluorophore	Other
1/1*	P1/P1*	H3K9me3/K9A	TAT	H3(1-15)	2	SiR	-
7*	P2*	H3K9A	TAT	H3(1-15)	2	SiR	HA2-TAT
10	P3	H3K9me3	-	H3(1-15)	2	SiR	-
11	P4	H3K9me3	NLS1	H3(1-15)	2	SiR	-
14	P5	H3K9me3	NLS1	H3(1-15)	2	SiR	cR10
16	P6	H3K9me3	NLS1	H3(1-15)	2	SiR	SeL
24/24*	P7/ P7*	H3K9me3/K9A	NLS2	H3(1-15)	3	SiR	-
25/25*	P8/P8*	H3K9me3/K9A	NLS2	H3(1-12)	3	SiR	-
26/26*	P9/P9*	H3K9me3/K9A	NLS2	H3(3-12)	3	SiR	-
27	P10	H3K9me3	neg	H3(1-15)	3	SiR	-
28	P11	H3K9me3	neg	H3(3-12)	3	SiR	-
29	P12	H3K9me3	neg	H3(5-12)	3	SiR	-

Table 1: Multivalent peptide probes and their characteristics. TAT: HIV-TAT protein derived. NLS1: SV40 T antigen derived. NLS2: human c-myc protein derived. Neg: negatively charged. SiR: Silicon-rhodamine. HA2: influenza viral hemagglutinin protein subunit HA2. cR10: cyclic poly-arginine. SeL: diselenolane.

5.3.1 Synthesis of heterochromatin multivalent peptide probes – 1st generation

The probe is composed of multiple and interchangeable peptide building blocks, that were synthesized by Fmoc-SPPS on a rink amide resin. The first-generation probes included a scaffold peptide, two copies of the H3 N-terminal tail and Silicon Rhodamine (SiR) fluorophore (**Figure 65**):

- *Scaffold peptide*

The peptide was functionalized at two sites with an azido residue, introduced during SPPS as azido lysine and separated by GS likers (**Figure 66A**). The azido groups can be used to perform a click reaction and attach two alkyne-containing modified histone tails (see below). A free cysteine close to the C-terminus serves as a reactive residue for the coupling of the fluorescent dye of choice, which in turn allows to image the probe localization in cells. In addition, a cysteine that is temporarily protected with an Acm group is added at the C-terminus. In case a further functionalization of the scaffold peptide is needed, the Acm group can be specifically removed and the thiol reacted with the desired moiety (**Figure 65**). Initially, at the N-terminus, a polyarginine cell penetrating peptide (CPP) derived from the HIV Tat protein[402,403] was added to facilitate the cell membrane penetration (**Figure 66A, B, C**).
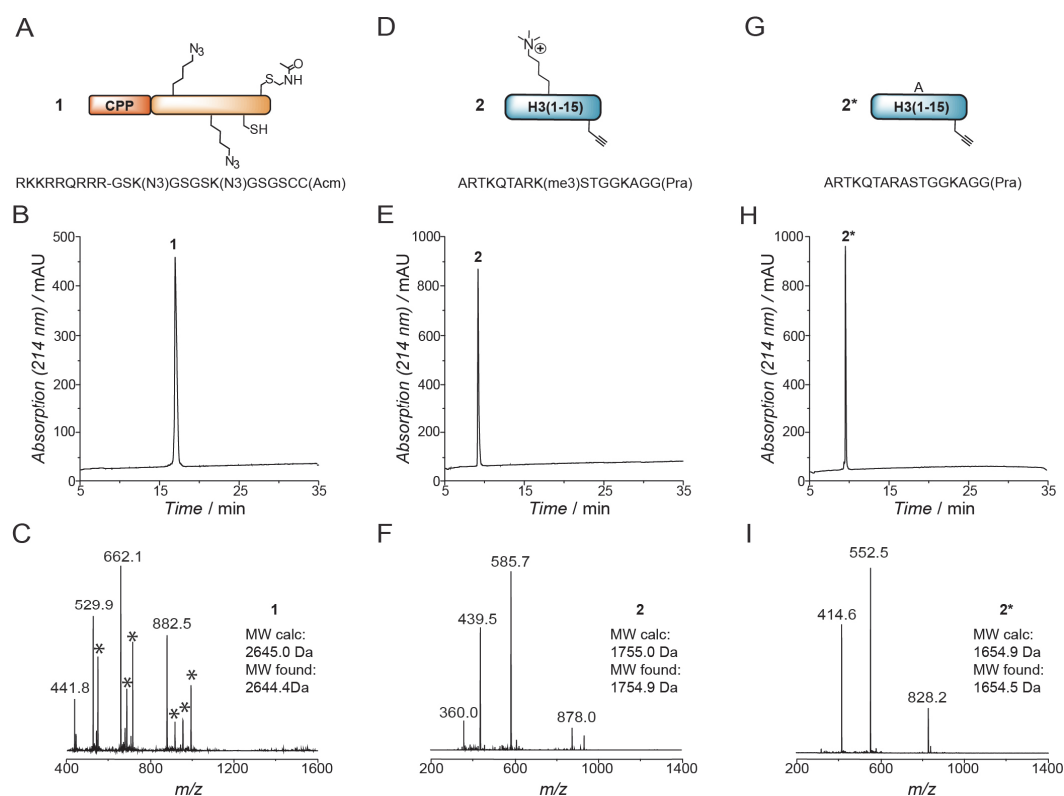
Figure 66: Fmoc-SPPS and analysis of peptide building blocks. A) Sequence of the scaffold peptide **1**. **B**) Analytical RP-HPLC of scaffold peptide **1**. **C**) ESI-MS analysis of scaffold peptide **1** (MW calculated: 2645.0 Da, MW found: 2644.4 Da). **D**) Sequence of H3(1-15)K9me3 propargyl histone peptide **2**. **E**) Analytical RP-HPLC of H3(1-15)K9me3 propargyl histone peptide **2**. **F**) ESI-MS analysis of H3(1-15)K9me3 propargyl histone peptide **2** (MW calculated: 1755.0 Da, MW found: 1754.9 Da). **G**) Sequence of H3(1-15)K9A propargyl histone control peptide **2***. **H**) Analytical RP-HPLC of H3(1-15)K9A propargyl histone control peptide **2***. **I**) ESI-MS analysis of H3(1-15)K9A propargyl histone control peptide **2*** (MW calculated: 1654.9 Da, MW found: 1654.5 Da). Asterisks are TFA adducts.

- *Histone tails*

For the detection of heterochromatin, the H3 N-terminal tail corresponding to residue 1-15 was synthesized and a trimethylated lysine residue was introduced at position K9 during Fmoc-SPPS. The peptide carried an alkyne moiety at the C-terminus to allow the click reaction with the scaffold peptide azido lysines. As a control, the same peptide was also synthesized with a K9A mutation. **Figure 66D-I** shows the sequences and analytics of the K9me3 and K9A histone tail peptides.

- *Fluorophore*

Silicon-rhodamine maleimide (SiR-mal) (**Figure 67A**). The maleimide moiety ensures the reactivity towards the cysteine residue of the scaffold peptide. Silicon-rhodamine was chosen as it is a non-toxic and highly permeable dye with far-red maximal excitation wavelength (652 nm). In addition, aggregation of SiR or interaction with hydrophobic surfaces favors the formation of a spirolactone moiety that turns off its fluorescence emission (OFF state). On the other hand, its fluorescence increases up to 10-100 fold upon binding to the desired target, as the equilibrium is shifted towards the zwitterionic form (ON state) (**Figure 67B**)[404] This fluorogenic behavior allows for an enhanced signal-to-noise ratio, a highly desirable advantage in fluorescence microscopy.
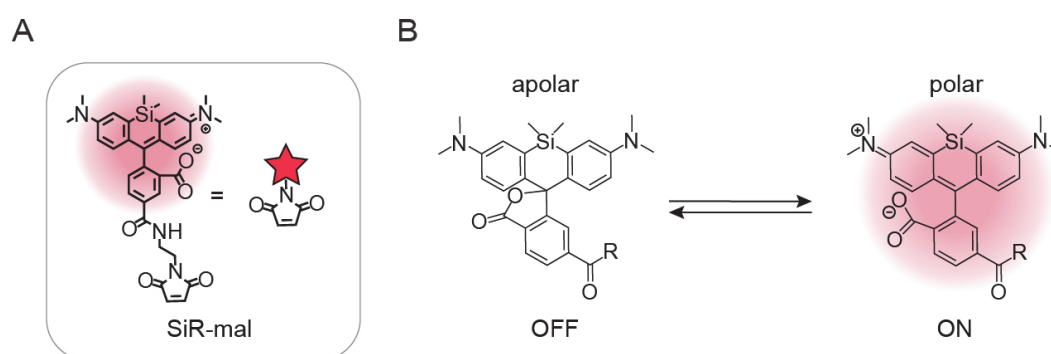
Figure 67: Silicon-rhodamine fluorophore. A) Scheme and structure of Silicon-rhodamine maleimide (SiR-mal). **B**) Silicon-rhodamine equilibrium between ON and OFF states. Two forms of the SiR molecule exist and the equilibrium between states depends on the polarity of the environment. In hydrophobic environments SiR forms a spirolactone that turns off the fluorescence emission. In polar media, the equilibrium is shifted towards the open state with an increase in fluorescence emission up to 10-100 fold.[404]

Having all the building blocks in hand, I set out to find the best conditions for the one-pot cysteine-maleimide labeling and copper-catalyzed azide-alkyne click reaction. First, I labeled the scaffold cysteine with SiR-mal by incubating 1.2 eq. SiR-mal with 1 eq. of scaffold peptide **1** in 100 mM phosphate buffer, pH 7.5 at RT for 15min. Subsequently, I added 4 eq. of histone tail peptide **2** or **2*** in the reaction mixture and initiated the click reaction with 4 eq. CuSO$_4$ and 20 eq. Na ascorbate (**Figure 68A**). The reaction was almost instantaneous and after 5 min the scaffold-SiR peptide intermediate **3** was totally consumed. The obtained probe **P1** (**4**) was purified by semipreparative RP-HPLC and lyophilized (53% average yield). Of note, control probe **P1*** (**4***), carrying the K9A mutation, was synthesized following the same procedure and

obtained with 59% yield. As an example, the synthesis of probe **P1** (**4**) was monitored by analytical RP-HPLC and reported in **Figure 68A-C** (see **Chapter 5.6** for the analytics of control peptide **P1***).

The cellular localization and permeability of the probe **P1** was subsequently tested in live NIH 3T3 mouse fibroblasts and detected by confocal microscopy. This cell line is known to exhibit well defined HP1α-enriched heterochromatin foci (chromocenters) in the nucleus that colocalize with DNA-dense regions.[310] Indeed, specific binding of the probe to endogenous HP1α would create an unambiguous spotty pattern in the cell nucleus (SiR channel) that overlaps with the DNA stain signal (NuclearGreen channel). However, after 1h incubation at a concentration of 1 μM, probe **P1** was exclusively found in discrete cytoplasmic spots, indicating endosomal trapping (**Figure 68D**).[405]
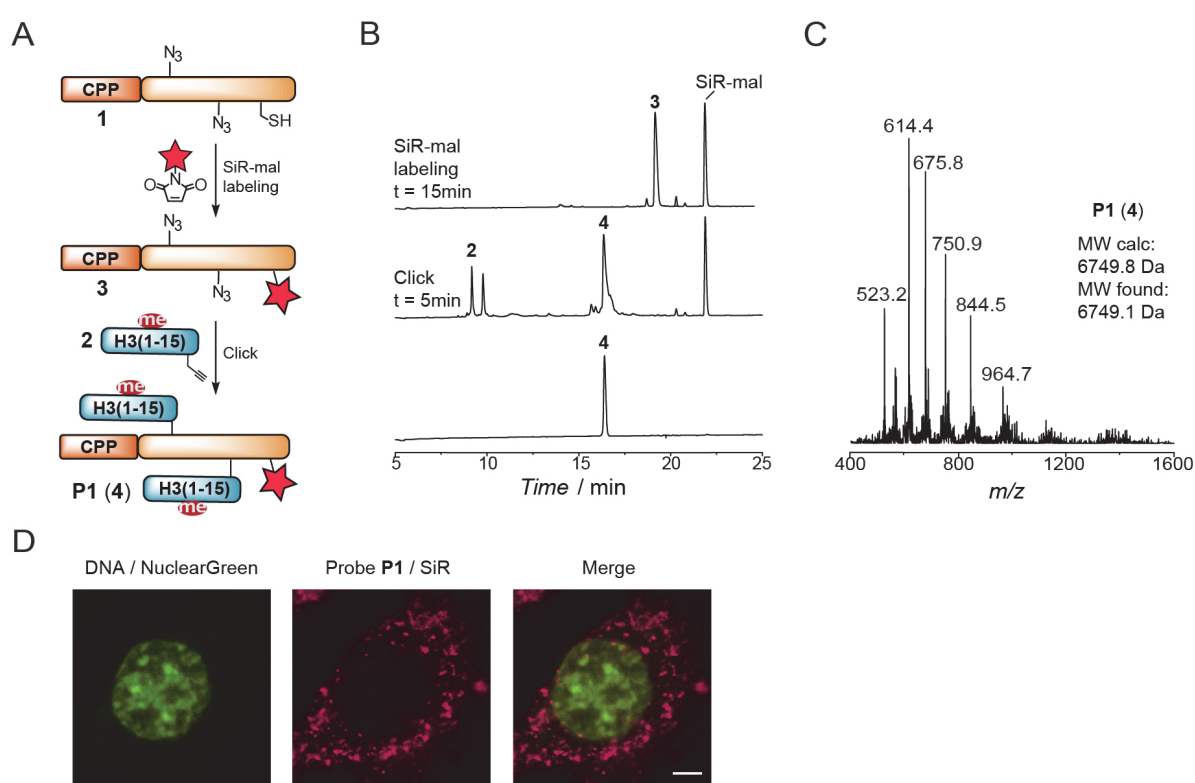
Figure 68: Synthesis of probe P1 and live cell imaging. A) Scheme of the one-pot SiR-mal labeling and copper catalyzed azide-alkyne click reaction to yield probe **P1** (**4**) and control **P1*** (**4***). **B**) Analytical RP-HPLC of the labeling reaction between scaffold peptide **1** and SiR-mal after 15min incubation at RT (top), followed by click reaction with modified histone tail **2** in the presence of CuSO$_4$ and sodium ascorbate (middle) and purified probe **P1** (**4**) (bottom). **C**) ESI-MS analysis of probe **P1** (**4**) (MW calculated: 6749.8 Da, MW found: 6749.1 Da). **D**) Confocal microscopy images of NIH 3T3 live cells incubated with 1 μM probe **P1** (**4**) for 1h at 37 °C. From left to right: DNA stain NuclearGreen; probe **P1** (**4**); merge. Scale bar: 5 μm.

In fact, although the pathways responsible for the cellular penetration of CPPs and the delivery of their cargo remain controversial, endocytosis is recognized to be the primary mode of entry of such positive and arginine-rich CPPs,[406,407,408] and endosomal escape the rate-limiting step for the effectiveness of the delivery.[405]

5.3.2 Overcoming endosomal sequestration

Several strategies have been developed to overcome endosomal sequestration. Among other techniques,[405] the coupling or coincubation of pH-sensitive endosomolytic sequences have shown to considerably improve the cytoplasmic release of the cargo molecules.[409,410] These sequences present multiple acidic residues that, while being deprotonated and negatively charged at neutral pH, become protonated in acidic environments. The protonation of these residues renders the peptide more hydrophobic and prone to be inserted in the lipidic membrane, triggering its rupture. Since endosomes are more acidic than the surrounding intercellular space and cytoplasm,[411] these peptides preferentially lyse endosomes leaving the plasma membrane intact.

Indeed, a pH-sensitive endosomolytic peptide derived from the HA2 subunit of the influenza viral hemagglutinin protein[412] was synthesized (**Figure 69A, B, C**) and coupled to control probe **P1*** through formation of a heterodisulfide bond. In a first step, the Acm protected C-terminal cysteine of probe **P1*** was deprotected by treatment with AgOAc in 50% AcOH over 6h at RT and activated with Ellman's reagent. Subsequently, short incubation of TNB-activated probe **P1*** (**5***) with the N-terminal cysteine-bearing HA2-TAT sequence **6** followed by RP-HPLC purification afforded probe **P2*** (**7***) (**Figure 69D, E, F**).
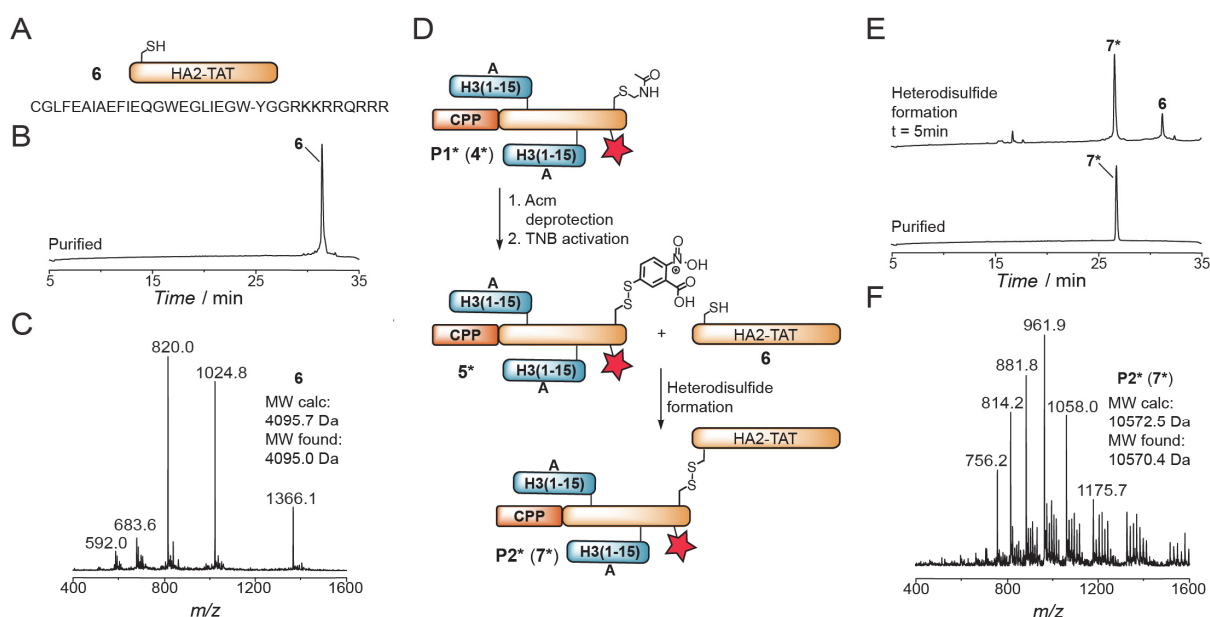
Figure 69: Coupling of HA2-TAT endosomolytic sequence to trigger endosomal escape. A) Sequence of the HA2-TAT peptide **6**. **B**) Analytical RP-HPLC of HA2-TAT peptide **6**. **C**) ESI-MS analysis of the HA2-TAT peptide **6** (MW calculated: 4095.7 Da, MW found: 4095.0 Da). **D**) Scheme of the reaction yielding control probe **P2*** (**7***). Acm deprotection of control probe **P1*** (**4***), followed by TNB activation and RP-HPLC purification yield intermediate **5***. Disulfide-thiol exchange between activated control probe **5*** and endosomolytic peptide **6** and RP-HPLC purification afford probe **P2*** (**7***). **E**) Analytical RP-HPLC of the disulfide-thiol exchange reaction after 5 min incubation at RT (top) and purified control probe **P2*** (**7***) (bottom). **F**) ESI-MS analysis of control probe **P2*** (**7***) (MW calculated: 10572.5 Da, MW found: 10570.4 Da).

Having this new probe in hand, I compared the cytoplasmic release and the nuclear staining of probe **P1*** (no HA2-TAT), probe **P2*** (disulfide-linked HA2-TAT) and coincubation of probe **P1*** and HA2-TAT **6** (not coupled through the disulfide bond) (**Figure 70**). Of note, initial tests of cell penetration were performed with the non-HP1α binding peptide, containing the K9A mutation. Although the endosomal escape ability of most CPPs is strictly dependent on concentration, an increase in control probe **P1*** concentration (5 μM, with respect to 1 μM **P1** in **Figure 68D**) did not help the endosomal escape (**Figure 70A**). On the other hand, the HA2-TAT-coupled probe **P2*** seemed to be toxic and led to cell death at concentrations > 1uM (**Figure 70B**). However, when the control probe **P1*** was coincubated together with the HA2-TAT sequence **6,** the endosomal escape ability was enhanced even at HA2-TAT concentrations as low as 1 μM (**Figure 70C**). Under coincubation conditions, control probe **P1*** was able to partially reach the nucleus and accumulated at defined sites that presumably correspond to nucleoli. Of note, due to the reducing environment found in the cytoplasm, the disulfide bond in probe **P2*** should be cleaved and the probe liberated upon endosomal rupture and cytoplasmic release. Indeed, the cause of such a striking difference in toxicity between disulfide-coupled probe **P2*** (**7***) (**Figure 70B**) and coincubated HA2-TAT **6** and probe **P1*** (**4***) (**Figure 7C**) is still under investigation.
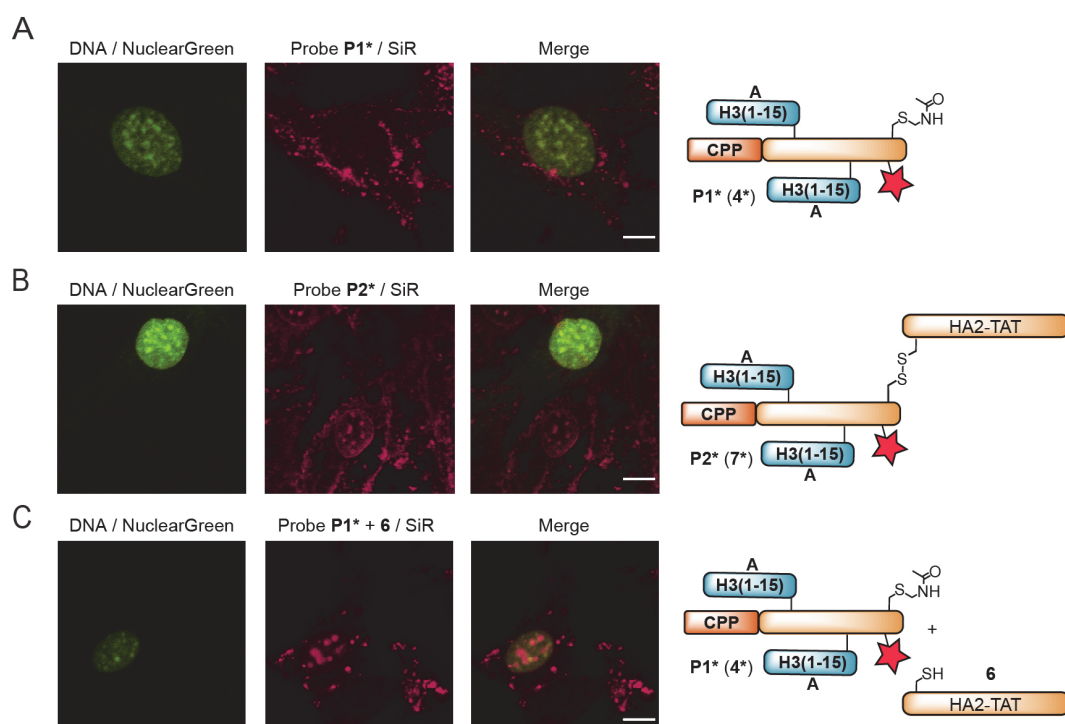
Figure 70: HA2-TAT and endosomal escape. Confocal microscopy images of NIH 3T3 live cells incubated with 5 μM control probe **P1*** (**A**), 5 μM control probe **P2*** (**B**) or 5 μM control probe **P1*** and 5 μM HA2-TAT **6** (**C**) for 1h at 37 °C. From left to right: DNA stain NuclearGreen; SiR probe**;** merge. Scale bar: 10 μm.

5.3.3 Testing heterochromatin foci detection ability of the peptide probe

To test whether probe **P1** was able to accumulate at heterochromatin foci, I compared the nuclear localization of probe **P1** and the control probe **P1*** in NIH 3T3 cells. These cells were also transiently

transfected with a plasmid to overexpress HP1α in fusion with mEos3.2, a monomeric photoswitchable green fluorescent protein.[413] This allows for a straightforward evaluation of the functioning of probe **P1**: if the probe was bound to HP1α proteins, the SiR fluorescence emission would colocalize with the green emission in the nucleus. On the other hand, the negative control peptide **P1***, lacking the K9me3 mark, should show a different localization. However, in both cases, the peptides seemed to be mostly trapped in endosomes, with some probe reaching the nucleus but accumulating in nucleoli (**Figure 71**). No evident difference was observed between the peptide bearing the mark of interest (**Figure 71A**) and the peptide carrying the alanine mutation (**Figure 71B**)
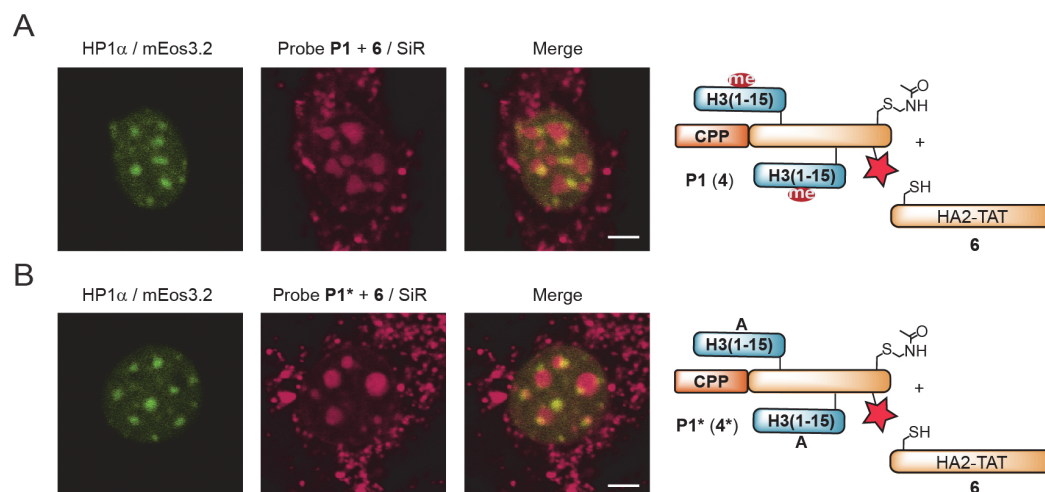
Figure 71: Comparison between nuclear localization of probe P1 and negative control P1*. Confocal microscopy images of NIH 3T3 live cells incubated with 5 μM probe **P1** (**A**) or 5 μM negative control probe **P1*** (**B**) for 1h at 37 °C in the presence of 5 μM HA2-TAT peptide **6.** From left to right: mEos3.2-HP1α; SiR probe**;** merge. Scale bar: 5 μm.

5.3.4 Avoiding nucleolar accumulation

Positively charged and arginine-rich peptide sequences are known to accumulate in nucleoli.[414] This significantly impacted the use of the previously synthesized probes (i.e. **Figure 71**). Thus, I investigated the role of the TAT-derived CPP sequence at the N-terminus of the scaffold peptide **1** in driving nucleolar accumulation. Two new scaffold peptides were synthesized. (**Figure 72A-F**) The CPP sequence was either deleted (scaffold peptide **8**, **Figure 72A, B, C**) or replaced with a nuclear localization sequence (NLS) derived from the SV40 T antigen,[415] that was shown to boost the nuclear localization without accumulation in nucleoli (scaffold peptide **9**, **Figure 72D, E, F**).[414] One-pot SiR-mal labeling followed by click reaction with modified histone tails **2** afforded probes **P3** (**10**) and **P4** (**11**) that were subsequently tested in cells under the same coincubation conditions as in **Figure 71** (see **Chapter 5.6** for the analytics of probes **P3** and **P4**). Indeed, probe **P3**, was now excluded from the nucleoli, but only a very low amount of probe was able to reach the nucleus (**Figure 72G**). On the other hand, probe **P4** improved the nuclear localization and even partially stained HP1α-dense spots (white arrows, **Figure 72H**), although still preferentially staining the nucleoli.
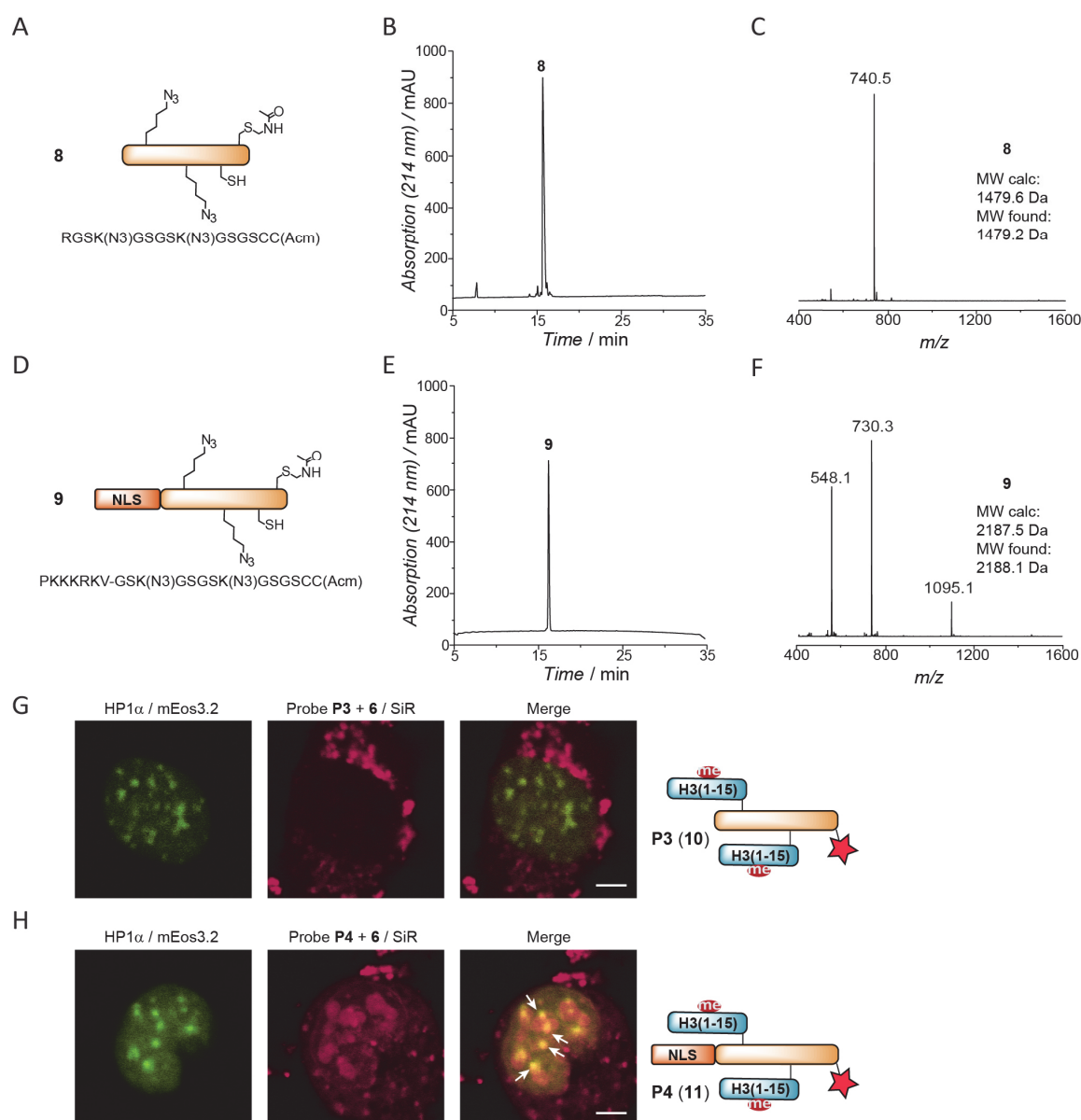
Figure 72: **Synthesis of probes P3 and P4 and live cell imaging. A**) Sequence of scaffold peptide **8**. **B**) Analytical RP-HPLC of scaffold peptide **8**. **C**) ESI-MS analysis of scaffold peptide **8** (MW calculated: 1479.6 Da, MW found: 1479.2 Da). **D**) Sequence of scaffold peptide **9**. **E**) Analytical RP-HPLC of scaffold peptide **9**. **F**) ESI-MS analysis of scaffold peptides **9** (MW calculated: 2187.5 Da, MW found: 2188.1 Da). **G**) Confocal images of NIH 3T3 live cells expressing mEos3.2-HP1α and incubated for 1h at 37 °C with 5 μM of probe **P3** (**10**) (top) or 5 μM of probe **P4** (**11**) (bottom), in the presence of 5 μM HA2-TAT peptide **6**. From left to right: mEos3.2-HP1α; SiR probe; merge. Scale bar: 5 μm.

Although the NLS sequence proved to be beneficial to the nuclear localization and moderately improved the HP1α targeting, two main challenges – cell penetration/endosomal release and selectivity – still needed to be tackled. In fact, following cell penetration, the majority of the peptide probe molecules cannot escape the endosomes, yet the tiny amount that freely diffuses in the cytoplasm and reaches the nucleus does not solely bind to the HP1α-rich heterochromatin phase but is sequestered by the nucleoli.

5.3.5 Optimization of cell penetration and endosomal escape

Two different delivery agents were tested for their ability to efficiently transport and release their cargo in the cytosol.

- Cyclic R10 peptide

Although most CPPs, including the TAT derived sequence used previously, prevalently enter the cells by endocytosis, arginine-rich hydrophilic peptides can cross the plasma membrane by a different mechanism called transduction.[416,417] Such a pathway results in the immediate bioavailability of the probe. In addition, the backbone rigidity is a key requirement for higher transduction efficiencies.[418] Indeed, cyclic CPP exhibit higher cellular uptake kinetics than the corresponding linear peptides.
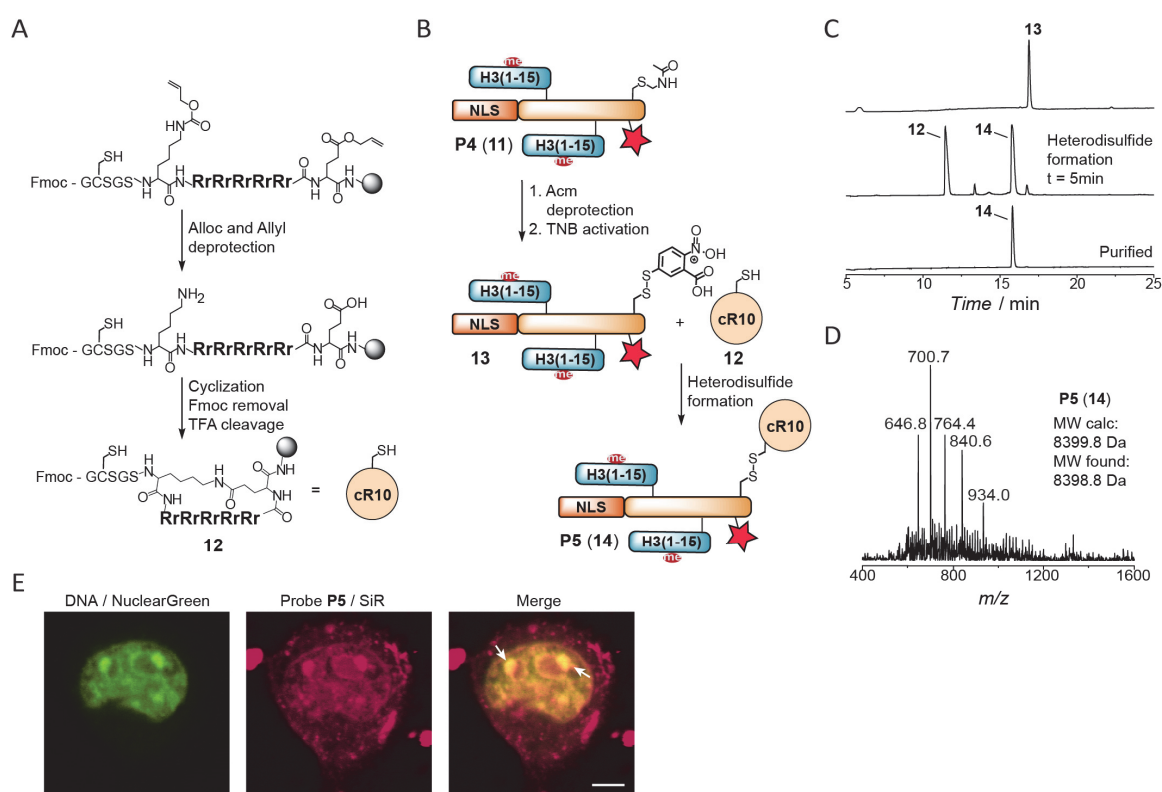
Figure 73: Disulfide-coupled cR10 improves cell penetration and probe bioavailability. A) Synthesis of cR10. cR10 was synthesized by Fmoc-SPPS on rink amide resin. An alloc-protected lysine and an allyl-glutamate were positioned at the N- and C-terminus, respectively. After their concomitant deprotection with Pd(PPh₃)₄ and PhSiH₃ in dichloromethane (DCM), the glutamate carboxylic acid was activated with HATU and DIEA and coupling step with the liberated lysine side chain yielded Fmoc-protected cyclic R10 on resin. The crude peptide was recovered in solution after Fmoc-deprotection and TFA cleavage from the resin, and subsequently purified by RP-HPLC to afford cR10 (**12**). Grey sphere: resin. **B)** Scheme of the reaction yielding probe **P5** (**14**). Acm deprotection of probe **P4** (**11**), followed by TNB activation and RP-HPLC purification yield intermediate **13**. Disulfide-thiol exchange between activated probe **13** and cR10 (**12**) and RP-HPLC purification afford probe **P5** (**14**). **C)** Analytical RP-HPLC chromatograms of the disulfide-thiol exchange reaction. Purified intermediate **13** before the reaction (top); after 5 min incubation at RT with cR10 (**12**) (middle) and purified probe **P5** (**14**) (bottom). **D)** ESI-MS analysis of probe **P5** (**14**) (MW calculated: 8399.8 Da, MW found: 8398.8 Da). **E)** Confocal images of NIH 3T3 live cells incubated for 1h at 37 °C with 5 µM of probe **P5**. From left to right: DNA stain NuclearGreen; SiR probe; merge. Scale bar: 5 µm. White arrows indicate regions where NuclearGreen and SiR signals overlap.

Here, I selected the cyclic R10 (cR10) peptide as a promising candidate for the cell delivery of my peptide probes (**Figure 73A**). The sequence presents ten alternate L- and D-arginine amino acids that confer a flat conformation, and cyclization maximizes the distance between guanidinium groups and favours their interaction with the membrane, leading to enhanced penetration.[418] cR10 (**12**) was synthesized by Fmoc-SPPS and cyclized as reported previously with minor modifications (**Figure 73A,** see **Chapter 5.6** for the analytics of peptide **cR10** (**12**)).[419] Subsequently, it was coupled to probe **P4** (**11**) through disulfide bond formation, to yield probe **P5** (**14**) (**Figure 73B, C, D**). The disulfide bond is readily cleaved in the reducing environment of the cytoplasm, ensuring the release of the probe and avoiding arginine-driven nucleolar accumulation. Incubation of NIH 3T3 live cells with 5 μM probe **P5** and successive confocal imaging revealed more diffused and less punctuate patterns of the SiR signal in the cytoplasm, corroborating the transduction ability of the cR10 carrier. Furthermore, most of the probe was able to stain the nucleus, in particular DNA-dense regions (**Figure 73E**). Thus, the disulfide-linked cR10 seemed to improve the probe cell penetration and nuclear staining ability without causing nucleolar accumulation.

- Diselenolanes

Diselenolanes (SeL) are small molecule delivery agents that rely on a strain-promoted thiol-mediated uptake mechanism.[420] The uptake is initiated by dynamic covalent thiol-disulfide exchange on the surface of the cell. This pathway does not rely on endocytosis, thus overcoming endosomal release problems. Furthermore, the diselenolane moiety was shown not to accumulate in nucleoli.

The SeL-NHS ester molecule was a kind gift by the Matile group (University of Geneva). In order to make it reactive towards cysteine residues, it was derivatized *in situ* with N-(2-aminoethyl)maleimide (mal-NH$_2$). The obtained maleimide-diselenolane (SeL-mal) was then reacted with the Acm-deprotected probe **15** to give probe **P6** (**16**) (**Figure 74A, B, C**). The probe behaviour was then tested in live cells as depicted in **Figure 74D.** First of all**,** in agreement with the literature, a general weaker SiR signal was detected at a probe concentration of 5 μM. In fact, SeL molecules were reported to quench the fluorescence emission of their cargo by 15.6-fold.[420] Interestingly, the spotty pattern typical of endosomal trapping was observed in the majority of cells, meaning that the penetration mechanisms might be influenced by the cargo molecule. Nevertheless, a portion of the probe could reach the nucleus and accumulated at HP1α dense spots (**Figure 74D**, white arrows). Since SeL transporters were shown to be less toxic than positively charged CPPs at high concentration,[420] the incubation with higher amounts of probe **P6** should be tested.

All in all, the utility of cR10 or SeL in transporting and delivering the heterochromatin peptide probes in the cell nucleus was investigated with preliminary experiments and appeared to be potentially promising, especially for cR10. However, further experiments need to be performed to assess to which extent the cell penetration ability is improved.
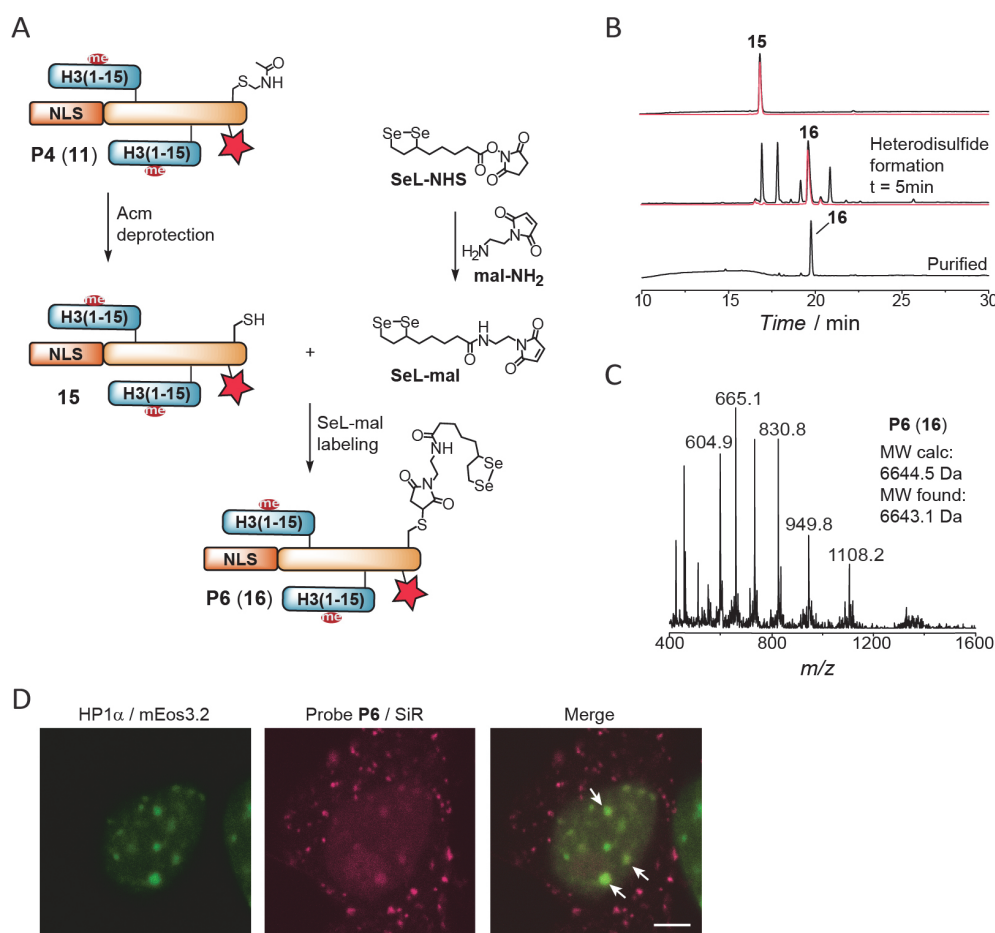
Figure 74: SeL-coupled probe P6 and live cell imaging. A) Synthesis of probe **P6** (**16**) . 2 eq. SeL-NHS were reacted with 1 eq. N-(2-aminoethyl)maleimide and the excess quenched with glycine. Concomitantly, the Acm group on probe **11** was removed and peptide **15** purified by RP-HPLC. Cysteine labeling of peptide **15** by SeL-mal and RP-HPLC purification afforded probe **P6** (**16**). **B**) Analytical RP-HPLC of the reaction yielding probe **P6** (**16**). Purified intermediate **15** before the labeling reaction (top); after 15 min incubation at RT with SeL-mal (middle) and purified probe **P6** (**16**) (bottom). Black: 214nm, red: 652 nm absorption. **C**) ESI-MS analysis of probe **P6** (**16**) (MW calculated: 6644.5 Da, MW found: 6643.1 Da). **D**) Confocal images of NIH 3T3 live cells incubated for 1h at 37 °C with 5 µM of probe **P6**. From left to right: mEos3.2-HP1α; SiR probe**;** merge. Scale bar: 5 µm. White arrows indicate regions where mEos3.2 and SiR signals overlap.

5.3.6 Optimization of probe selectivity

The second major challenge in the development of a probe for chromatin states detection in live cells is its selectivity. Currently, the specific interaction between H3K9me3 containing probe and HP1α-dense sites is hindered by the nucleolar accumulation, which results in a lower actual probe concentration in the nucleus. Furthermore, the unspecific binding of the probe to the surrounding environment, such as the negatively charged DNA, might explain why the probes and control peptides show comparable staining patterns in the cell nucleus (i.e **Figure 71**).

Indeed, electrophoresis mobility shift assay (EMSA) with 170bp 601 DNA reveled a strong binding affinity (in the order of 200-400 nM) between probe **P4** and DNA (**Figure 75A, B (top)**). The interaction was only

slightly weaker when probe **P3**, lacking the NLS sequence, was used instead (**Figure 75A, B (bottom)**). This indicates that the positively charged residues present in the NLS sequence are not the main contributors to the interaction, but rather the scaffold sequence or, most likely, the targeting histone tails.
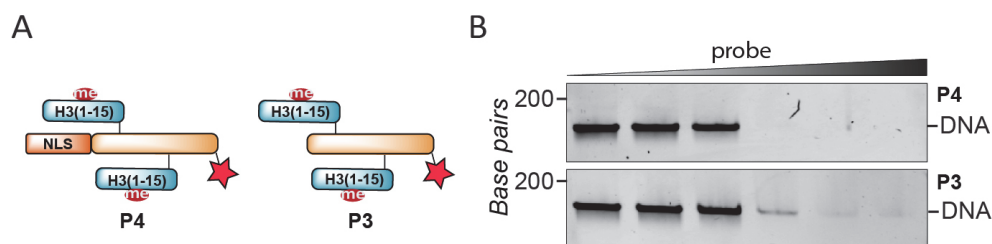
Figure 75: Probes P4 and P3 bind DNA. A) Scheme of the probes used in the EMSA experiment. **B**) EMSA with 20 nM 170bp 601 DNA and increasing concentration of probe **P4** (top) or probe **P3** (bottom), after 30 min incubation at RT. Tested probe concentrations (from left to right): 50, 100, 200, 400, 800, 1600 nM. Note: the shifted DNA-probe complex did not enter the gel and was detected at the wells bottom in the SiR-fluorescence channel.

To face these two main challenges, that is, nucleolar trapping and DNA unspecific staining, two new scaffold peptides were synthesized (**Figure 76**).
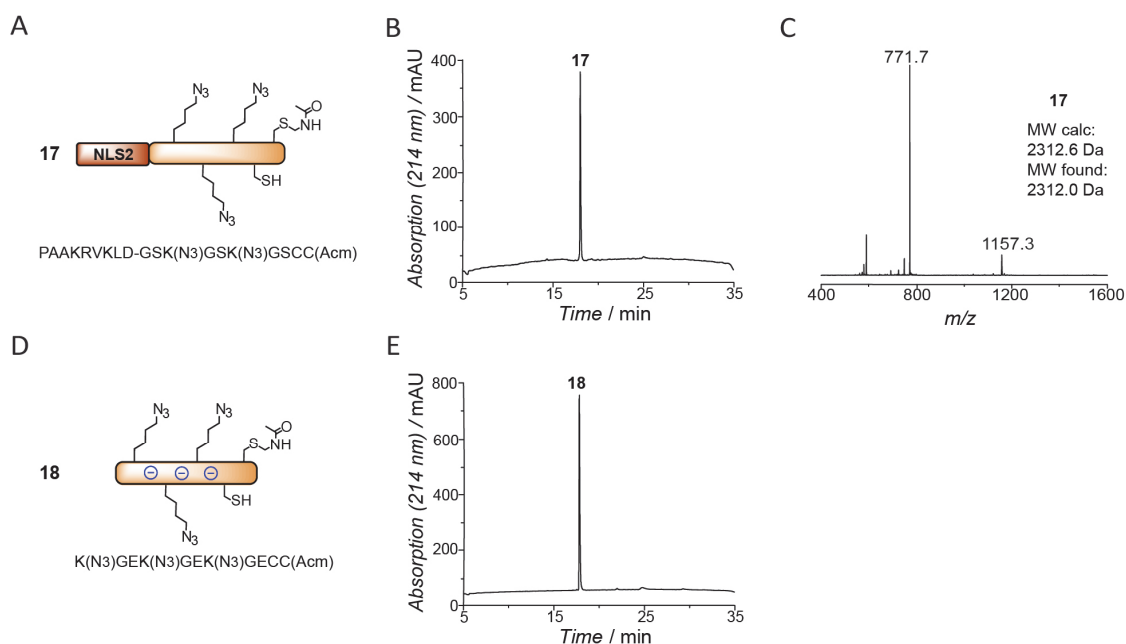
Figure 76: Synthesis of scaffold peptide 17 and scaffold peptide 18. A) Sequence of scaffold peptide **17** synthesized by Fmoc-SPPS. **B**) Analytical RP-HPLC chromatogram of scaffold peptide **17**. **C**) ESI-MS analysis of scaffold peptide **17** (MW calculated: 2312.6 Da, MW found: 2312.0 Da). **D**) Sequence of scaffold peptide **18**. **E**) Analytical RP-HPLC chromatogram of scaffold peptide **18**. The identity of peptide S5 could not be confirmed by ESI-MS because it is poorly ionizable (MW calculated: 1315.4 Da).

In order to make the specific interaction stronger, the multivalency of the probes was increased to three. To this aim, three azidolysines (instead of two), were included in the Fmoc-SPPS of the peptides, allowing subsequent click of three modified alkyne-containing histone tails. To prevent the nucleolar staining, an alternative NLS sequence (NLS2), derived from the human c-myc protein,[421] was added at the N-terminus of the scaffold peptide **17** (**Figure 76A**). On the other hand, the second scaffold peptide **18** was designed

so as to carry three negatively charged glutamate residues to oppose the affinity of the probe for DNA (**Figure 76B**).

In addition, shorter versions of the alkyne-histone tails were synthesized (**Figure 77**). The histone tails are highly basic peptide sequences with multiple positively charged amino acids. As a consequence, I reasoned that shortening the recognition sequence surrounding the key K9me3 mark would reduce the probe unspecific binding while preserving the specific recognition by HP1α. Previous structural studies have shown that the minimal HP1α binding motif on the trimethylated H3 tail are residues 5-11, and the QTAR peptide sequence preceding K9me3 is the major contributor to the additional interactions with the HP1α chromodomain.[422] Indeed, the histone tail was progressively chopped to give: H3(1-12)K9me3 (**19**, **Figure 77A**), H3(3-12)K9me3 (**20**, **Figure 77B**) and H3(5-12)K9me3 (**21**, **Figure 77C**) alkyne peptides. Similarly, short histone tails carrying the K9A mutation were also synthesized (see **Chapter 5.6** for the analytics of peptides **19*** and **20***).
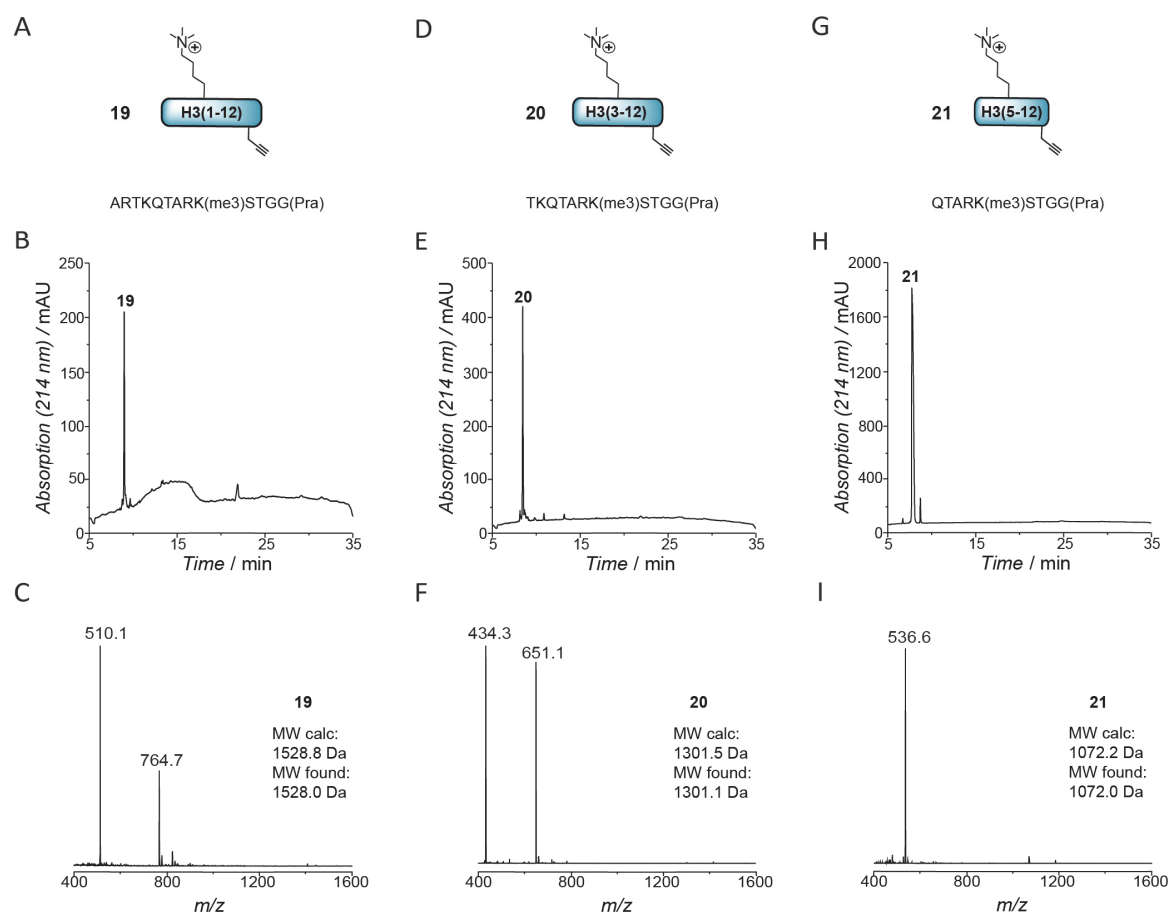
Figure 77: Synthesis of shorter H3K9me3 tails. A) Sequence of H3(1-12)K9me3-alkyne peptide **19** synthesized by Fmoc-SPPS. **B**) Analytical RP-HPLC of H3(1-12)K9me3-alkyne peptide **19**. **C**) ESI-MS analysis of H3(1-12)K9me3-alkyne peptide **19** (MW calculated: 1528.8 Da, MW found: 1528.0 Da). **D**) Sequence of H3(3-12)K9me3-alkyne peptide **20** synthesized by Fmoc-SPPS. **E**) Analytical RP-HPLC of H3(3-12)K9me3-alkyne peptide **20**. **F**) ESI-MS analysis of H3(3-12)K9me3-alkyne peptide **20** (MW calculated: 1301.5 Da, MW found: 1301.1 Da). **G**) Sequence of H3(5-12)K9me3-alkyne peptide **21** synthesized by Fmoc-SPPS. **H**) Analytical RP-HPLC of H3(5-12)K9me3-alkyne peptide **21**. **I**) ESI-MS analysis of H3(5-12)K9me3-alkyne peptide **21** (MW calculated: 1072.2 Da, MW found 1072.0 Da).

Due to the high modularity of the synthetic strategy, the newly synthesized scaffolds and histone tails could be combinatorially reacted to assemble probes **P7 – P12** and some of the relative negative controls (**Table 1**), by one-pot SiR-mal labeling and click reaction. As an example, analytical RP-HPLC traces and ESI-MS spectra are reported for probes **P8** (**25**, **Figure 78A, B, C**) and **P12** (**29, Figure 78D, E, F**) (**see Chapter 5.6** for the analytics of probes **P7**, **P9**, **P10**, **P11** and control peptides **P7***, **P8*** and **P9***).
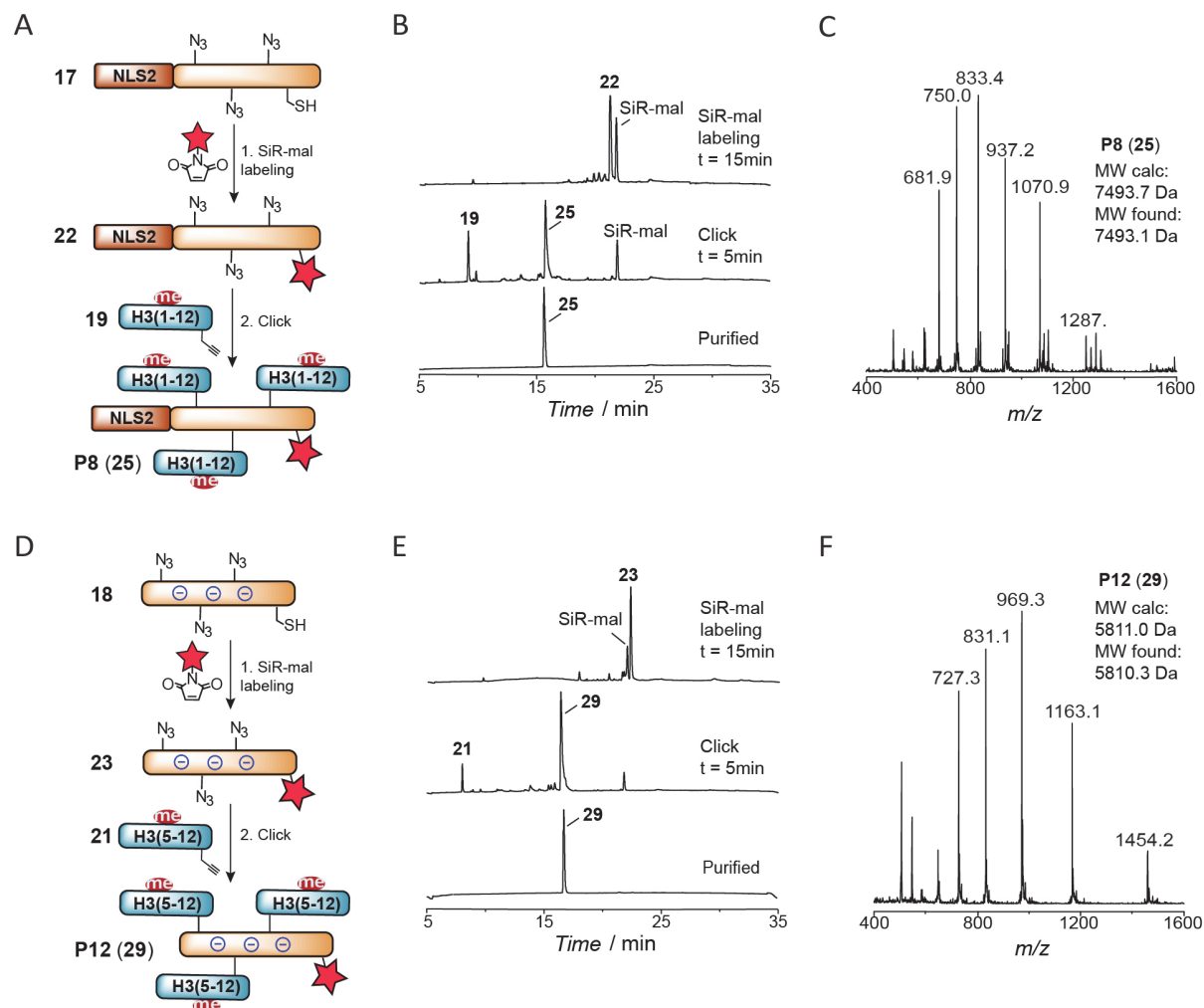
Figure 78: Synthesis of optimized probes. A) Scheme of the one-pot SiR-mal labeling and copper catalyzed azide-alkyne click reaction to yield probe variant **P8** (**25**). **B**) Analytical RP-HPLC of the labeling reaction between the scaffold peptide **17** and SiR-mal after 15min incubation at RT (top), followed by click reaction with modified short histone tail **19** in the presence of CuSO₄ and sodium ascorbate (middle) and purified probe **P8** (**25**). **C**) ESI-MS analysis of probe **P8** (**25**) (MW calculated: 7493.7 Da, MW found: 7493.1 Da). **D**) Scheme of the one-pot SiR-mal labeling and copper catalyzed azide-alkyne click reaction to yield probe variant **P12** (**29**). **E**) Analytical RP-HPLC of the labeling reaction between the scaffold peptide **18** and SiR-mal after 15 min incubation at RT (top), followed by click reaction with modified short histone tail **21** in the presence of CuSO₄ and sodium ascorbate (middle) and purified probe **P12** (**29**). **C**) ESI-MS analysis of probe **P12** (**29**) (MW calculated: 5811.0 Da, MW found: 5810.3 Da).

The newly synthesized probes were first checked for their unspecific DNA binding ability by EMSA (**Figure 79**). The negative charges on scaffold peptide **18** could not counterbalance the strong affinity of the positively charged long histone tails (H3(1-15)) for DNA, confirming the pivotal role of the histone tails in driving the DNA interaction (probe **P10**, **Figure 79A, B (top)**). Indeed, probes with shorter histone

sequences, such as probe **P11** or probe **P12** (**Figure 79A, B (middle-top and middle)**) did not show any DNA binding ability in the probe concentration range tested (150-1575 nM). Regarding the probes with scaffold peptide **17** (with NLS2), shorter H3 tails led to weaker DNA interactions, even though the DNA band was partially shifted at the highest probe concentrations studied (probe **P8** and probe **P9**, **Figure 79A, B (middle-bottom and bottom)**). This indicates that the sequence of the scaffold peptide does influence the DNA binding affinity, albeit playing a minor role with respect to the histone tail length.
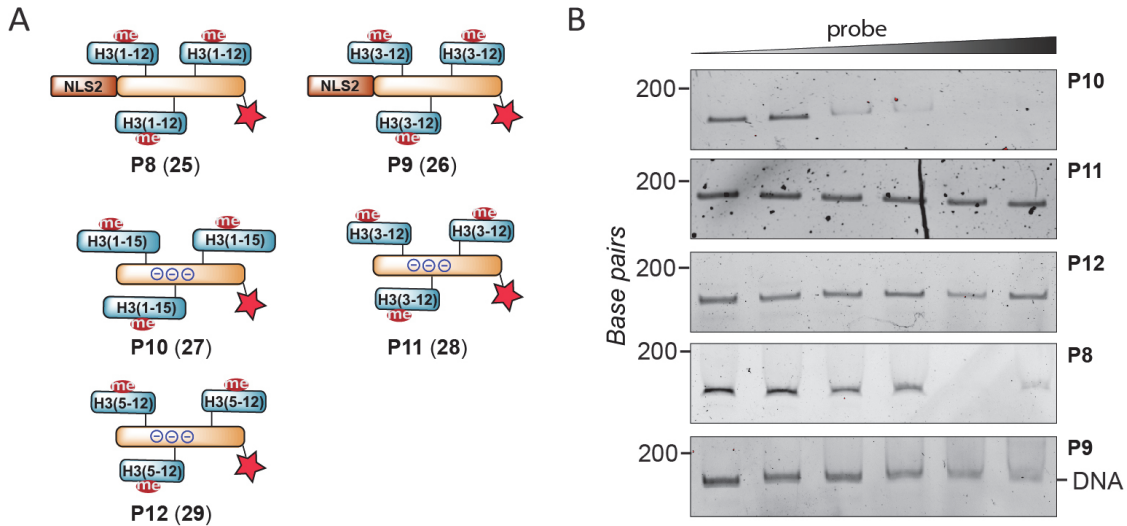
Figure 79: Histone tails are the main contributors to DNA binding. A) Scheme of the newly synthesized probes tested for their DNA binding ability. **B**) EMSA with 10 nM 170bp 601 DNA and increasing concentration of probe **P10** (top), probe **P11** (middle-top), probe **P12** (middle), probe **P8** (middle-bottom) or probe **P9**, after 30 min incubation at RT. Tested probe concentrations (from left to right): 150, 240, 385, 615, 985, 1575 nM. Note: the shifted DNA-probe complex did not enter the gel and was detected at the wells bottom in the SiR-fluorescence channel.

5.3.7 Testing HP1α binding ability of peptide probes *in vitro*

Since the EMSA analysis revealed that DNA unspecific binding could be prevented by shortening the histone tails, it was of fundamental importance to ensure that such short motifs could still efficiently bind to HP1α. I thus set up an *in vitro* liquid-liquid phase separation assay, in which the interaction between the multivalent peptide probes and phosphorylated HP1α (phHP1α) causes the formation of phase-separated fluorescent droplets. The size and number of such droplets represent a straightforward read-out that correlates with the strength of the interaction. In fact, phHP1α alone was found to phase-separate in solutions as concentrated as ~200 μM.[157] At lower concentrations, addition of a multivalent binder, such as our probe, would induce a decrease in the saturation concentration with consequent droplets formation.

Indeed, the phase separation induction ability of the probe was exploited as a readout of HP1α binding. We expect that the peptides only sense, but not induce, phase separation for the following two reasons: 1) the peptide that reaches the nucleus has a lower concentration than the one tested *in vitro*, 2) in the nucleus, HP1α phase separates with chromatin prior to addition of the peptide probe, thus the saturation

concentration for HP1α has already been reached. The peptide probes are therefore not expected to generate artificial phase separated HP1α compartments in live cells.

First of all, as a proof-of-concept, the binding ability of probe **P7**, carrying the NLS2 sequence and long histone tails, was compared with its respective negative control peptide **P7***, carrying the K9A mutation (**Figure 80**). Immediately after mixing phHP1α and probe **P7**, at a respective final concentration of 60 and 10 μM at 4 °C, the clear solution became turbid and viscous. On the other hand, addition of the K9A corresponding control peptide P7* did not cause any perceptible changes (**Figure 80A**). The solutions were then inspected by confocal microscopy (**Figure 80B**). Big and copious SiR fluorescent phHP1α-probe droplets (mean number: 37.67 ± 5.86; mean diameter: 2.91 ± 0.15 μm) were detected with 10 μM probe **P7** (**Figure 80B left**). As expected, considerably fewer and smaller phase separated compartments were imaged for the control peptide **P7*** at the same concentration (mean number: 23.33 ± 4.5; mean diameter: 0.87 ± 0.13 μm) and the size of the droplets was comparable to the solution containing a 10-fold diluted probe **P7** (1 μM, mean number: 10.33 ± 3.51; mean diameter: 0.75 ± 0.02 μm) (**Figure 80B, middle and right** and **Figure 80C**). This indicates that the NLS2 scaffold combined with the long modified histone tails is able to bind to HP1α and that specific mark recognition is driving the demixing of the solution.
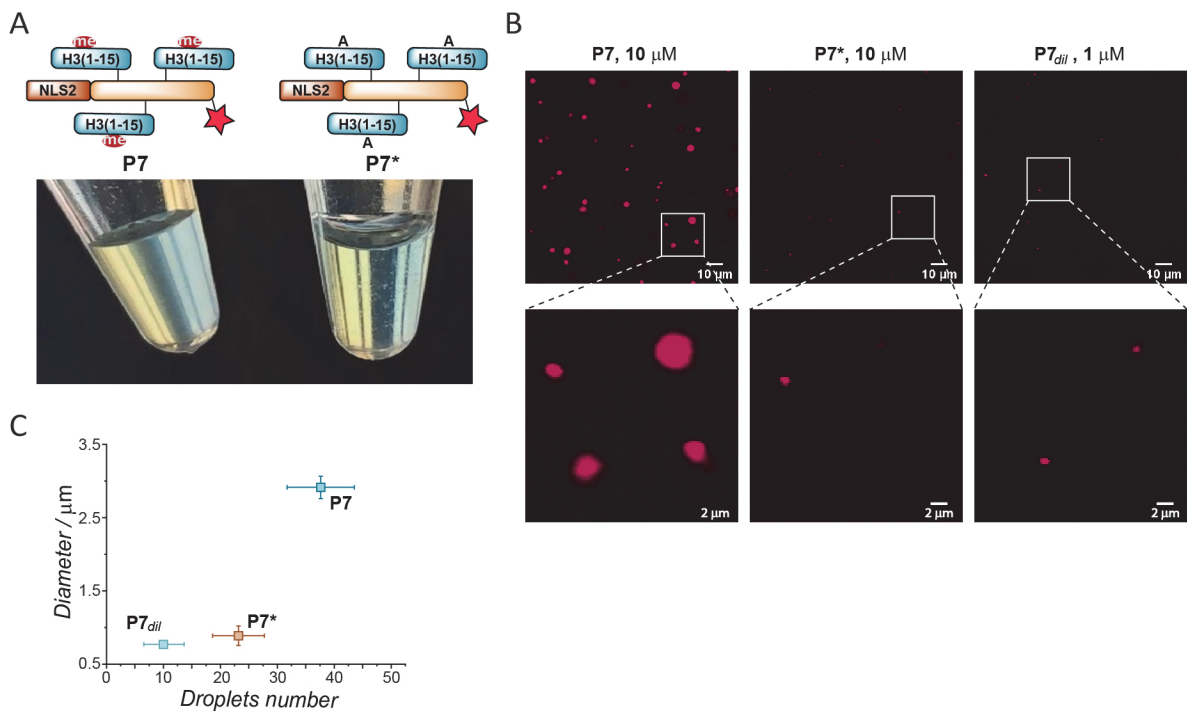
Figure 80: *In vitro* phase separation assay with phHP1α and probe P7 or negative control peptide P7*. A) Image of the turbid phase separated phHP1α-probe **P7** solution obtained after mixing 60 μM phHP1α and 10 μM probe **P7** at 4 °C in PBS (left). 60 μM phHP1α were also mixed with 10 μM control peptide **P7***, under the same conditions, to obtain a clear solution (right). **B**) Confocal images of the SiR-fluorescent droplets obtained by mixing 60 μM phHP1α and 10 μM probe **P7** (left), 10 μM control peptide **P7*** (middle) or 1 μM probe **P7** (**P7_dil**) (right) and incubating the mixture for 5 min at 4 °C. **C**) Calculation of droplets mean diameter and number. Error bars represent standard deviation from 3 movies, obtained from a single experiment.

Next, I investigated if the short histone tails could still bind to their target, and whether the scaffold peptide played a role in the phase-separation process. To this end, 70 µM phHP1α were incubated for 5 min at 4 °C with 10 µM of probe **P7**, **P8** or probe **P9** and the resulting mixture imaged by confocal microscopy (**Figure 81A**).
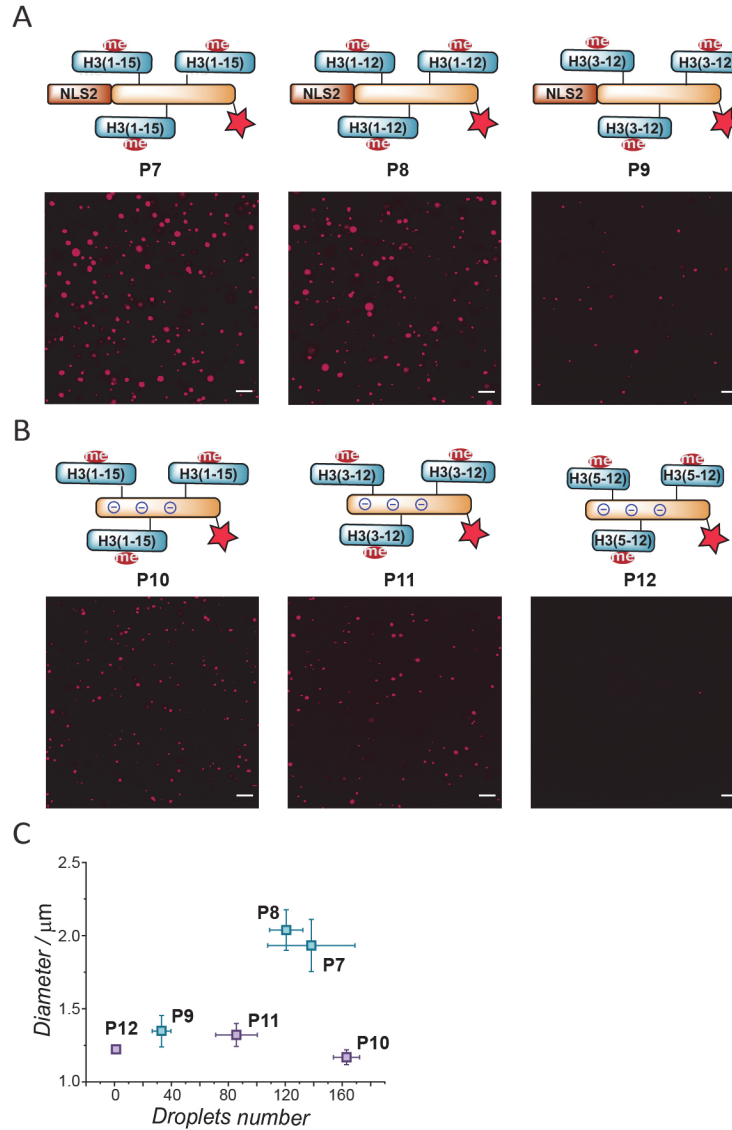
Figure 81: *In vitro* **phase separation assay with phHP1α and probes P7–P12. A**) Confocal images of the SiR-fluorescent droplets obtained by mixing 70 µM phHP1α and 10 µM probe **P7** (left), 10 µM probe **P8** (middle) or 10 µM probe **P9** (right) and incubating the mixture for 5 min at 4 °C. **B**) Confocal images of the SiR-fluorescent droplets obtained by mixing 70 µM phHP1α and 10 µM probe **P10** (left), 10 µM probe **P11** (middle) or 10 µM probe **P12** (right) and incubating the mixture for 5 min at 4 °C. Scale bar: 10 µM. **C**) Plot of droplets mean diameter and number. Error bars represent standard deviation from 3 movies, obtained from a single experiment. For image analysis method, see **Chapter 6.2.18**.

Probes with long and intermediate H3 tail lengths (**P7** and **P8**) appeared to induce phase separation to a similar extent (**P7**, mean number: 120.67 ± 11.68; mean diameter: 2.04 ± 0.14 µm; **P8**, mean number: 138.33 ± 30.66; mean diameter: 1.93 ± 0.18 µm), suggesting that the H3 residues above position 12 are not involved in HP1α recognition (**Figure 81A, left** and **middle panels** and **Figure 81C**). Further chopping

the histone tail at the N-terminus (**P9**) resulted in the formation of smaller and fewer droplets (mean number: 33.00 ± 6.56; mean diameter: 1.35 ± 0.11 µm) (**Figure 81A, right** and **Figure 81C**), indicating a weaker interaction between phHP1α and this probe. Nevertheless, the presence of droplets corroborated that even this short binding motif (H3(3-12)K9me3) is sufficiently long to be recognized by HP1α. Probes bearing the negatively charged scaffold peptide **P10**, **P11** and **P12** were then considered. Long histone tail-probe **P10** triggered phase separation (mean number: 163.00 ± 9.16; mean diameter: 1.17 ± 0.05 µm), but to a weaker extent than **P7** in terms of droplet size (**Figure 81B, left** and **Figure 81C**). In addition, probe **P11** (with histone sequence H3(3-12)) formed droplets with diameters comparable to **P9** (same histone length but different scaffold peptide) and **P10** (same scaffold peptide but longer histone tail) **(Figure 81B, middle** and **Figure 81C**) although the number of the droplets (mean number: 85.67 ± 14.57; mean diameter: 1.32 ± 0.08 µm) was lower relative to probe **P10**. Interestingly, the probe with the shortest histone sequence, comprising residues 5-12, did not show any droplets formation, indicating that the H3(5-12)K9me3 sequence is not a good binder of HP1α (**Figure 81B, right** and **Figure 81C**).

This *in vitro* phase separation assay thus provides a visual and straightforward indication of how strongly the peptide probe binds to its target. However, in order to quantify this binding affinity, I will employ other methods, such as microscale thermophoresis (MST),[423] to quantitatively measure the dissociation constant of the interaction.

5.3.8 Testing optimized probes in living cells

An ideal heterochromatin probe should have a strong binding affinity for its target (HP1α), while showing minimal unspecific interactions (DNA). Indeed, probes exhibiting *in vitro* phase separation propensity, as well as no or low DNA band shift ability were selected as promising candidates and tested in live cells (probes **P8**, **P9** and **P11**) together with their respective control peptides (**P8***, **P9*** and **P11***) under coincubation conditions with HA2-TAT **6**.

Probe **P11**, possibly due to its multiple negative charges, could not cross the cell membrane at the conditions tested (data not shown). On the other hand, both probes **P8** and **P9** efficiently penetrated the cells and showed colocalization with the HP1α chromocenters (Pearson's correlation coefficient (PCC) for probe **P8** = 0.632 ± 0.10; PCC for **P9** = 0.56 ± 0.10) (**Figure 82A, C**). Furthermore, to our delight, the control peptides stained HP1α dense regions significantly less effectively (PCC for **P8*** = 0.30 ± 0.11; PCC for **P9*** = 0.072 ± 0.17). **P8*** preferentially stained nucleoli over heterochromatin, while **P9*** did not show strong accumulation into the nucleus (**Figure 82B, D**). Therefore, probes **P8** and **P9** are specific and cell permeable heterochromatin multivalent peptide probes, and their development sets the stage for the investigation of LLPS heterochromatin domains in live cells.
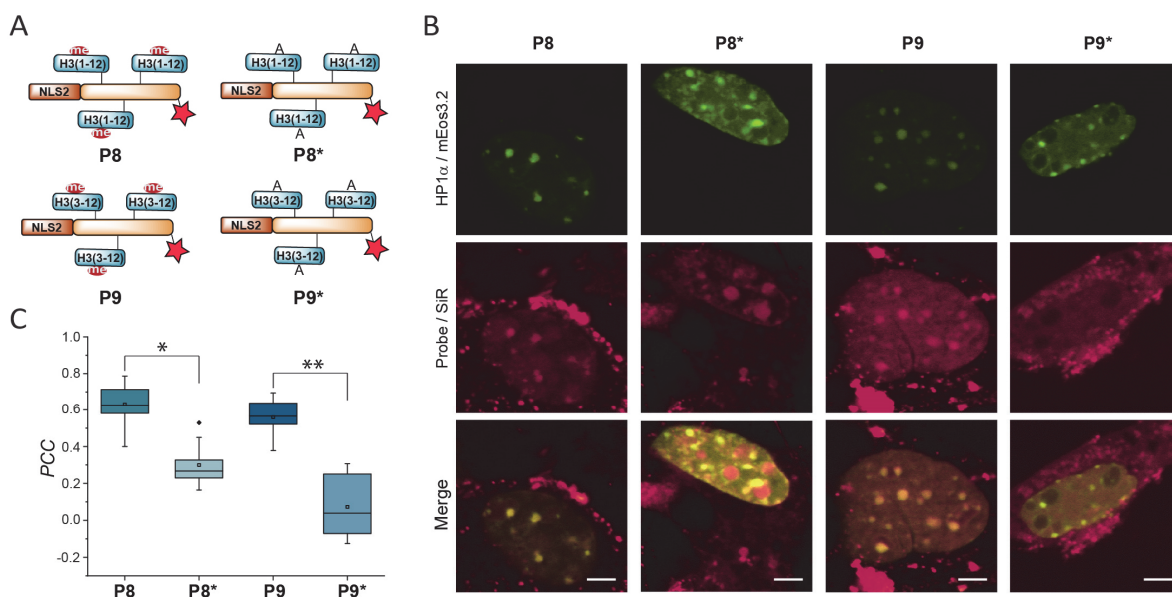
Figure 82: Probes P8, P9 and negative control peptides P8*, P9* in live cells. **A**) Schematic representation of the probes tested. **B**) Confocal images of NIH 3T3 live cells expressing mEos3.2-HP1α and coincubated for 1h at 37 °C with 5 μM endosomolytic sequence **6** and 5 μM of probe **P8** (first column), 5 μM control peptide **P8*** (second column), 5 μM of probe **P9** (third column) or 5 μM of control peptide **P9*** (fourth column). From top to bottom: mEos3.2-HP1α; SiR probe**;** merge. Scale bar: 5 μm. **C**) Colocalization of HP1α and peptide probe quantified by measuring Pearson's correlation coefficients (P8, n=21; P8*, n=10; P9, n=9; P9*, n=6. n represents the number of cells considered. Two-sample t-test, *****p<4.10^{-7}, ******p<0.0005). For image analysis method, see **Chapter 6.2.16**.

5.4 Conclusions

A modular strategy for the synthesis of peptide-based heterochromatin probes was developed. Two specific probes for heterochromatin detection were produced after multiple rounds of optimization. Indeed, the two main obstacles encountered were probe specificity and cell penetration. Both were investigated and preliminary results showed that the lack of specificity could be overcame by shortening the histone tail peptides and by selecting an NLS that prevents nucleolar accumulation. Cell penetration and probe release remain challenging as further studies would be needed to assess whether the molecular carriers considered (notably HA2-TAT, cR10 and SeL) significantly help. It would be interesting to couple them to the last -generation probes (including the negatively charged scaffold probes that did not show any cell penetration) and check their nuclear staining ability. Thus, the results reported here are promising starting points for the development of specific peptide-based LPPS probes for live cell manipulation.

Currently, to demonstrate unequivocally that a biomolecular compartment is formed via liquid-liquid phase separation, a multitude of criteria have to be satisfied. Each of these requirements needs to be addressed with a different method. Indeed, it is of fundamental importance to perform accurate experimental characterization of LLPS both *in vivo* and *in vitro*. Often, to visualize and investigate LLPS droplets in cells, the phase separating protein is overexpressed in fusion to a fluorescent tag. However, since LLPS transition is strictly dependent on concentration, the behavior of the endogenous protein might

not be correctly represented. On the other hand, our probes do not require any transfection steps, do not alter the endogenous protein concentration and do not perturb the system with genetically encoded tags.

In addition, to make our probes function as a sensor for LLPS, a fluorogenic dye, whose spectral features undergo a change depending on the environment, might be coupled to the probe instead of Silicon-rhodamine. For instance, Nile Red excitation and emission maxima were shown to undergo a shift towards shorter wavelengths in apolar solvents.[424] However, initial *in vitro* tests did not highlight any spectral shifts when the dye was included into the phHP1α droplets or remained in solution (data not shown), indicating that the polarity inside and outside these droplets is comparable. However, likely due to the droplet higher viscosity, differences in Nile Red fluorescence exponential decay rate were observed by fluorescence lifetime imaging microscopy (FLIM) (data not shown). Indeed, FLIM might be a precious technique to further investigate our probe's ability to detect LLPS in live cells.

Finally, another strength of such a mode of detection, that is, a peptide probe composed of multiple and interchangeable building blocks, is that the selectivity of the probe can be tuned by replacing the histone tails sequences with the ones characterizing the specific chromatin state of interest. For instance, one could target euchromatin sites by functionalizing the probe with acetylated histone tails, or direct the probe to DNA repair compartments to monitor the live DNA damage response. By labeling different probes with different fluorophores, this strategy potentially enables the concomitant visualization of multiple chromatin states and to monitor their dynamic behavior over time in live cells. In addition, it allows to investigate the spatial 3D arrangement of the targeted genome domains upon treatment with specific epigenetic modulators, inhibitors and drugs.

5.5   Acknowledgements

5.6   Appendix

RP-HPLC chromatograms and ESI-MS spectra of purified peptides and peptide probes synthesized in this work and not shown in **Chapter 5.3**:
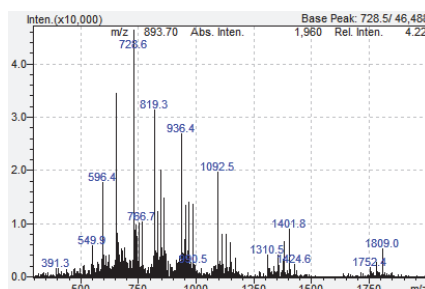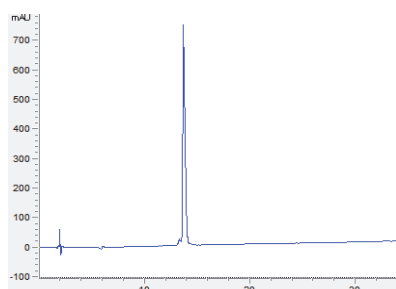
P1* (**4**): MW calculated: 6549.5 Da, MW found: 6548.8 Da
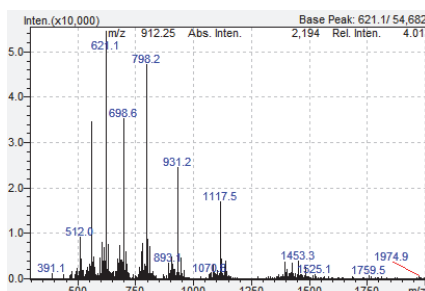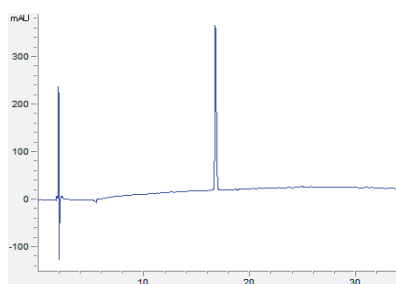
P3 (**10**): MW calculated: 5584.4 Da, MW found: 5583.3 Da
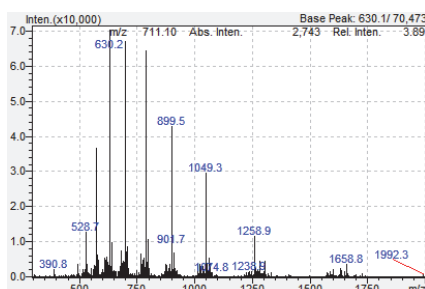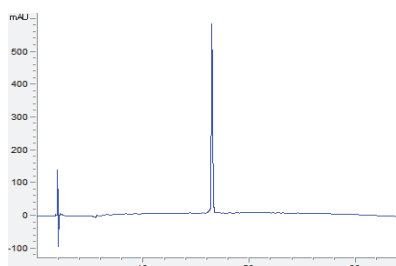
P4 (**11**): MW calculated: 6289.8 Da, MW found: 6289.4 Da
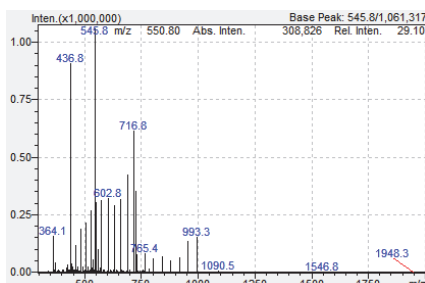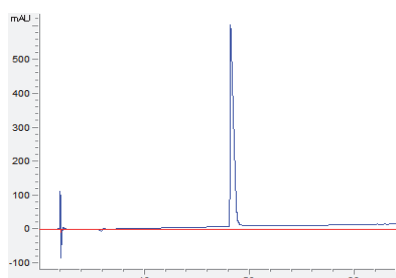
cR10 (**12**): MW calculated: 2179.2 Da, MW found: 2179.2 Da
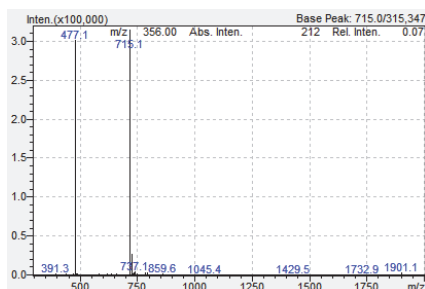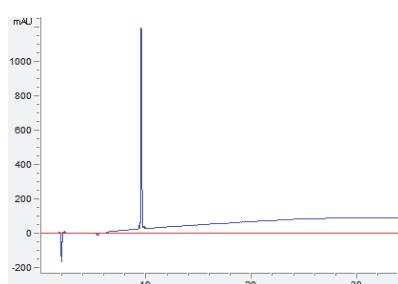
19*: MW calculated: 1428.6 Da, MW found: 1428.2 Da
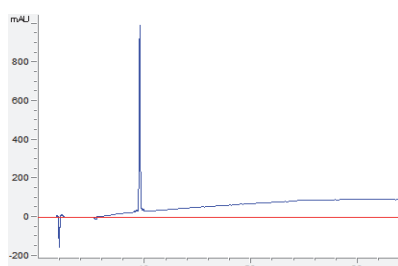
20*: MW calculated: 1201.3 Da, MW found: 1201.0 Da
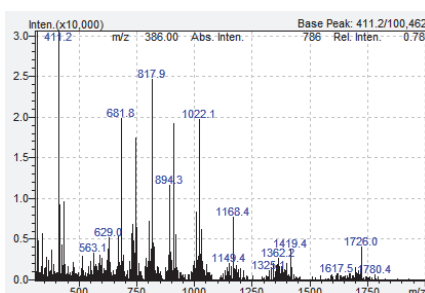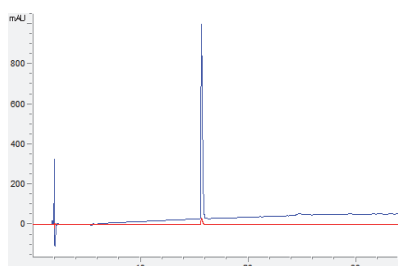
P7 (**24**): MW calculated: 8172.5 Da, MW found: 8171.8 Da
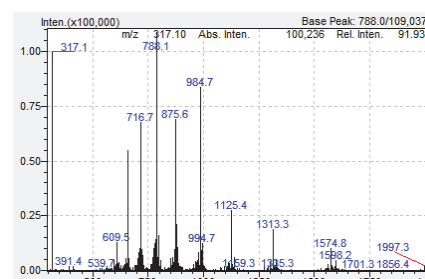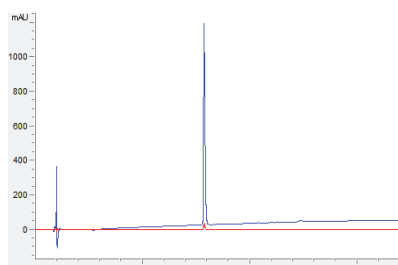
P7* (**24***): MW calculated: 7871.9 Da, MW found: 7870.8 Da
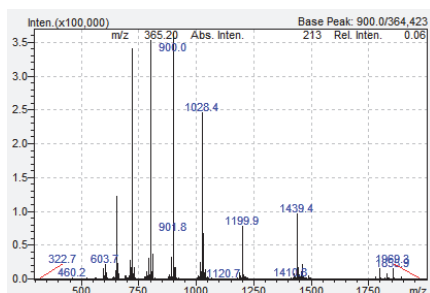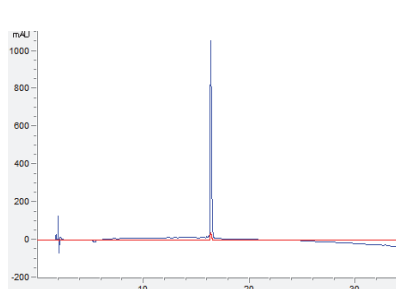
P8* (**25***): MW calculated: 7193.7 Da, MW found: 73192.4 Da
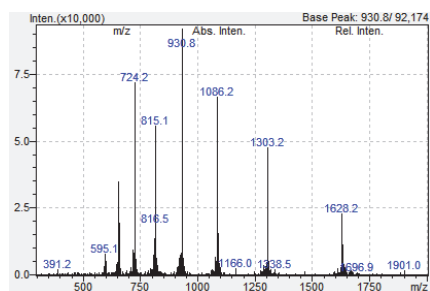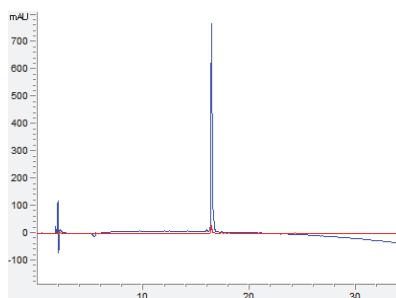
P9* (**26***): MW calculated: 6510.7 Da, MW found: 6511.2 Da
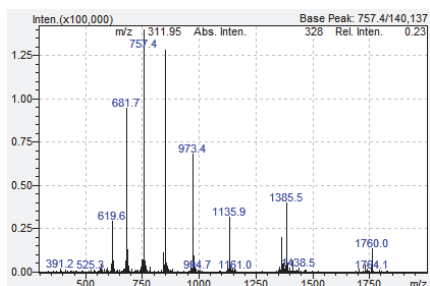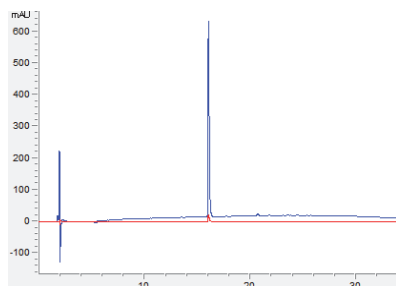
P9 (**26**): MW calculated: 6808.8 Da, MW found: 6807.4 Da
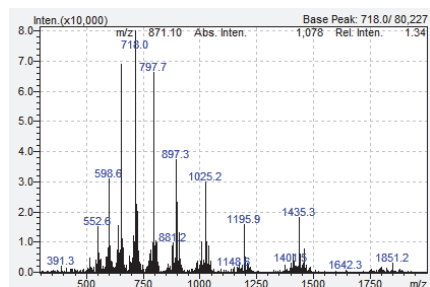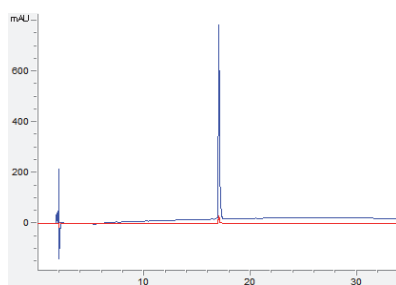
P10 (**27**): MW calculated: 7175.3 Da, MW found: 7174.2 Da

107
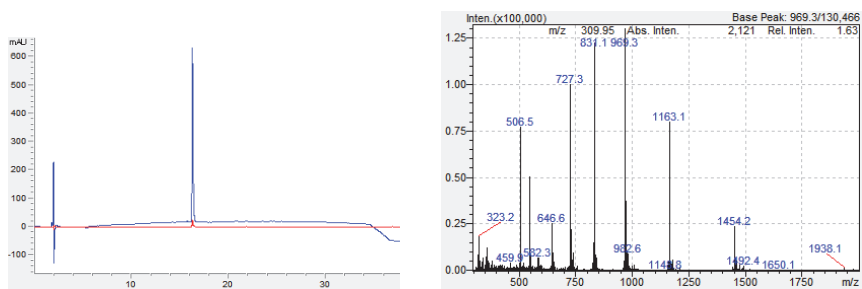
P11 (**28**): MW calculated: 5810.5 Da, MW found: 5810.4 Da

Chapter 6: Conclusions

Chemical biology methods provide precious tools for the deciphering of fundamental biological processes, such as gene regulation or disease. The understanding of the mechanisms underlying such signaling pathways sets the stage for the investigation of the molecular basis of life and is essential for the development of diagnostics and therapeutics.

Chromatin is a central signaling node that orchestrates the interpretation of the genetic information encoded in the DNA sequence. On one hand, its tight packaging and hierarchical organization controls the accessibility of the genes through compaction or transient exposure of the DNA. On the other hand, it can be heavily modified by PTMs that recruit chromatin effectors and trigger downstream biological responses. Patterns of histone PTMs define a complex language that establishes complex and intricated networks of interactions and signaling pathways. Elucidating the link between chromatin-effector interactions and the consequent occurrence of a defined phenotype in an organism is not straightforward. Indeed, for dissecting such mechanisms in detail, methods that allow to create controlled model systems are thus highly needed, both *in vitro* and in the cellular context.

Recent progress has shown that a special nucleosome configuration exists in cells, where histone pairs within a nucleosome carry different PTMs. However, the implications of such asymmetry remain enigmatic, as these substrates are not easily generated synthetically. Therefore, in the first part of my PhD project, I set out to generate pure, defined asymmetric nucleosomes. Based on a previously developed crosslinking strategy that forces two histone proteins to be included into the same nucleosome, I developed a semisynthetic procedure to produce nucleosomes carrying asymmetric H3K36me3 or H4K20me marks. These PTMs are found in stem cells, cancer cells and fibroblasts and play key roles in cell development and regulation. The strategy made use of the *lnc*-tag either at the N-terminus or at an internal position of the histone tail and enabled the efficient synthesis and reconstitution of asymmetric nucleosomes. Importantly, the strategy being very modular, it allows the generation of a library of asymmetric nucleosomes, where any amino acid (up to residue 46 in H3 and 37 in H4) can carry the PTM of interest. In comparison to previous methods relying on successive pulldowns with antibodies, our strategy can be used to generate asymmetric nucleosomes bearing PTMs for which no or insufficiently specific antibodies exist. Of note, the methodology can also be applied to the formation of any other multivalent complexes, not limited to histone proteins.

These substrates were used to investigate the activity of PRC2 and SET8, two key methyltransferases. Our asymmetric nucleosomes allowed to uncover crucial aspects of their regulation. SET8 appeared to access both faces of the nucleosome independently, as the presence of one H4K20me neither hindered nor

boosted the deposition of the modification on the other tail. On the other hand, PRC2 was found to be partially inhibited by the presence of asymmetric H3K36me3. In fact, the mark abolished the deposition of H3K27me3 on the same tail, but did not hinder the writing activity on the cognate histone tail. Surprisingly, although inhibitory, the H3K36me3 modification prolonged the PRC2 residence time on chromatin, thus suggesting an allosteric mechanism for the inhibition of the enzyme. Since PTMs are involved in all aspects of gene regulation and when dysregulates they can lead to many diseases including cancer, the details concerning the mode of action of chromatin effectors are of fundamental importance. These studies provide a powerful platform for deciphering the complex crosstalk between PTMs and their interacting partners and reveal mechanisms that govern human health and growth. Indeed, such mechanistic investigations pave the way for the development of novel therapeutics, such as allosteric inhibitors.

To gain further insights in gene regulation and PTM readout, biophysics represents a valuable tool that allows to study the binding dynamics and kinetics characterizing a defined interaction. Very often, such methods rely on the site-specific labeling of the interacting partners with bright and stable fluorescent dyes. Protein chemical modification is indeed a core discipline of the chemical biology field. Therefore, as a second project of my PhD, I collaborated on the development of a cysteine-specific modification chemistry that allows the dual orthogonal labeling of peptide and protein complexes, such as histone octamers. Using hypervalent iodine reagents (EBX) as a trivalent venue for double functionalization, thiol moieties in small molecules, peptides and proteins can be modified in a fast, efficient and chemoselective way. The reaction occurs under physiological conditions at room temperature and the formed vinylbenziodoxolone (VBX) intermediates are stable, biocompatible and do not interfere with protein function and structure, as exemplified by the reconstitution of labeled nucleosomes. Of note, this is the first time a hypervalent iodine reagent is reported to be bound to a protein substrate. Furthermore, the VBX intermediate contains two reactive functional groups, notably an azide and a hypervalent iodine, that orthogonally react with functionalized alkyne and boronic acid molecules via azide-alkyne cycloaddition and cross-coupling. Herein, this powerful dual-labeling methodology was exemplified by the incorporation of a triplet state quencher (TSQ) and cyanine fluorophore on a peptide substrate. Fluorescence microscopy has become a widely used and highly informative imaging technology to investigate and monitor biological and cellular processes in time and space. Therefore, ensuring the photo-stability of fluorophores throughout long-term laser exposure, as typically needed for live cell or single-molecule experiments is key. Our single-molecule studies revealed that the proximity of the TSQ significantly decreased the photobleaching rates of the fluorophore, corroborating the utility and applicability of our chemical modification approach. In addition to biophysics and chemical biology, protein conjugates have become essential tools for human medicine. Antibody–drug conjugates are a good example of such therapeutic agents and constitute an active area of research in the pharmaceutical industry. Our dual labeling method might find promising applications in such a field. Owing to its simplicity, this three-step modular procedure should facilitate the generation of a new class of bioconjugates with theranostic potential.

As emphasized above, understanding the mechanisms involved in gene expression and repression is of fundamental relevance for the design of effective drugs. Recently, the concept of liquid-liquid phase separation, playing a major role in cellular organization, has come into sharp focus. In fact, these membraneless assemblies compartmentalize the cell environment, speed up some reactions or physically separate other compounds and biomolecules. Phase separation thus represents a novel intriguing notion to explain the existence of stable but highly dynamic domains in cells. Specific chromatin states, such as heterochromatin, euchromatin and DNA damages sites, were also found to phase separate. This organizational concept has an enormous impact on how the processes involved in health and disease can be considered and provides a new point of view for the understanding of gene regulation. The field is developing fast and new biomolecules and cellular compartments are found to undergo phase separation on a regular basis. A tool that can assess the material properties of the condensates inside the cells, that allows to unequivocally detect liquid-liquid phase separation and distinguish it from gel-like states, is however currently lacking. Hence, I dedicated the last part of my PhD to the design of peptide-based probes for the investigation of chromatin organization in living cells. Designed for heterochromatin visualization, the probe can potentially be tuned to recognize other chromatin states, as the synthetic strategy is highly modular. This would enable the simultaneous detection of multiple chromatin states in space and time. The probe consists of a scaffold peptide functionalized with multiple histone tails bearing the PTMs that define a particular chromatin state. As chromatin effectors are known to be multivalent, they dynamically bind to the peptide probe causing its accumulation at chromatin spots that are enriched with the same endogenous PTM. Confocal imaging of live cells showed that two cell permeable probes, specific for the H3K9me3-enriched constitutive heterochromatin were generated, as confirmed by colocalization analysis with overexpressed fluorescent HP1α. These preliminary results are promising starting points for the generation of LLPS-specific probes for live cell chromatin visualization. As a potential outlook, a sensor of the cellular environment might be coupled to the probes. For instance, a fluorogenic dye whose photophysical properties are modulated according to the viscosity and/or polarity of the surroundings would provide precious information about the material properties of the targeted droplet.

In conclusion, I believe that this thesis adds some useful handles to the chemical biology toolbox and contributes to the general understanding of the histone code, providing new insight into the readout of key PTMs and bringing novel appreciation of gene regulation.

Chapter 7: Materials and Methods

6.1   Materials

All solvents and reagents were purchased from commercial sources and used without further purification. Fmoc-L-Lys(Boc,Me)-OH was from Iris Biotech (Marktredwitz, Germany). Fmoc-L-Lys(Me)$_3$-OH chloride, Fmoc-Lys(N$_3$)-OH, Fmoc-Pra-OH and Rink amide MBHA were from GL Biochem (Shanghai, China). Boc-L-thiazolidine-4-carboxylic acid was from Bachem (Bubendorf, Switzerland). All other amino acid derivatives, 2-chlorotrityl chloride resin and 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) were purchased from Novabiochem, Merck (Darmstadt, Germany). N,N-Dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), piperidine and phenol were from Acros Organics (Geel, Belgium). 2-(1H-benzotriazol-1-yl)-1,1,3,3- tetramethyluronium hexafluorophosphate (HBTU) was from Protein Technologies Inc. (Tucson, USA). (6-Chlorobenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HCTU) was from Carl Roth GMbH (Karsruhe, Netherlands). Hydrazine monohydrate was purchased from Alfa Aesar (Heysham, UK), acetonitrile (ACN) from Avantor Performance Materials (USA). 4-mercaptophenyl acetic acid (MPAA), Diethylether, acetic anhydride, phenylsilane, tetrakis(triphenylphosphine)palladium(0), hydroxybenzotriazole (HOBt), 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB), Dithiothreitol (DTT), silver acetate, trifluoroacetic acid (TFA), dichloromethane (DCM), triisopropylsilane (TIS), ethanedithiol (EDT), thioanisole, sodium nitrite, L-glutathione reduced (GSH), sodium diethyldithiocarbamate trihydrate and methyl thioglycolate (MTG), Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), dimethyl sulfoxide (DMSO), sodium ascorbate, DBCO-Cy5, Copper sulfate, Trolox, glucose oxidase and catalase were from Sigma Aldrich (Taufkirchen, Germany). 2,2'-Azobis[2-(2- imidazolin-2-yl)propane]dihydrochloride (VA-044) was purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). All other commonly used chemical reagents and buffer components were from Applichem (Darmstadt, Germany) and Fisher Scientific (Reinach, Switzerland). Restriction enzymes, Q5 DNA polymerase, Phusion polymerase, T5 exonuclease, Taq DNA ligase, dTNPs, DNA ladders and DNA loading dyes were from New England Biolabs (Ipswich, MA, USA). QiaQuick spin column for PCR purification, gel extraction and nucleotide removal as well as QiaPrep spin columns for miniprep plasmid purification were from Qiagen (Hilden, Germany). Primers were ordered from and synthesized by Integrated DNA technologies (Leuwen, Belgium) and gene sequencing was performed at GATC Biotech (Constance, Germany). Chemicals and solutions for preparation of agarose and SDS polyacrylamide gels (agarose, acrylamide, Precision Plus Protein™ All Blue Prestained Protein Standard) were purchased from BioRad (Hercules, CA, USA). Slide-A-Lyzer™ dialysis cassettes and Slide-A-Lyzer™ MINI dialysis devices were from Thermo Scientific (Rockford, IL, USA). Recombinant Set8 and PRC2 complex was purchased from Active Motif (La Hulpe, Belgium). Amersham Amplify Fluorographic Reagent was from GE Healthcare, S-[Methyl-^{3}H]-adenosyl-L-methionine (^{3}H-SAM) and Ultima Gold F liquid

scintillation cocktail from Perkin Elmer (Waltham, MA, USA). P81 Ion Exchange Cellulose Chromatography Paper was from Reaction Biology (Malvern, PA, USA). Mammalian cell culture media and components, such as Dulbecco's modified eagle's medium (DMEM)-GlutaMAX, Opti-MEM reduced serum medium, Phosphate-buffered saline (PBS), Trypsine-EDTA, Penicillin-Streptomycin (Pen-Strep) and Lipofectamine 2000 were from Gibco-Invitrogen (Basel, Switzerland).

6.2 Methods

6.2.1 Expression and purification of recombinant proteins

- *Histone proteins:*

Human wild-type core histones H2A, H2B, H3(C110A) and H4 and H4E63C mutant were expressed and purified as previously described.[310] N-terminal truncated H4 protein (H4(38-102)A38C) and H3 protein (H3(47-135)A47C) were recombinantly expressed in *E. coli* BL21(DE3) cells as N-terminal fusion to a His6-SUMO tag.[357] Bacterial cultures were grown in LB medium (with 50 μg/mL kanamycin and 35 μg/mL chloramphenicol) at 37 °C to an OD_{600} of 0.6-0.8. Protein expression was induced with 0.25 mM (isopropylthio-β-galactoside) IPTG and cells were incubated at 37 °C for 2-4 h. Cells were harvested by centrifugation (4000 x g, 4 °C, 15 min) and cell pellets resuspended in lysis buffer (200 mM NaCl, 20 mM Tris-HCl, 1 mM EDTA, 2 mM 2-mercaptoethanol, pH 7.5). Cells were lysed by sonication and centrifuged (15000 x g, 4 °C, 15 min). The pelleted insoluble inclusion bodies were washed twice with lysis buffer containing 1% Triton X-100 and once with lysis buffer without detergent. Inclusion bodies were solubilized in resolubilization buffer (6 M GdmHCl, 100 mM NaCl, 50 mM Tris-HCl, 5 mM imidazole, 250 mM L-ArgHCl, 2 mM 2-mercaptoethanol, pH 7.5), centrifuged (15000 x g, 4 °C, 15 min) to remove insoluble material and applied to Ni-NTA resin previously washed with 10 mM imidazole and equilibrated in resolubilization buffer. The protein was bound to the resin under gentle agitation overnight at 4 °C. The flow-through was collected and the resin was washed with 1 column volume (CV) resolubilization buffer and 1 CV urea wash buffer (6 M Urea, 150 mM NaCl, 25 mM Tris-HCl, 5 mM imidazole, 2 mM 2-mercaptoethanol, pH 7.5). The protein was eluted with 5 x 0.5 CV urea elution buffer (6 M Urea, 150 mM NaCl, 25 mM Tris-HCl, 300 mM imidazole, 2 mM 2-mercaptoethanol, pH 7.5). Washing and elution fractions were analyzed by SDS-PAGE (12% polyacrylamide gel) and protein containing fractions were pooled.

The SUMO protease His6-Ulp1 needed for cleavage of the His6-SUMO-H4(38-102)A38C and His6-SUMO-H3(47-135)A47C was expressed in *E. coli* BL21(DE3) cells. Bacterial cultures were grown in LB medium (with 50 μg/ml kanamycin) at 37 °C to an OD_{600} of 0.6-0.8. Protein expression was induced with 0.25 mM IPTG and cells were incubated at 37 °C for 3.5 h. Harvesting and lysis of cells, isolation of inclusion bodies and purification was performed as described above.

Dialysis of both SUMO protease and histone protein against dialysis buffer (1 M Urea, 150 mM NaCl, 75 mM Tris-HCl, 25 mM L-ArgHCl, 5 mM L-Cys, 2 mM 2-mercaptoethanol, pH 7.5) allowed the Ulp1 refolding and cleavage of the His$_6$-SUMO tag from the fusion protein. The protease and the protein of interest were mixed to a ratio 2:1 and dialyzed over night at 4 °C. In the case of truncated H4 preparation, the cleaved H4(38-102)A38C as well as His$_6$-Ulp1 precipitated during the dialysis, as confirmed by analytical RP-HPLC and ESI-MS. The precipitate was recovered by centrifugation (47000 x g, 4 °C, 15 min) and solubilized in resolubilization buffer. The proteins were applied to Ni-NTA resin previously washed with 10 mM imidazole and equilibrated in resolubilization (H4) or dialysis buffer (H3). His$_6$-Ulp1 was bound to the resin under gentle agitation for 20 min at 4 °C. The flow-through was collected and the resin was washed with 6 x 0.1 CV resolubilization buffer (supplemented with 15mM imidazole). The flow-through and washing fractions were analyzed by RP-HPLC, histone containing fractions were pooled and 10mM TCEP was added to avoid histone - 2-mercaptoethanol adducts. Crude truncated histone proteins were purified by semipreparative RP-HPLC and collected fractions were analyzed by analytical RP-HPLC and ESI-MS. Pure fractions were pooled, flash-frozen, lyophilized and stored at -20 °C.

- *Cys-ubiquitin*

Cys-ubiquitin was recombinantly expressed as N-terminal fusion to a His$_6$-tag in E. coli BL21(DE3)pLysS cells. Cells were grown in LB medium (supplemented with 100 μg/mL ampicillin as selection marker) at 37 °C until reaching an OD$_{600}$ of 0.6. Protein expression was induced with 0.2 mM IPTG and cells were further incubated at 37 °C for 4 hours. Subsequently, cells were harvested by centrifugation (4000 g, 4 °C, 10 minutes) and the cell pellet was resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8). Cells were lysed by sonication and cell lysate was centrifuged (15000 g, 4 °C, 15 minutes). The lysate supernatant was applied to Ni- NTA resin previously equilibrated with lysis buffer. The protein was bound to the resin by gentle nutating for 1 hour at 4 °C. The flow through was collected and the resin was washed with 2 x column volumes (CV) wash buffer I (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, 1 mM 2-mercaptoethanol, pH 8) followed by 2 x CV wash buffer II (50 mM sodium phosphate, 300 mM sodium chloride, 50 mM imidazole, 1 mM 2-mercaptoethanol, pH 8). Finally, the protein was eluted with 6 x 500 μL elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 250 mM imidazole, 1 mM 2-mercaptoethanol, pH 8). Fractions were analyzed by SDS-PAGE, protein-containing fractions were pooled. The protein was incubated at 4 °C for 30 minutes with 10 mM TCEP and submitted to preparative RP-HPLC purification. Collected fractions were analyzed by analytical RP-HPLC and ESI-MS. Pure fractions were pooled, lyophilized and stored at - 20 °C.

6.2.2 Fmoc-SPPS

The sequences of the synthesized peptides are summarized in **Table 2.** The general synthetic procedure and protocols for peptides that required special manipulations are described in the following.

Chapter	#	Sequence	Peptide
2	1	Thz-PATGGVK(me3)KPHRYRPGTV-NHNH$_2$	H3(29-46)C29Thz,K36me3
2	4	H-C(Acm)GSGSENLYFQARTKQTARKSTGGKAPRKQLATKAARKS-NHNH$_2$	lnc[Acm]H3(1-28)
2	4'	H-ARTKQTARKSTGGKAPRQLATKAARKS-NHNH$_2$	H3(1-28)
3	1	Ac-SGRGK(^{iso}lnc[Acm])-GGKGLGKGG-NHNH$_2$	acH4(1-14)K5^{iso}lnc[Acm]
3	1'	Ac-SGRGKGGKGLGKGG-NHNH$_2$	acH4(1-14)
3	2	H-CKRHRKVLRDNIQGITKPAIRRL-NHNH$_2$	H4(15-37)A15C
3	2a	H-CKRHRK(me)VLRDNIQGITKPAIRRL-NHNH$_2$	H4(15-37)A15C,K20me
4	4	Ac-GCRPKPQQFFGLM-NH$_2$	Substance P (SP)
5	1	H-RKKRRGRRRGSK(N$_3$)GSGSK(N$_3$)GSGSCC(Acm)-NH$_2$	CPP-scaffold
5	2	H-ARTKQTARK(me3)STGGKAGG(Pra)-NH$_2$	H3(1-15)K9me3,Pra
5	2*	H-ARTKQTARASTGGKAGG(Pra)-NH$_2$	H3(1-15)K9A,Pra
5	6	H-CGLFEAIAEFIEQGWEGLIEGWYGGRKKRRQRRR-NH$_2$	HA2-TAT
5	8	H-RGSK(N$_3$)GSGSK(N$_3$)GSGSCC(Acm)-NH$_2$	Scaffold (w/o CPP or NLS)
5	9	H-PKKKRKVGSK(N$_3$)GSGSK(N$_3$)GSGSCC(Acm)-NH$_2$	NLS-scaffold
5	12	H-GCSGSK(Alloc)RrRrRrRrRrE-NH$_2$	cR10
5	17	H-PAAKRVKLDGSK(N$_3$)GSK(N$_3$)GSK(N$_3$)GSCC(Acm)-NH$_2$	NLS2-scaffold
5	18	H-K(N$_3$)GEK(N$_3$)GEK(N$_3$)GECC(Acm)-NH$_2$	Scaffold (with negative charges)
5	19	H-ARTKQTARK(me3)STGG(Pra)-NH$_2$	H3(1-12)K9me3,Pra
5	19*	H-ARTKQTARASTGG(Pra)-NH$_2$	H3(1-12)K9A,Pra
5	20	H-TKQTARK(me3)STGG(Pra)-NH$_2$	H3(3-12)K9me3,Pra
5	20*	H-TKQTARASTGG(Pra)-NH$_2$	H3(3-12)K9A,Pra
5	21	H-QTARK(me3)STGG(Pra)-NH$_2$	H3(5-12)K9me3,Pra

Table 2: Sequences of peptides synthesized by Fmoc-SPPS.

- *Preparation of Fmoc-aa-hydrazine-Cl-trityl-resin*

Preparation of Fmoc-aa-NHNH-Cl-Trt-resin was performed as reported previously.[227] In general, 0.5 g 2-Cl-Trt-resin (0.82 mmol, 1 eq., substitution: 1.63 mmol/g) were swollen in 3 mL DMF for 15 min at room temperature. Subsequently, the resin was cooled to 0 °C and 1 mL of a solution containing DIEA (2.45 mmol, 427 μl, 3 eq.) and hydrazine monohydrate (1.64 mmol, 80 μl, 2 eq.) in DMF was added dropwise. The reaction mixture was stirred 1 h at room temperature. Then 100 μl methanol (MeOH) were added, the resin was stirred 10 min at room temperature, transferred to a reaction vessel for manual peptide synthesis and washed with DMF.

Due to the low stability of the hydrazine resin, the first amino acid Fmoc-aa-OH was coupled manually by standard Fmoc chemistry.[241] The loading of the 2-Cl-Trt-resin with hydrazine was assumed to be 50% (0.41 mmol, 1 eq.). Fmoc-aa-OH (2.05 mmol, 5 eq.) was first activated by addition of 3.9 mL of a 0.5 M HBTU solution (1.95 mmol, 4.76 eq.) in DMF followed by 2 min incubation at room temperature. Then, 714 μL DIEA (4.1 mmol, 10 eq.) were added and the reaction mixture was incubated another 1 min at room temperature. The activated amino acid was added to the dry resin, incubated 30 min at room temperature and washed with DMF. To ensure high-yield, the coupling procedure was repeated. Finally, the resin was washed with DMF, DCM and MeOH and dried under vacuum. Resin substitution was determined by treating a defined amount of resin with 20% piperidine in DMF for 30 min at RT, followed by spectrophotometrical quantification of released dibenzofulven-related species.

- *Automated Solid Phase Peptide Synthesis (SPPS)*

General protocol: the peptides were synthesized by the Tribute peptide synthesizer (PTI) on the previously prepared Fmoc-aa-hydrazine-Cl-Trt-resin or Rink amide resin to yield peptides with C-terminal hydrazide or amide, respectively. The syntheses were performed on 0.1 mmol scale using Fmoc chemistry. The standard base-resistant groups employed to protect amino acid side chains are listed below: Arg(Pbf), Lys(Boc), Thr(tBu), Gln(Trt), Ans(Trt), Asp(OtBu), His(Trt), Cys(Trt), Ser(tBu), Tyr(tBu), Glu(tBu). In addition, the following orthogonal protecting groups were employed: Cys(Acm), Lys(Alloc), Lys(Me), Lys(Me3), Lys(N3), Gly(Pra), Glu(OAll). To maximize synthesis yield, amino acids were double coupled and pseudoproline dipeptide building blocks were used where necessary.

Briefly, the N-terminal Fmoc-group was deprotected by incubating the resin with 20% (v/v) piperidine in DMF. Activation of amino acid (0.5 mmol, 5 eq.) was achieved by addition of HBTU or HCTU (0.48 mmol, 4.8 eq.) and DIEA (1 mmol, 10 eq.). The coupling step was performed by adding the activated amino acid to the resin, followed by 30 min incubation at room temperature. When the full-length peptide was assembled, the peptidyl-resin was washed with DMF, DCM and MeOH and dried under vacuum.

The peptide was cleaved from the resin using either 95% TFA, 2.5% TIS, 2.5% H_2O or 87.5% TFA, 5% phenol, 5% thioanisole, 2.5% ethanedithiol, 5% H_2O. The crude peptide was precipitated by addition of ice-cold diethyl ether, recovered by centrifugation, dissolved in 50% (v/v) acetonitrile in H_2O, flash-frozen and lyophilized.

Purification of the crude peptide was achieved with preparative RP-HPLC. Fractions were analyzed, pooled and lyophilized. Purified peptides ware analyzed and characterized by analytical RP-HPLC and ESI-MS.

- *Alloc deprotection*

The peptidyl-resin was swollen for 30 min in DCM. 1 mL of dry DCM and PhSiH$_3$ (2.4 mmol, 24 eq.), followed by Pd(PPh$_3$)$_4$ (0.025 mmol, 0.25 eq.) in 3 mL dry DCM were added to the resin and incubated for 30 min at room temperature. The resin was washed with DCM and the deprotection reaction with PhSiH$_3$ and Pd(PPh$_3$)$_4$ was repeated two more times. The resin was thoroughly washed with DCM followed by washing with 0.5% (v/v) DIEA in DMF; 0.5% (w/v) sodium-diethyldithiocarbamate in DMF; 50% (v/v) DCM in DMF; 0.5% (w/v) HOBt in DMF and intensively washed with DMF.

- *N-terminal acetylation*

Acetylation of the N-terminus was achieved by 3x30 min incubation of the peptidyl-resin with 15 mL of a solution containing 10% (v/v) acetic anhydride, 10% (v/v) DIEA in DMF.

- *Synthesis of acH4(1-14)K5(^{iso}lnc[Acm])*

This peptide was synthesized by L. C. C.. The H4 N-terminal 14 residues were synthesized by automated Fmoc-SPPS as reported above, on the pre-loaded Fmoc-Gly-hydrazine Cl-Trt resin and introducing the commercially available Fmoc-Lys(Alloc)-OH at position 5. Acetylation of the N-terminus was achieved as described above, followed by extensive washings with DMF and DCM. After Alloc deprotection of lysine 5, the glutamine residue was coupled manually to the ε-amino group of Lys5. Fmoc-Gln(Trt)-OH (0.5 mmol, 5 eq.) was activated by addition of 0.95 mL of a 0.5 M HATU solution (0.48 mmol, 4.8 eq.) in DMF followed by 2 min incubation at room temperature. Then, 172 µL DIEA (1 mmol, 10 eq.) were added and the reaction mixture was incubated another 1 min at room temperature. The activated amino acid was added to the peptidyl-resin, incubated 30 min at room temperature and washed with DMF. To ensure high- yield, the coupling procedure was repeated. Finally, the resin was washed with DMF. The rest of the synthesis was carried out by automated Fmoc-SPPS on the Fmoc-Gln(Trt) coupled to Lys5 side chain. Subsequent cleavage from the resin and purification was performed as described above.

- *Synthesis of cR10*

The linear peptide was synthesized by automated Fmoc-SPPS as reported above, on rink amide resin, with alternating Fmoc-L-Arg(Pbf)-OH and Fmoc-R-Arg(Pbf)-OH amino acids.
Alloc/Allyl deprotection: the peptidyl-resin was swollen for 30 min in DCM. 3 mL of dry DCM and PhSiH$_3$ (5 mmol, 50 eq.), followed by Pd(PPh$_3$)$_4$ (0.05 mmol, 0.5 eq.) in 0.5 mL dry DCM were added to the resin and incubated for 30 min at room temperature. The resin was washed with DCM and the deprotection reaction with PhSiH$_3$ and Pd(PPh$_3$)$_4$ was repeated two more times. The resin was thoroughly washed with DCM followed by washing with 0.5% (v/v) DIEA in DMF; 0.5% (w/v) sodium-diethyldithiocarbamate in

DMF; 50% (v/v) DCM in DMF; 0.5% (w/v) HOBt in DMF and intensively washed with DMF.

On-resin cyclization: glutamate side chain carboxylic acid was activated with a 0.5 M stock solution of HATU in DMF (1 eq., 0.1 mmol) and DIEA (8 eq., 0.8 mmol). The cyclization reaction was allowed to occur for 2h, under N_2 bubbling at room temperature. After extensive DMF washes, 2x500 µL 20% (v/v) piperidine were added to the resin for 5 min at room temperature to remove the N-terminal Fmoc. The resin was extensively washed with DMF, DCM and MeOH and dried under vacuum. The peptide was cleaved and purified as described above.

6.2.3 Preparation of asymmetric nucleosomes:

- *One-pot ligation and desulfurization*

One-pot ligation and desulfurization was achieved following Fang et al.,[266] but with some modifications. Typically, peptide hydrazides (1 µmol, 1 eq.) were dissolved in ligation buffer (6 M GdmHCl, 0.2 M NaH_2PO_4, pH 3) to a final concentration of 10 mM. The peptide solution was cooled to -20 °C in an ice/salt bath, followed by addition of $NaNO_2$ (1 µmol, 1 eq., aqueous 0.5 M stock solution) and incubation at -20 °C for 20 min. MTG was added to a final concentration of 75 mM (75 µmol, 75 eq.) in order to form the thioester *in situ*, the pH was adjusted to 6.8 with NaOH and the reaction mixture was stirred at room temperature for 10 min. Subsequently, the solution containing the peptide-MTG thioester was transferred to a tube containing the N-terminal cysteine-containing peptide (0.67 µmol, 0.67 eq.). TCEP was added to a final concentration of 25 mM (25 µmol, 25 eq.), the pH was adjusted to 6.8 and the ligation reaction was allowed to proceed for 3-16 h at 25 °C. The progress of the reaction was monitored by RP-HPLC and ESI-MS analysis. When the ligation was complete, desulfurization of the cysteine at the ligation site was performed in the same reaction tube without prior purification of the ligation product. TCEP desulfurization buffer (0.5 M TCEP, 6 M GdmHCl, 0.2 M phosphate, pH 7) was added to a final TCEP concentration of 0.25 M. The desulfurization reaction was initiated by addition of VA-044 and GSH to a final concentration of 30 mM and 40 mM, respectively. The reaction mixture was incubated at 42 °C for 6-16 h and the progress of the reaction was monitored by RP-HPLC and ESI-MS analysis. In the case of full length H4, a tenfold dilution of the reaction mixture in combination with a decreased incubation temperature (30 °C) prevented precipitation. When the desulfurization was complete, 2 volumes of ligation buffer (pH 3) were added and the peptide was purified by semipreparative RP-HPLC.

- *Conversion to MPAA thioester and purification*

This reaction was performed by L. C. C.. Typically, peptide hydrazide (1 µmol, 1 eq.) was dissolved in ligation buffer (6 M GdmHCl, 0.2 M NaH2PO4, pH 3) to a final concentration of 10 mM. The peptide solution was cooled to -18 °C in an ice/salt bath, followed by addition of NaNO2 (1 µmol, 1 eq., aqueous 0.5 M stock

solution) and incubation at -18 °C for 20 min. MPAA was added to a final concentration of 87 mM (8.7 µmol, 8.7 eq.) and 100 nM (10 µmol, 10 eq.), respectively, in order to form the thioester *in situ*, the pH was adjusted to 6.8 with NaOH and the reaction mixture was stirred at room temperature for 10 min. The peptide thioester was purified by semipreparative RP-HPLC. Fractions were analyzed, pooled accordingly, lyophilized and analyzed and characterized by analytical RP-HPLC and ESI-MS.

- *One-pot ligation and Thz deprotection*

This reaction was performed by L. C. C.. The thiazolidine protected fragment MPAA thioester (1.2 µmol, 3 eq.) was dissolved in ligation buffer (6 M GdmHCl, 0.2 M NaH2PO4, pH 7.4) to a final concentration of 15mM, TCEP was added to a final concentration of 50 mM (4 µmol, 10 eq.) and the pH was re-adjusted to 7.4. The recombinant truncated N-terminal cysteine-containing truncated histone was added to the reaction mixture (0.4 µmol, 1 eq.) to a final concentration of 5mM and the ligation reaction was allowed to proceed for 16 h at 25 °C. The progress of the reaction was monitored by RP-HPLC and ESI-MS analysis. When the ligation was complete, the *in-situ* conversion of Thz to Cys was initiated in the same reaction tube without prior purification of the ligation product. TCEP (2 µmol, 5eq.), MPAA (9 µmol, 22.5eq) and [Pd(allyl)Cl]$_2$ (9 µmol, 22.5 eq.) were added and the reaction mixture was stirred at 37 °C for 45min. In order to precipitate the Pd prior to purification, aqueous DTT to a final concentration of 250 mM was added and the solution was vortexed for 1 min and centrifuged (15000xg, 4 °C, 10min). The supernatant was separated and the pellet washed twice with 200 µL of ligation buffer. The combined fractions were pooled and purified by semipreparative RP-HPLC. Fractions were analyzed, pooled accordingly, lyophilized and analyzed and characterized by analytical RP-HPLC and ESI-MS.

- *Acm deprotection*

Silver acetate (AgOAc) was employed to remove the Acetamidomethyl (Acm) protecting group. The peptides were dissolved in 50% (v/v) acetic acid in H$_2$O to a final concentration of 0.5mM and AgOAc was added to a final concentration of 15mM from a stock solution in 50% (v/v) acetic acid in H$_2$O. The reaction mixture was incubated at 37 °C for 6h. Progression of the reaction was monitored by RP-HPLC and ESI-MS. When the reaction was complete, aqueous solution of DTT was added to a final concentration of 50mM in order to precipitate silver. The reaction was thoroughly vortexed, centrifuged (15000xg, 4 °C, 10min) and the supernatant was mixed with 3 volumes of ligation buffer (pH 3). After centrifugation (15000xg, 4 °C, 10min) the peptide was purified by semipreparative RP-HPLC and characterized by analytical RP-HPLC and ESI-MS.

- *Deprotection of Cys(Acm) and in situ activation of Cys with DTNB*

After treatment with AgOAc in 50% (v/v) AcOH for 6 h at 37 °C, an equal volume of a solution of 60 mM DTNB in 6 M GmdHCl, 0.2 M phosphate, pH 7 was added to the reaction mixture. The solution was thoroughly vortexed, incubated 5 min at room temperature and centrifuged (5 min, 4 °C, 15000 x g). Acetonitrile was added to the supernatant to a final concentration of 30% (v/v) and the solution was immediately subjected to semipreparative RP-HPLC purification. The pure protein was characterized by analytical RP-HPLC and ESI-MS.

- *Generation of heterodisulfide dimer proteins*

Proteins were dissolved in reaction buffer (6 M GdmHCl, 0.2 M phosphate, pH 6, degassed and flushed with argon) to a final concentration of 2.5 mM for the Cys(TNB)-protein and 2.3mM for the Cys-protein. Equal volumes of the two proteins were mixed together, incubated for 30s at room temperature and quenched by addition of 6M GmdHCl pH 2 and disulfide-linked asymmetric histone dimers were immediately purified by semipreparative RP-HPLC. Heterodisulfide formation was confirmed by analytical RP-HPLC and ESI-MS. Purified proteins were flash frozen, lyophilized and stored at -20 °C.

- *Refolding and purification of histone octamers*

Lyophilized recombinant or semisynthetic modified histone proteins were dissolved in unfolding buffer (6 M GdmHCl, 20 mM Tris-HCl, 1 mM EDTA, pH 7.5). 1eq. of H3 or H4 heterodisulfide histone dimers was mixed with 2 eq. of hH4 or hH3_C110A, respectively, and 2.2 eq. of hH2A and hH2B to a final protein concentration 0.5 mg/mL in unfolding buffer. For refolding of unmodified or symmetric octamers, 1eq. of H3 or H4 variants was mixed with 1 eq. of hH4 or hH3_C110A, respectively, and 1.1 eq. of hH2A and hH2B to a final protein concentration of 1 mg/mL in unfolding buffer. The solution was transferred to a Slide-A-Lyzer™ dialysis cassette (7k MWCO) and dialyzed at 4 °C overnight against refolding buffer (2 M NaCl, 20 mM Tris-HCl, 1 mM EDTA, pH 7.5). The refolded octamers were removed from the dialysis cassette and concentrated with a centrifugal concentrator Vivaspin500 (10K MWCO) to a final concentration of 40 μM. Subsequently, they were purified by size exclusion chromatography on a Superdex S200 10/300GL column and collected fractions were analyzed by SDS-PAGE (15% polyacrylamide gel). Octamer containing fractions were pooled and concentrated to a final concentration of 30-40 μM. After addition of glycerol to a final concentration of 50% (v/v), octamers were analyzed again by SDS-PAGE (15% polyacrylamide gel) under reducing (in presence of DTT) and non-reducing conditions (in the absence of DTT) and stored at -20 °C.

- *Reconstitution of nucleosomes*

Nucleosomes were reconstituted as described before.[104] Typically, the 153 base pair 601 nucleosome positioning DNA sequence [303] (75 pmol, 1 eq.) was mixed with the refolded octamers (1-1.2 eq.) at high salt concentration and reconstitution was achieved by gradient dialysis into low salt conditions (10 mM KCl, 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.5). Dialysis was performed in Slide-A-Lyzer™ MINI dialysis devices using a peristaltic pump at a flow rate of 1ml/min over 18h at 4 °C. Subsequently, DTT to a final concentration of 1mM was added and nucleosomes incubated at 4 °C for 30 min.

- *Test of TEV isopeptide bond cleavage on H4 peptide*

This reaction was performed by L. C. C.. 0.2 mg peptide (79 nmol) were dissolved in 100 μL TEV solution (1 mg/ml TEV protease in 200 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl pH8, 2 mM DTT) and incubated at 30 °C under gentle agitation. This resulted in a molar ratio of 1:22 TEV/peptide. The cleavage reaction was monitored by analytical RP-HPLC and MS.

- *TEV cleavage on nucleosomes*

H4 asymmetric nucleosomes: multiple additions of TEV protease (4 x 1 μg per 100 pmol nucleosomes) over a total time of 24 h at room temperature allowed to remove the *isolnc*-tag.

H3 asymmetric nucleosomes: TEV protease was added (1 μg per 60 pmol of nucleosomes) and allowed to cleave the lnc-tag overnight at 4 °C.

Final nucleosomes concentrations were determined by UV quantification and the reconstitution was checked by native gel electrophoresis. Removal of the lnc-tag by TEV protease was confirmed by SDS-PAGE (17% polyacrylamide gel).

6.2.4 <u>Methyltransferase assays with Set8 or PRC2</u>

- *Endpoint experiments*

1 pmol of recombinant Set8 or 1.6 pmol PRC2 complex was incubated at RT for 2 h with 15 pmol of nucleosomes and 60 pmol ^{3}H-SAM (1 μCi) in 50 mM Tris-HCl pH 8.5, 10 mM NaCl, 2 mM MgCl$_2$, 1 mM EDTA, 5 mM DTT (30 μL total volume). Salt concentration was adjusted to 30 mM. Subsequently, histones were separated by SDS-PAGE (13% polyacrylamide gel, 20 μL) and stained with Coomassie Brilliant Blue G250. After destaining, the gel was incubated for 30 min in Amersham Amplify Fluorographic Reagent and dried for 2 h at 80 °C. Incorporation of ^{3}H in histone substrates was detected by overnight exposure to an X-ray

film at -80 °C. The activity of each reaction was quantified by scintillation counting. 5 µL reaction mixture were applied to P81 Ion Exchange Paper. The filters were dried at RT for 40 min, washed 3 x with 50 mM NaHCO$_3$ pH 9 and dried on a gel dryer for 2 h at 80 °C. Scintillation counting was performed with 5 mL Ultima Gold F scintillation cocktail on a Packard Tri-Carb Liquid Scintillation Counter.

- *Kinetic studies*

0.25 pmol of recombinant Set8 or 0.5 pmol PRC2 complex were incubated at RT for 2 h with 15 pmol or 22.5pmol of nucleosomes, respectively, and 60 pmol ^{3}H-SAM (1 µCi) in 50 mM Tris-HCl pH 8.5, 10 mM NaCl, 2 mM MgCl$_2$, 1 mM EDTA, 5 mM DTT (30 µL total volume). Salt concentration was adjusted to 30 mM. 5 µL time point samples were quenched with 5 µL 0.2% TFA and applied to P81 Ion Exchange Paper. Filter washing and scintillation counting were performed as described above.

6.2.5   Synthesis of N$_3$-EBX reagent JW-RF-010

The synthesis of JW-RF-010 was performed by T. R..

- *Preparation of (4-Azidobut-1-yn-1-yl)trimethylsilane*
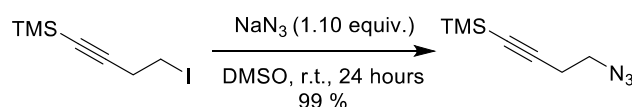
Following a slightly modified procedure,[220] (4-iodobut-1-yn-1-yl)trimethylsilane (34.9 g, 138 mmol, 1.00 eq.) was added to a 0.5 M solution of sodium azide in dimethyl sulfoxide (NaN$_3$, 304 mL, 152 mmol, 1.10 eq.). The reaction mixture was stirred for 24 hours at room temperature, then slowly poured into a mixture of ice:water (800 mL). The aqueous layer was extracted with diethyl ether (3 x 300 mL) and the combined organic layers were washed with water (2 x 200 mL), brine (200 mL), dried over magnesium sulfate, filtered and concentrated under reduced pressure. The light-yellow crude liquid was purified through a plug of silica, eluting with pentane, to afford pure (4-azidobut-1-yn-1-yl)trimethylsilane (22.8 g, 136 mmol, 99% yield) as a colorless liquid.

- *(4-azidobut-1-ynyl)-1,2-benziodoxol-3(1H)-one*
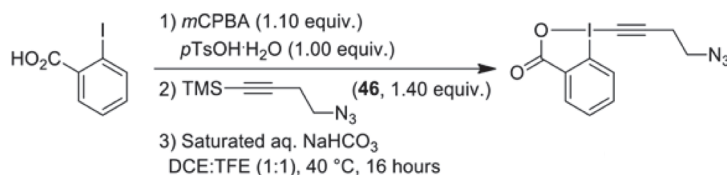
Following a reported procedure,[220] 2-iodobenzoic acid (24.1 g, 97.0 mmol, 1.00 eq.), *para*-toluene sulfonic acid monohydrate (*p*TsOH.H$_2$O, 18.5 g, 97.0 mmol, 1.00 eq.) and *meta*-chloroperoxybenzoic acid (*m*CPBA-

77%, 23.9 g, 107 mmol, 1.10 eq.) were dissolved in a mixture of dichloroethane (81 mL) and 2,2,2-trifluoroethanol (81 mL). After 1 hour stirring at 40 °C, (4-azidobut-1-yn-1- yl)trimethylsilane (22.7 g, 136 mmol, 1.40 eq.) was added in one portion. The reaction mixture was stirred for an additional 14 hours at the same temperature, then the resulting suspension was filtered and the volatiles were removed under reduced pressure. The resultant residue was dissolved in dichloromethane (1000 mL) and treated with a solution of saturated aqueous sodium bicarbonate (1000 mL). The mixture was vigorously stirred for 1 hour, then the two layers were separated and the aqueous layer was extracted with additional portions of dichloromethane (3 x 500 mL). The organic layers were combined, dried over magnesium sulfate; filtered and concentrated under reduced pressure. Purification by column chromatography (SiO$_2$, ethyl acetate) afforded the N$_3$-EBX reagent JW-RF-010 (4-azidobut-1-ynyl)-1,2-benziodoxol-3(1H)-one (5.23 g, 15.3 mmol, 16% yield) as a white solid.

6.2.6 One-pot EBX labeling and click reaction

- *"Post-click" reaction on peptide HAS*

HSA was dissolved in 10 mM Tris buffer pH 8.2 (1 eq., 1.20 μmol) to a final concentration of 1 mM. The resulting solution was vortexed few seconds and a 200 mM solution of N$_3$-EBX reagent JW-RF-010 in DMSO (0.91 eq., 1.09 μmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Subsequently, a 20.0 mM solution of DBCO-Cy5 in DMSO (1 eq., 1.20 μmol) was added in one portion. The resulting mixture was vortexed and shaken at 300 rpm at room temperature for 60 minutes. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis. No effort was made to exclude oxygen.

- *"Post-click" reaction on peptide SP*

SP (1 eq., 645 nmol) was dissolved in 50 mM phosphate buffer pH 8.2 (containing 10% (v/v) DMSO) to a final concentration of 0.5 mM. The resulting solution was vortexed few seconds and a 200 mM solution of N$_3$-EBX reagent JW-RF-010 in DMSO (1.3 eq., 838.5 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Subsequently, a 20.0 mM solution of DBCO-Cy5 in DMSO (1.5 eq., 967.5 nmol) was added in one portion. The resulting mixture was vortexed and shaken at 300 rpm at room temperature for 15 minutes. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis. No effort was made to exclude oxygen.

- *"Pre-click" reaction on cys-ubiquitin*

200 mM solution of N_3-EBX reagent JW-RF-010 in DMSO (2.00 eq., 100 nmol) was diluted with 10 mM Tris buffer pH 8.2 to a final concentration of 2 mM. The resulting solution was vortexed few seconds and a 20 mM solution of DBCO-Cy5 in DMSO (2.5 eq., 125 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Subsequently, a 1 mM solution of cys-ubiquitin in water (1 eq., 50 nmol) was added to the reaction mixture. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis. No effort was made to exclude oxygen.

- *"Post-click" reaction on cys-ubiquitin under native conditions*

Cys-ubiquitin (1 eq., 50 nmol) was dissolved in 10 mM Tris buffer pH 8.2 to a final concentration of 0.5 mM. The resulting solution was vortexed few seconds and a 10 mM solution of N_3-EBX reagent JW-RF-010 in Tris buffer pH 8.2 (containing 5% v/v DMSO) (2 eq., 100 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Subsequently, a 20.0 mM solution of DBCO-Cy5 in DMSO (1 eq., 50 nmol) was added in one portion. The resulting mixture was vortexed and shaken at 300 rpm at room temperature for 60 minutes. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis. No effort was made to exclude oxygen. Labeling was further confirmed by SDS-PAGE (17% polyacrylamide gel) and fluorescence microscopy analysis.

- *"Post-click" reaction on cys-ubiquitin under denaturing conditions*

Cys-ubiquitin (1 eq., 50 nmol) was dissolved in denaturing phosphate buffer (6 M GmdHCl, 200 mM phosphate, pH 8.2) to a final concentration of 0.5 mM. The resulting solution was vortexed few seconds and a 10 mM solution of N_3-EBX reagent JW-RF-010 in denaturing phosphate buffer (containing 5% v/v DMSO) (1.1 eq., 55 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Subsequently, a 20.0 mM solution of DBCO-Cy5 in DMSO (1 eq., 50 nmol) was added in one portion. The resulting mixture was vortexed and shaken at 300 rpm at room temperature for 60 minutes. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis. No effort was made to exclude oxygen.

- *Control reaction on non-mutated ubiquitin*

Ubiquitin (1 eq., 50 nmol) was dissolved in denaturing phosphate buffer (6 M GmdHCl, 200 mM phosphate, pH 8.2) to a final concentration of 0.5 mM. The resulting solution was vortexed few seconds and a 10 mM

solution of N$_3$-EBX reagent JW-RF-010 in denaturing phosphate buffer (containing 5% v/v DMSO) (1.1 eq., 55 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis.

- *"Post-click" reaction on histone octamers*

A 40 µM H4E63C octamers solution stock (1 eq., 0.8 nmol) in refolding buffer (2.00. M NaCl, 10.0 mM Tris-HCl, 1.00 mM EDTA, pH 8.3 at 4 °C) were incubated with a 1 mM DTT solution in water (1 eq., 0.8 nmol), for 20 min at room temperature. Subsequently, a 5 mM N$_3$-EBX reagent JW-RF-010 stock solution in DMSO (5 eq., 3.2 nmol) was added to the reaction mixture and incubated for 1 h at room temperature. The reaction was monitored by analytical RP-HPLC and product identity confirmed by ESI-MS analysis. When the reaction was complete, excess N$_3$-EBX reagent was quenched with the addition of a 10 mM GSH aqueous solution (10 eq. 8 nmol) Subsequently, to 10 µL of the crude H4-EBX octamers, a 10 mM DBCO-Cy5 stock solution in DMSO (8 eq., 6.4 nmol) was added and the reaction mixture and incubated for 15 min at room temperature. The reaction was monitored by analytical RP-HPLC and product identity confirmed by ESI-MS analysis. Specific cysteine labeling was confirmed by SDS-PAGE (17% polyacrylamide gel) and fluorescence microscopy analysis. Cy5-labeled octamers were transferred to a Slide-A-Lyzer™ MINI dialysis device and dialyzed overnight against refolding buffer (2 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, pH 8.3 at 4 °C). H4 EBX-labeled octamers were stored at 4 °C and used to reconstitute nucleosomes.

6.2.7   Synthesis of ligand and preparation of Pd complex for Suzuki-Miyaura cross-coupling reaction

The synthesis of ligand and Pd complex was performed by T. R..

- *Synthesis of 2-(Dimethylamino)pyrimidine-4,6-diol*
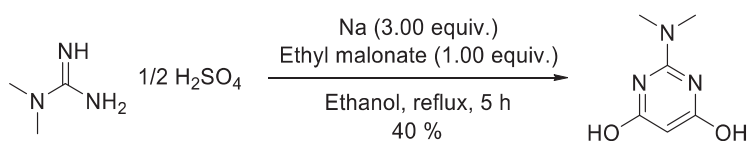
Following a reported procedure,[425] to a solution of sodium (Na, 93.0 mg, 4.04 mmol, 1.10 eq.) in ethanol (1.9 mL, 2.12 M), dimethyl guanidine sulfate (1.00 g, 3.67 mmol, 1.00 eq.) was added. The resulting solution was added to another solution of sodium (Na, 0.160 g, 6.98 mmol, 1.90 eq.) and diethyl malonate (0.560 mL, 3.67 mmol, 1.00 eq.) in ethanol (1.96 mL, 1.87 M). The combined solution was refluxed for 5 hours. The reaction was then evaporated to dryness, dissolved in water (5 mL) and taken to pH 6 with acetic acid.

Collection of the white solid formed by filtration under vacuum afforded the ligand product (0.225 g, 1.45 mmol, 40%).

- *Synthesis of Pd complex*
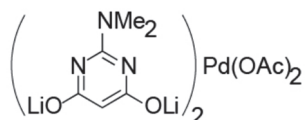
Palladium catalysts was synthesized following reported procedures by T. R..[383] 2-(Dimethylamino)pyrimidine-4,6-diol based ligand (20.0 μmol) was dissolved in an aqueous solution of LiOH (0.4 mL, 0.1m) in an ultrasonic bath for 2 minutes. The palladium source (10.0 μmol) was added and the mixture was magnetically stirred at 65 °C for 30 minutes. Deionized water (0.6 mL) was then added to afford a 20.0 mM catalyst solution.

6.2.8 One-pot EBX labeling and cross-coupling on cys-ubiquitin

A 1.4 mM solution of cys-ubiquitin in water (1 eq., 7 nmol) was diluted with 100 mM phosphate buffer pH 8.2 to a final concentration of 0.7 mM. The resulting solution was vortexed and a 1.4 mM solution of N_3-EBX reagent JW-RF-010 in 50 uM phosphate buffer pH 8.2 (containing 5% (v/v) DMSO) (2 eq. 14 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Separately, a 25 mM stock solution of (*para*-methoxyphenyl)boronic acid in 50 mM phosphate buffer pH 8.2 (containing 2.5% (v/v) DMSO) was prepared by heating up the mixture to 50 °C, until obtaining a clear solution. After cooling down to 37 °C, an aliquot of the phenyl boronic acid solution (350 nmol, 50.0 eq.) was added to the protein sample. The resulting solution was vortexed to ensure proper reagent mixing and incubated at 37 °C for 10 minutes. Then, a 40 mM solution of bis-lithium-2-(dimethylamino)-4,6-dihydroxylate-pyrimidine palladium diacetate complex in water (20 eq., 140 nmol) was added in one portion. The resulting solution was vortexed to ensure proper reagent mixing and incubated 30 minutes at 37 °C. The reaction was quenched with a 1.14 M solution of 3-mercaptopropionic acid in water (0.8 μmol, 5.70 eq. per eq. of palladium) and shaken at room temperature for 10 minutes. No effort was made to exclude oxygen. The reaction was monitored by analytical HPLC and products identity confirmed by ESI-MS analysis.

6.2.9 One-pot EBX labeling/click/cross-coupling on cys-ubiquitin

A 1.4 mM solution of cys-ubiquitin in water (1 eq., 7 nmol) was diluted with 100 mM phosphate buffer pH 8.2 to a final concentration of 0.7 mM. The resulting solution was vortexed few seconds and a 1.4 mM solution of N_3-EBX reagent JW-RF-010 in 50 uM phosphate buffer pH 8.2 (containing 5% (v/v) DMSO) (2 eq. 14 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. Subsequently, a 20 mM solution of

DBCO-Cy3 in DMSO (5 eq., 35 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 2 hours. Separately, a 25 mM stock solution of (*para*-methoxyphenyl)boronic acid in 50 mM phosphate buffer pH 8.2 (containing 2.5% (v/v) DMSO) was prepared by heating up the mixture to 50 °C, until obtaining a clear solution. After cooling down to 37 °C, an aliquot of the phenyl boronic acid solution (350 nmol, 50.0 eq.) was added to the protein sample. The resulting solution was vortexed to ensure proper reagent mixing and incubated at 37 °C for 10 minutes. Then, a 40.0 mM solution of bis-lithium-2-(dimethylamino)-4,6-dihydroxylate-pyrimidine palladium diacetate complex in water (20 eq., 140 nmol) was added in one portion. The resulting solution was vortexed to ensure proper reagent mixing and incubated 30 minutes at 37 °C. The reaction was quenched with a 1.14 M solution of 3-mercaptopropionic acid in water (0.8 µmol, 5.7 eq. per eq. of palladium) and shaken at room temperature for 10 minutes. The reaction was monitored by analytical HPLC and product identity confirmed by ESI-MS analysis. No effort was made to exclude oxygen.

6.2.10 Synthesis of Trolox boronic acid

The synthesis of Trolox boronic acid was performed by C. J..
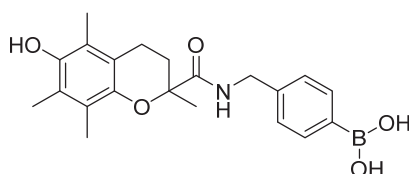
To a round bottom flask charged with 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (20.0 mg, 0.08 mmol, 1.0 eq.) in DMF (0.47 mL, 0.17 M), TSTU (2-(2,5-dioxopyrrolidin-1-yl)-1,1,3,3-tetramethylisouronium tetrafluoroborate) (28.9 mg, 0.096 mmol, 1.20 eq.) and NEt$_3$ (32.3 µL, 0.320 mmol, 4.00 eq.) were added. 4-boronic acid phenylmethanamine (12.1 mg, 0.080 mmol, 1.00 eq.) at room temperature was then added and the reaction mixture was stirred for 16 hours. The reaction was then submitted to preparative HPLC to obtain (4-((6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxamido)methyl)phenyl)boronic acid as a white solid (15.3 mg, 0.040 mmol, 50%).

6.2.11 EBX labeling, cross-coupling and click reaction on peptide SP

This reaction was performed by C. J.. SP (1 eq., 2.66 µmol) was dissolved in 2.250 mL water and 2.25 mL of 100 mM phosphate buffer pH 8.2. A 30 mM solution of N$_3$-EBX (**1a**) (1 eq., 2.66 µmol) was added and the reaction was vortexed and shaken at room temperature for 3 hours. A 20 mM solution of bis-lithium-2-(dimethylamino)-4,6-dihydroxylate-pyrimidine palladium diacetate complex in water (5 eq., 13.3 µmol) was added in one portion, followed by a 7.5 mM solution of the Trolox boronic acid in DMSO (10 eq., 26.6 µmol). The reaction was then shaken for 4 hours at room temperature before quenching it with a 7.95 mM solution of mercaptopropionic acid (28.5 eq., 75.8 µmol) at 37 °C for 10 minutes. The reaction was then diluted with 100 mL of water and left to dry on the lyophilizer. The crude product was resuspended in 2 mL of DMSO, passed through an HPLC filter and purified by preparative RP-HPLC. No effort was made to

exclude oxygen at any step. To 20 mM solution of the Trolox-SP conjugate in DMSO (11.1 µL), another 20 mM solution of DBCO-Cy5 in DMSO (11.1 µL) was added. The reaction was vortexed and incubated at room temperature for 1.5 hours. No effort was made to exclude oxygen. The reaction was analyzed by HPLC-MS. The crude product was used for subsequent studies as a 10 mM solution without need for further purification.

6.2.12 SP Biotinylation

A 5 mM Cy5-SP or Cy5-Trolox-SP solution in DMSO (1 eq., 65 nmol) was diluted to 1 mM in 50 mM phosphate buffer pH 8.2. The resulting solution was vortexed and a 10 mM solution of biotin-NHS in DMSO (1 eq., 65 nmol) was added in one portion. The resulting mixture was vortexed to ensure proper reagent mixing and shaken at 300 rpm at room temperature for 60 minutes. 10 µL of the reaction mixture were diluted to 100 µL with 6.00 M GdmHCl, pH 3 and purified by analytical RP-HPLC. Fractions were analyzed, pooled accordingly and lyophilized. Purified SP-Cy5-bt conjugate and SP-Cy5-Trolox-bt were dissolved in DMSO to a final concentration of 1 mM, aliquoted and stored at -20 °C.

6.2.13 One-pot SiR-mal cysteine labeling and Copper catalyzed click reaction

Scaffold peptide (1 eq., 200 nmol) was dissolved in 50 mM phosphate buffer pH 7.3 to a final concentration of 0.5 mM. A 20 mM SiR-mal solution in DMSO (1.2 eq., 240 nmol) was added, the reaction was vortexed and shaken at room temperature for 15min. Alkyne-containing histone tail peptide was added (1:2 azide:alkyne molar ratio), followed by addition of a 100 mM solution of $CuSO_4$ (4 eq., 800 nmol) and a 500 mM solution of sodium ascorbate (20 eq., 4 µmol). The reaction mixture was vortexed and shaken at room temperature for 5min and immediately purified by semipreparative RP-HPLC. Product identity was confirmed by RP-HPLC and ESI-MS analysis.

6.2.14 SeL labeling

5 µL of 100 mM SeL-NHS solution in DMSO (2 eq., 500 nmol) were mixed with 2.5 µL of 100 mM *N*-(2-aminoethyl)maleimide in DMSO (1 eq., 250 nmol). The solution was diluted to 25 µL with 100 mM phosphate buffer pH 8.2. After 1h incubation at room temperature, excess glycine (5 eq., 1.25 µmol) was added to quench the remaining SeL-NHS. Acm-deprotected peptide probe (1 eq., 100 nmol) was dissolved in 100 mM phosphate buffer, pH 8.2, to a final concentration of 0.5 mM in a siliconized 1.5 mL tube. Crude SeL-mal (2 eq., 200 nmol) was added and the reaction mixture was vortexed and shaken at room temperature for 15min. The SeL-probe peptide was purified by semipreparative RP-HPLC and lyophilized. Product identity was confirmed by RP-HPLC and ESI-MS analysis.

6.2.15 smTIRF imaging and data processing

- *PRC2 studies*

These experiments were performed by B. A. L.. Chromatin fibers were reconstituted similarly as nucleosomes and as previously described.[310,331] As DNA template, a 12-mer repeat of the 177 bp 601 sequence was used, which was ligated to an anchor oligonucleotide containing a biotin for immobilization and an Atto647N dye for localization.[310,331] To avoid overloading chromatin fibers with histone octamers, 0.5 eq. MMTV buffer DNA was added to the reconstitution mixture. After reconstitution, the chromatin fibers were immobilized within polyethylene-glycol (PEG) passivated flow channels using neutravidin–biotin attachment chemistry. The imaging buffer composition was 30 mM HEPES, pH 7.5, 50 mM NaCl, 10% (w/v) glycerol, 0.005% (v/v) Tween 20, 2 mM Trolox, 0.32% (w/v) glucose, 1 mg/ml BSA, 1 mg/ml glucose oxidase and 40 μg/ml catalase in degassed water. Molecules were imaged using a Nikon Eclipse Ti-E microscope with a manual TIRF illuminator and equipped with an Andor iXon EMCCD camera. Movies were recorded with a 50-ms integration time per frame. Positions of chromatin fibers were determined every 200 images using a 640-nm laser line (Coherent Obis). PRC2 binding dynamics were observed using a 532-nm laser line (Coherent Obis). Typical movies contained 5,000 frames. During the measurements, laser intensities of 40 W/cm^2 were employed for 532 nm and 20 W/cm^2 for 640 nm. Data processing was performed as described[310,331] using custom-made Matlab scripts (Mathworks). After background and drift correction, individual kinetic traces were extracted for each chromatin array, and binding events were detected by a thresholding algorithm. Each detected binding event was fitted with a 2D Gaussian function. Events were excluded if the width of the point-spread function was found to exceed a threshold value for a single-molecule detection or if the center of the fitted Gaussian was offset above a distance cutoff.

To determine residence times, cumulative histograms were generated from t_{bright}, from traces corresponding to individual chromatin arrays and corrected for DY-547 photobleaching by dividing the histogram by a single-exponential decay with a time constant τ = 18s.[331] Finally, the corrected lifetime histograms were fitted using double-exponential function. Each reported value was obtained from at least three independent experimental replicates.

- *SP-Cy5-Trolox-bt studies*

SP-Cy5-Trolox-bt or SP-Cy5-bt were immobilized within polyethylene-glycol (PEG) passivated flow channels using neutravidin–biotin attachment chemistry.[310] The imaging buffer composition was 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM 2-mercaptoethanol, 3.2% (w/v) glucose, 1 mg/mL glucose oxidase, 40 μg/mL catalase in ultrapure degassed H$_2$O, filtered. Surface-immobilized molecules were imaged using a fully automated Nikon Eclipse Ti-E inverted fluorescence microscope with a manually controlled TIRF illuminator arm and equipped with a CFI Apo TIRF 100x Oil immersion objective (NA 1.49). Data was acquired using an Andor iXon EMCCD camera. Movies were recorded with a 100-ms

integration time per frame. Typical movies contained 1800 frames. SP-Cy5-Trolox-bt or SP-Cy5-bt photobleaching was observed using a 640-nm laser line (Coherent Obis). During measurements, a laser intensity 0.74 mW was employed.

For data analysis, a custom-made Matlab script was employed.[310] In short, a baseline correction and a drift correction was performed. A peak-finding algorithm was employed to detect individual SP peptide positions. Fluorescence intensity traces for each SP peptide position were obtained by integrating over a circle of 2 pixel radius. Traces over 200-300 individual molecules were averaged and analyzed by a single-exponential decay to obtain bleaching rates or time constants.

6.2.16 <u>Cell culture and confocal microscopy</u>

- *HEK 293T*

HEK 293T NK1-GFP stable cell lines were plated on poly-lysine coated 12-well plates (seeding density 10^5 cell/well, surface area 3.5 cm^2) and cultured for 36 h at 37 °C with 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and Penicillin-Streptomycin antibiotics (Penicillin G Sodium Salt: 100 µg/mL, Streptomycin Sulfate: 100 µg/mL). Cells were washed with Opti-MEM Reduced Serum Medium and incubated with 1 mL of 20 nM SP-Cy5 solution in Opti-MEM. Time-lapse images were acquired on a confocal microscope LSM 510 Meta (Zeiss), using a C-Apochromat 63x/1.2 water corrected objective. Imaging was performed using the following parameters: pixel time: 3.2 µs, excitation lasers: 488 nm (16% emission power) and 633 nm (26% emission power), detector gain: 696.

- *NIH 3T3*

NIH 3T3 cell lines were plated on uncoated 96-well plates (seeding density 7*10^3 cell/well, surface area 0.32 cm^2) and cultured for 24 h at 37 °C with 5% CO_2 in 100 µL DMEM supplemented with 10% FBS and P-S antibiotics (Penicillin G Sodium Salt: 100 µg/mL, Streptomycin Sulfate: 100 µg/mL). After 24h, the medium was replaced with DMEM/FBS (10%) (without antibiotics) and the cells were transfected with pcDNA3-mEOS3.2-HP1α plasmid (0.4 µg/well) and Lipofectamine2000 (1.6 µL/well) in Opti-MEM Reduced Serum Medium. After 4h the Lipofectamine2000-plasmid complex was removed and cells incubated in 100 µL DMEM/FBS (10%)/P-S for 36h. Cells were washed with PBS (supplemented with $MgCl_2$ and $CaCl_2$) and incubated for 1h with 5 µM peptide probe in PBS (supplemented with $MgCl_2$ and $CaCl_2$). Images were acquired on a confocal microscope LSM 510 Meta (Zeiss), using a C-Apochromat 63x/1.2 water corrected objective. Imaging was performed using the following parameters: pixel time: 1.28 µs, excitation lasers: 488 nm (13% emission power) and 633 nm (25% emission power), detector gain: 550.

Image analysis was performed with Fiji software. Colocalization analysis of mEos3.2-HP1α and probe-SiR signals was performed on single cells. First, a nuclear mask was obtained by smoothing the mEos3.2-HP1α channel signal with a Median filter and applying the IsoData thresholding algorithm, followed by "Fill holes" command. The nuclear mask was applied to SiR-probe channel to exclude any SiR signal outside of the nucleus from the analysis. Background correction was performed in the mEos3.2-HP1α and probe-SiR channels setting a rolling ball radius of 20 and 50 pixels, respectively. Pearson's correlation coefficients were calculated with the BIOP JACOBS plugin, using the nuclear mask as a ROI, to exclude pixels outside of the nucleus to be considered in the analysis.

6.2.17 <u>Electrophoretic mobility shift assay</u>

170 bp 601 DNA (10 or 20 nmol) was incubated with increasing concentration of peptide probe (50-1600 nM) in 20 μL TEK buffer (10 mM Tris, 0.1 mM EDTA, 100 mM KCl, pH 7.5) and incubated 15 min at RT. 10 μL were mixed with 2μL 25% (w/v) sucrose and analyzed by Native-PAGE (200 V, 30 min, 4 °C).

6.2.18 <u>*In vitro* HP1α phase separation and confocal microscopy</u>

Phosphorylated HP1α frozen aliquot (in 50 mM HEPES, 150 mM NaCl, 1 mM DTT, 30% (w/v) glycerol, pH 7.3, 5 μL) was defrosted on ice and mixed with the probe of interest to a final concentration of 60 or 70 μM phHP1α and 10 μM probe. The mixture was incubated 5 min at 4 °C and imaged with a confocal microscope. Movies (10 frames, 0.2 fps) were acquired on a confocal microscope LSM 510 Meta (Zeiss), using a Plan Apochromat 100x/1.45 oil corrected objective. Imaging was performed using the following parameters: pixel time: 2.56 μs, excitation lasers: 633 nm (15% emission power), detector gain: 534.

Image analysis: calculation of the diameter and number of phHP1α-probe droplets was performed with Fiji software. A mask of the SiR signal was obtained by generating a binary image with Intermodes thresholding algorithm and the total area was calculated. Watershed segmentation was performed to separate droplets in proximity. Find Maxima algorithm was then applied to get the number of droplets. Average droplet size was then obtained by dividing the total area by the number of droplets.

Chapter 8: References

(1)   Holde, K. E. van. *Chromatin*; Springer Science & Business Media, **2012**.

(2)   Luger, K. Crystal Structure of the Nucleosome Core Particle at 2.8 A° Resolution. **1997**, *389*, 10.

(3)   Luger, K.; Rechsteiner, T. J.; Flaus, A. J.; Waye, M. M. Y.; Richmond, T. J. Characterization of Nucleosome Core Particles Containing Histone Proteins Made in Bacteria11Edited by A. Klug. *J. Mol. Biol.* **1997**, *272* (3), 301–311. https://doi.org/10.1006/jmbi.1997.1235.

(4)   Li, G.; Levitus, M.; Bustamante, C.; Widom, J. Rapid Spontaneous Accessibility of Nucleosomal DNA. *Nat. Struct. Mol. Biol.* **2005**, *12* (1), 46–53. https://doi.org/10.1038/nsmb869.

(5)   Koopmans, W. J. A.; Brehm, A.; Logie, C.; Schmidt, T.; van Noort, J. Single-Pair FRET Microscopy Reveals Mononucleosome Dynamics. *J. Fluoresc.* **2007**, *17* (6), 785–795. https://doi.org/10.1007/s10895-007-0218-9.

(6)   Böhm, V.; Hieb, A. R.; Andrews, A. J.; Gansen, A.; Rocker, A.; Tóth, K.; Luger, K.; Langowski, J. Nucleosome Accessibility Governed by the Dimer/Tetramer Interface. *Nucleic Acids Res.* **2011**, *39* (8), 3093–3102. https://doi.org/10.1093/nar/gkq1279.

(7)   Ngo, T. T. M.; Ha, T. Nucleosomes Undergo Slow Spontaneous Gaping. *Nucleic Acids Res.* **2015**, *43* (8), 3964–3971. https://doi.org/10.1093/nar/gkv276.

(8)   Hall, M. A.; Shundrovsky, A.; Bai, L.; Fulbright, R. M.; Lis, J. T.; Wang, M. D. High-Resolution Dynamic Mapping of Histone-DNA Interactions in a Nucleosome. *Nat. Struct. Mol. Biol.* **2009**, *16* (2), 124–129. https://doi.org/10.1038/nsmb.1526.

(9)   Park, Y.-J.; Dyer, P. N.; Tremethick, D. J.; Luger, K. A New Fluorescence Resonance Energy Transfer Approach Demonstrates That the Histone Variant H2AZ Stabilizes the Histone Octamer within the Nucleosome. *J. Biol. Chem.* **2004**, *279* (23), 24274–24282. https://doi.org/10.1074/jbc.M313152200.

(10)   Widom, J. Role of DNA Sequence in Nucleosome Stability and Dynamics. *Q. Rev. Biophys.* **2001**, *34* (3), 269–324. https://doi.org/10.1017/S0033583501003699.

(11)   Li, M.; Hada, A.; Sen, P.; Olufemi, L.; Hall, M. A.; Smith, B. Y.; Forth, S.; McKnight, J. N.; Patel, A.; Bowman, G. D.; et al. Dynamic Regulation of Transcription Factors by Nucleosome Remodeling. *eLife* **2015**, *4*, e06249. https://doi.org/10.7554/eLife.06249.

(12)   Olins, A. L.; Olins, D. E. Spheroid Chromatin Units (ν Bodies). *Science* **1974**, *183* (4122), 330–332.

(13)   Woodcock, C. L.; Safer, J. P.; Stanchfield, J. E. Structural Repeating Units in Chromatin. I. Evidence for Their General Occurrence. *Exp. Cell Res.* **1976**, *97*, 101–110. https://doi.org/10.1016/0014-4827(76)90659-5.

(14)   Thoma, F. Involvement of Histone H1 in the Organization of the Nucleosome and of the Salt-Dependent Superstructures of Chromatin. *J. Cell Biol.* **1979**, *83* (2), 403–427. https://doi.org/10.1083/jcb.83.2.403.

(15)   Dorigo, B.; Schalch, T.; Kulangara, A.; Duda, S.; Schroeder, R. R.; Richmond, T. J. Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science* **2004**, *306* (5701), 1571–1573. https://doi.org/10.1126/science.1103124.

(16)   Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. X-Ray Structure of a Tetranucleosome and Its Implications for the Chromatin Fibre. *Nature* **2005**, *436* (7047), 138–141. https://doi.org/10.1038/nature03686.

(17)   Song, F.; Chen, P.; Sun, D.; Wang, M.; Dong, L.; Liang, D.; Xu, R.-M.; Zhu, P.; Li, G. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science* **2014**, *344* (6182), 376–380. https://doi.org/10.1126/science.1251413.

(18)   Finch, J. T.; Klug, A. Solenoidal Model for Superstructure in Chromatin. *Proc. Natl. Acad. Sci.* **1976**, *73* (6), 1897–1901. https://doi.org/10.1073/pnas.73.6.1897.

(19)   Routh, A.; Sandin, S.; Rhodes, D. Nucleosome Repeat Length and Linker Histone Stoichiometry Determine Chromatin Fiber Structure. *Proc. Natl. Acad. Sci.* **2008**, *105* (26), 8872–8877. https://doi.org/10.1073/pnas.0802336105.

(20)   Grigoryev, S. A.; Bascom, G.; Buckwalter, J. M.; Schubert, M. B.; Woodcock, C. L.; Schlick, T. Hierarchical Looping of Zigzag Nucleosome Chains in Metaphase Chromosomes. *Proc. Natl. Acad. Sci.* **2016**, *113* (5), 1238–1243. https://doi.org/10.1073/pnas.1518280113.

(21)   Hsieh, T.-H. S.; Weiner, A.; Lajoie, B.; Dekker, J.; Friedman, N.; Rando, O. J. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **2015**, *162* (1), 108–119. https://doi.org/10.1016/j.cell.2015.05.048.

(22)   Hsieh, T.-H. S.; Fudenberg, G.; Goloborodko, A.; Rando, O. J. Micro-C XL: Assaying Chromosome Conformation from the Nucleosome to the Entire Genome. *Nat. Methods* **2016**, *13* (12), 1009–1011. https://doi.org/10.1038/nmeth.4025.

(23)   Risca, V. I.; Denny, S. K.; Straight, A. F.; Greenleaf, W. J. Variable Chromatin Structure Revealed by in Situ Spatially Correlated DNA Cleavage Mapping. *Nature* **2017**, *541* (7636), 237–241. https://doi.org/10.1038/nature20781.

(24)   Ou, H. D.; Phan, S.; Deerinck, T. J.; Thor, A.; Ellisman, M. H.; O'Shea, C. C. ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells. *Science* **2017**, *357* (6349). https://doi.org/10.1126/science.aag0025.

(25)   Ricci, M. A.; Manzo, C.; García-Parajo, M. F.; Lakadamyali, M.; Cosma, M. P. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **2015**, *160* (6), 1145–1158. https://doi.org/10.1016/j.cell.2015.01.054.

(26)   Lieberman-Aiden, E.; Berkum, N. L. van; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B. R.; Sabo, P. J.; Dorschner, M. O.; et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **2009**, *326* (5950), 289–293. https://doi.org/10.1126/science.1181369.

(27)   Szabo, Q.; Bantignies, F.; Cavalli, G. Principles of Genome Folding into Topologically Associating Domains. *Sci. Adv.* **2019**, *5* (4), eaaw1668. https://doi.org/10.1126/sciadv.aaw1668.

(28)   Nuebler, J.; Fudenberg, G.; Imakaev, M.; Abdennur, N.; Mirny, L. A. Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation. *Proc. Natl. Acad. Sci.* **2018**, *115* (29), E6697–E6706. https://doi.org/10.1073/pnas.1717730115.

(29)   Wang, S.; Su, J.-H.; Beliveau, B. J.; Bintu, B.; Moffitt, J. R.; Wu, C.; Zhuang, X. Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes. *Science* **2016**, *353* (6299), 598–602. https://doi.org/10.1126/science.aaf8084.

(30)   Greenberg, M. V. C.; Bourc'his, D. The Diverse Roles of DNA Methylation in Mammalian Development and Disease. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (10), 590–607. https://doi.org/10.1038/s41580-019-0159-6.

(31)   Zaratiegui, M.; Irvine, D. V.; Martienssen, R. A. Noncoding RNAs and Gene Silencing. *Cell* **2007**, *128* (4), 763–776. https://doi.org/10.1016/j.cell.2007.02.016.

(32)   Bannister, A. J.; Kouzarides, T. Regulation of Chromatin by Histone Modifications. *Cell Res.* **2011**, *21* (3), 381–395. https://doi.org/10.1038/cr.2011.22.

(33)   Griffith, J. S.; Mahler, H. R. DNA Ticketing Theory of Memory. *Nature* **1969**, *223* (5206), 580–582. https://doi.org/10.1038/223580a0.

(34)   Lister, R.; Pelizzola, M.; Dowen, R. H.; Hawkins, R. D.; Hon, G.; Tonti-Filippini, J.; Nery, J. R.; Lee, L.; Ye, Z.; Ngo, Q.-M.; et al. Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* **2009**, *462* (7271), 315–322. https://doi.org/10.1038/nature08514.

(35)   Razin, A.; Cedar, H. DNA Methylation and Genomic Imprinting. *Cell* **1994**, *77* (4), 473–476. https://doi.org/10.1016/0092-8674(94)90208-9.

(36)   Li, E.; Beard, C.; Jaenisch, R. Role for DNA Methylation in Genomic Imprinting. *Nature* **1993**, *366* (6453), 362–365. https://doi.org/10.1038/366362a0.

(37)   Auclair, G.; Weber, M. Mechanisms of DNA Methylation and Demethylation in Mammals. *Biochimie* **2012**, *94* (11), 2202–2211. https://doi.org/10.1016/j.biochi.2012.05.016.

(38)   Kohli, R. M.; Zhang, Y. TET Enzymes, TDG and the Dynamics of DNA Demethylation. *Nature* **2013**, *502* (7472), 472–479. https://doi.org/10.1038/nature12750.

(39)   Britten, R. J.; Davidson, E. H. Gene Regulation for Higher Cells: A Theory. *Science* **1969**, *165* (3891), 349–357. https://doi.org/10.1126/science.165.3891.349.

(40)   Moazed, D. Small RNAs in Transcriptional Gene Silencing and Genome Defence. *Nature* **2009**, *457* (7228), 413–420. https://doi.org/10.1038/nature07756.

(41)   Guttman, M. Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals. **2009**, *458*, 5.

(42)   Whitehead, J.; Pandey, G. K.; Kanduri, C. Regulation of the Mammalian Epigenome by Long Noncoding RNAs. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2009**, *1790* (9), 936–947. https://doi.org/10.1016/j.bbagen.2008.10.007.

(43)   Bernstein, E.; Allis, C. D. RNA Meets Chromatin. *Genes Dev.* **2005**, *19* (14), 1635–1655. https://doi.org/10.1101/gad.1324305.

(44)   Huang, H.; Lin, S.; Garcia, B. A.; Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* **2015**, *115* (6), 2376–2418. https://doi.org/10.1021/cr500491u.

(45)   Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modifications. **2000**, *403*, 5.

(46)   Jenuwein, T.; Allis, C. D. Translating the Histone Code. *Science* **2001**, *293* (5532), 1074–1080. https://doi.org/10.1126/science.1063127.

(47)   Xhemalce, B.; Dawson, M. A.; Bannister, A. J. Histone Modifications. In *Reviews in Cell Biology and Molecular Medicine*; American Cancer Society, 2011. https://doi.org/10.1002/3527600906.mcb.201100004.

(48)   Sims, R. J.; Nishioka, K.; Reinberg, D. Histone Lysine Methylation: A Signature for Chromatin Function. *Trends Genet.* **2003**, *19* (11), 629–639. https://doi.org/10.1016/j.tig.2003.09.007.

(49)   Simon, M.; North, J. A.; Shimko, J. C.; Forties, R. A.; Ferdinand, M. B.; Manohar, M.; Zhang, M.; Fishel, R.; Ottesen, J. J.; Poirier, M. G. Histone Fold Modifications Control Nucleosome Unwrapping and Disassembly. *Proc. Natl. Acad. Sci.* **2011**, *108* (31), 12711–12716. https://doi.org/10.1073/pnas.1106264108.

(50)   Flanagan, J. F.; Mi, L.-Z.; Chruszcz, M.; Cymborowski, M.; Clines, K. L.; Kim, Y.; Minor, W.; Rastinejad, F.; Khorasanizadeh, S. Double Chromodomains Cooperate to Recognize the Methylated Histone H3 Tail. *Nature* **2005**, *438* (7071), 1181–1185. https://doi.org/10.1038/nature04290.

(51)   Chen, G.; Li, W.; Yan, F.; Wang, D.; Chen, Y. The Structural Basis for Specific Recognition of H3K14 Acetylation by Sth1 in the RSC Chromatin Remodeling Complex. *Structure* **2019**, S0969212619303569. https://doi.org/10.1016/j.str.2019.10.015.

(52)   Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T. Methylation of Histone H3 Lysine 9 Creates a Binding Site for HP1 Proteins. *Nature* **2001**, *410* (6824), 116–120. https://doi.org/10.1038/35065132.

(53)   Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T. Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain. *Nature* **2001**, *410* (6824), 120–124. https://doi.org/10.1038/35065138.

(54)   Laugesen, A.; Højfeldt, J. W.; Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* **2019**, *74* (1), 8–18. https://doi.org/10.1016/j.molcel.2019.03.011.

(55)   Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M.-M. Structure and Ligand of a Histone Acetyltransferase Bromodomain. **1999**, *399*, 6.

(56)   Paro, R.; Hogness, D. S. The Polycomb Protein Shares a Homologous Domain with a Heterochromatin-Associated Protein of Drosophila. *Proc. Natl. Acad. Sci.* **1991**, *88* (1), 263–267. https://doi.org/10.1073/pnas.88.1.263.

(57)   Sanchez, R.; Zhou, M.-M. The PHD Finger: A Versatile Epigenome Reader. *Trends Biochem. Sci.* **2011**, *36* (7), 364–372. https://doi.org/10.1016/j.tibs.2011.03.005.

(58)   Li, L.; Zhang, H.; Zhang, M.; Zhao, M.; Feng, L.; Luo, X.; Gao, Z.; Huang, Y.; Ardayfio, O.; Zhang, J.-H.; et al. Discovery and Molecular Basis of a Diverse Set of Polycomb Repressive Complex 2 Inhibitors Recognition by EED. *PLOS ONE* **2017**, *12* (1), e0169855. https://doi.org/10.1371/journal.pone.0169855.

(59)   Bonasio, R.; Lecona, E.; Reinberg, D. MBT Domain Proteins in Development and Disease. *Semin. Cell Dev. Biol.* **2010**, *21* (2), 221–230. https://doi.org/10.1016/j.semcdb.2009.09.010.

(60)   Lu, R.; Wang, G. G. Tudor: A Versatile Family of Histone Methylation 'Readers.' *Trends Biochem. Sci.* **2013**, *38* (11), 546–555. https://doi.org/10.1016/j.tibs.2013.08.002.

(61)   Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. Recognition of Trimethyllysine by a Chromodomain Is Not Driven by the Hydrophobic Effect. *Proc. Natl. Acad. Sci.* **2007**, *104* (27), 11184–11188. https://doi.org/10.1073/pnas.0610850104.

(62)   Ma, J. C.; Dougherty, D. A. The Cation–π Interaction. *Chem. Rev.* **1997**, *97* (5), 1303–1324. https://doi.org/10.1021/cr9603744.

(63)   Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. How Chromatin-Binding Modules Interpret Histone Modifications: Lessons from Professional Pocket Pickers. *Nat. Struct. Mol. Biol.* **2007**, *14* (11), 1025–1040. https://doi.org/10.1038/nsmb1338.

(64)   Li, H.; Fischle, W.; Wang, W.; Duncan, E. M.; Liang, L.; Murakami-Ishibe, S.; Allis, C. D.; Patel, D. J. Structural Basis for Lower Lysine Methylation State-Specific Readout by MBT Repeats of L3MBTL1 and an Engineered PHD Finger. *Mol. Cell* **2007**, *28* (4), 677–691. https://doi.org/10.1016/j.molcel.2007.10.023.

(65)   Botuyan, M. V.; Lee, J.; Ward, I. M.; Kim, J.-E.; Thompson, J. R.; Chen, J.; Mer, G. Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell* **2006**, *127* (7), 1361–1373. https://doi.org/10.1016/j.cell.2006.10.043.

(66)   Jacobs, S. A. Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science* **2002**, *295* (5562), 2080–2083. https://doi.org/10.1126/science.1069473.

(67)   Huang, Y.; Fang, J.; Bedford, M. T.; Zhang, Y.; Xu, R.-M. Recognition of Histone H3 Lysine-4 Methylation by the Double Tudor Domain of JMJD2A. *Science* **2006**, *312* (5774), 748–751. https://doi.org/10.1126/science.1125162.

(68)   Peña, P. V.; Davrazou, F.; Shi, X.; Walter, K. L.; Verkhusha, V. V.; Gozani, O.; Zhao, R.; Kutateladze, T. G. Molecular Mechanism of Histone H3K4me3 Recognition by Plant Homeodomain of ING2. *Nature* **2006**, *442* (7098), 100–103. https://doi.org/10.1038/nature04814.

(69)   Li, H.; Ilin, S.; Wang, W.; Duncan, E. M.; Wysocka, J.; Allis, C. D.; Patel, D. J. Molecular Basis for Site-Specific Read-out of Histone H3K4me3 by the BPTF PHD Finger of NURF. *Nature* **2006**, *442* (7098), 91–95. https://doi.org/10.1038/nature04802.

(70)   Trievel, R. C.; Beach, B. M.; Dirk, L. M. A.; Houtz, R. L.; Hurley, J. H. Structure and Catalytic Mechanism of a SET Domain Protein Methyltransferase. *Cell* **2002**, *111* (1), 91–103. https://doi.org/10.1016/S0092-8674(02)01000-0.

(71)   Rea, S.; Eisenhaber, F.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. Regulation of Chromatin Structure by Site-Speci®c Histone H3 Methyltransferases. **2000**, *406*, 7.

(72)   Peters, A. H.; O'Carroll, D.; Scherthan, H.; Mechtler, K.; Sauer, S.; Schöfer, C.; Weipoltshammer, K.; Pagani, M.; Lachner, M.; Kohlmaier, A.; et al. Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* **2001**, *107* (3), 323–337. https://doi.org/10.1016/s0092-8674(01)00542-6.

(73)   Cao, R.; Wang, L.; Wang, H.; Xia, L.; Erdjument-Bromage, H.; Tempst, P.; Jones, R. S.; Zhang, Y. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **2002**, *298* (5595), 1039–1043. https://doi.org/10.1126/science.1076997.

(74)   Nishioka, K.; Rice, J. C.; Sarma, K.; Erdjument-Bromage, H.; Werner, J.; Wang, Y.; Chuikov, S.; Valenzuela, P.; Tempst, P.; Steward, R.; et al. PR-Set7 Is a Nucleosome-Specific Methyltransferase That Modifies Lysine 20 of Histone H4 and Is Associated with Silent Chromatin. *Mol. Cell* **2002**, *9* (6), 1201–1213. https://doi.org/10.1016/S1097-2765(02)00548-8.

(75)   Edmunds, J. W.; Mahadevan, L. C.; Clayton, A. L. Dynamic Histone H3 Methylation during Gene Induction: HYPB/Setd2 Mediates All H3K36 Trimethylation. *EMBO J.* **2008**, *27* (2), 406–420. https://doi.org/10.1038/sj.emboj.7601967.

(76)   Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004**, *119* (7), 941–953. https://doi.org/10.1016/j.cell.2004.12.012.

(77)   Hodawadekar, S. C.; Marmorstein, R. Chemistry of Acetyl Transfer by Histone Modifying Enzymes: Structure, Mechanism and Implications for Effector Design. *Oncogene* **2007**, *26* (37), 5528–5540. https://doi.org/10.1038/sj.onc.1210619.

(78)   Kuo, M.-H.; Brownell, J. E.; Sobel, R. E.; Ranalli, T. A.; Cook, R. G.; Edmondson, D. G.; Roth, S. Y. Transcription-Linked Acetylation by GcnSp of Histones H3 and H4 at Specific Lysines. **1996**, *383*, 4.

(79)   Chen, H. P.; Zhao, Y. T.; Zhao, T. C. Histone Deacetylases and Mechanisms of Regulation of Gene Expression (Histone Deacetylases in Cancer). **2016**, 17.

(80)   Downs, J. A.; Lowndes, N. F.; Jackson, S. P. A Role for Saccharomyces Cerevisiae Histone H2A in DNA Repair. *Nature* **2000**, *408* (6815), 1001–1004. https://doi.org/10.1038/35050000.

(81)   Lau, P. N. I.; Cheung, P. Histone Code Pathway Involving H3 S28 Phosphorylation and K27 Acetylation Activates Transcription and Antagonizes Polycomb Silencing. *Proc. Natl. Acad. Sci.* **2011**, *108* (7), 2801–2806. https://doi.org/10.1073/pnas.1012798108.

(82)   Choi, H. S.; Choi, B. Y.; Cho, Y.-Y.; Mizuno, H.; Kang, B. S.; Bode, A. M.; Dong, Z. Phosphorylation of Histone H3 at Serine 10 Is Indispensable for Neoplastic Cell Transformation. *Cancer Res.* **2005**, *65* (13), 5818–5827. https://doi.org/10.1158/0008-5472.CAN-05-0197.

(83)   Sauvé, D. M.; Anderson, H. J.; Ray, J. M.; James, W. M.; Roberge, M. Phosphorylation-Induced Rearrangement of the Histone H3 NH$_2$-Terminal Domain during Mitotic Chromosome Condensation. *J. Cell Biol.* **1999**, *145* (2), 225–235. https://doi.org/10.1083/jcb.145.2.225.

(84)   Ruthenburg, A. J.; Li, H.; Patel, D. J.; Allis, C. D. Multivalent Engagement of Chromatin Modifications by Linked Binding Modules. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (12), 983–994. https://doi.org/10.1038/nrm2298.

(85)   Maison, C.; Almouzni, G. HP1 and the Dynamics of Heterochromatin Maintenance. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (4), 296–305. https://doi.org/10.1038/nrm1355.

(86) Panier, S.; Boulton, S. J. Double-Strand Break Repair: 53BP1 Comes into Focus. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (1), 7–18. https://doi.org/10.1038/nrm3719.

(87) Schuettengruber, B.; Bourbon, H.-M.; Di Croce, L.; Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **2017**, *171* (1), 34–57. https://doi.org/10.1016/j.cell.2017.08.002.

(88) Simon, J. A.; Kingston, R. E. Mechanisms of Polycomb Gene Silencing: Knowns and Unknowns. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (10), 697–708. https://doi.org/10.1038/nrm2763.

(89) Park, P. J. ChIP–Seq: Advantages and Challenges of a Maturing Technology. *Nat. Rev. Genet.* **2009**, *10* (10), 669–680. https://doi.org/10.1038/nrg2641.

(90) Grzybowski, A. T.; Chen, Z.; Ruthenburg, A. J. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Mol. Cell* **2015**, *58* (5), 886–899. https://doi.org/10.1016/j.molcel.2015.04.022.

(91) Martello, G.; Smith, A. The Nature of Embryonic Stem Cells. *Annu. Rev. Cell Dev. Biol.* **2014**, *30* (1), 647–675. https://doi.org/10.1146/annurev-cellbio-100913-013116.

(92) Bernstein, B. E.; Mikkelsen, T. S.; Xie, X.; Kamal, M.; Huebert, D. J.; Cuff, J.; Fry, B.; Meissner, A.; Wernig, M.; Plath, K.; et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **2006**, *125* (2), 315–326. https://doi.org/10.1016/j.cell.2006.02.041.

(93) Azuara, V.; Perry, P.; Sauer, S.; Spivakov, M.; Jørgensen, H. F.; John, R. M.; Gouti, M.; Casanova, M.; Warnes, G.; Merkenschlager, M.; et al. Chromatin Signatures of Pluripotent Cell Lines. *Nat. Cell Biol.* **2006**, *8* (5), 532–538. https://doi.org/10.1038/ncb1403.

(94) Zhao, X. D.; Han, X.; Chew, J. L.; Liu, J.; Chiu, K. P.; Choo, A.; Orlov, Y. L.; Sung, W.-K.; Shahab, A.; Kuznetsov, V. A.; et al. Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. *Cell Stem Cell* **2007**, *1* (3), 286–298. https://doi.org/10.1016/j.stem.2007.08.004.

(95) Pan, G.; Tian, S.; Nie, J.; Yang, C.; Ruotti, V.; Wei, H.; Jonsdottir, G. A.; Stewart, R.; Thomson, J. A. Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell* **2007**, *1* (3), 299–312. https://doi.org/10.1016/j.stem.2007.08.003.

(96) Dou, Y.; Milne, T. A.; Tackett, A. J.; Smith, E. R.; Fukuda, A.; Wysocka, J.; Allis, C. D.; Chait, B. T.; Hess, J. L.; Roeder, R. G. Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. *Cell* **2005**, *121* (6), 873–885. https://doi.org/10.1016/j.cell.2005.04.031.

(97) Delachat, A. M.-F.; Guidotti, N.; Bachmann, A. L.; Meireles-Filho, A. C. A.; Pick, H.; Lechner, C. C.; Deluz, C.; Deplancke, B.; Suter, D. M.; Fierz, B. Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. *Cell Chem. Biol.* **2018**, *25* (1), 51-56.e6. https://doi.org/10.1016/j.chembiol.2017.10.008.

(98) Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A. A Variant of Yellow Fluorescent Protein with Fast and Efficient Maturation for Cell-Biological Applications. *Nat. Biotechnol.* **2002**, *20* (1), 87–90. https://doi.org/10.1038/nbt0102-87.

(99) Fischle, W.; Wang, Y.; Jacobs, S. A.; Kim, Y.; Allis, C. D.; Khorasanizadeh, S. Molecular Basis for the Discrimination of Repressive Methyl-Lysine Marks in Histone H3 by Polycomb and HP1 Chromodomains. *Genes Dev.* **2003**, *17* (15), 1870–1881. https://doi.org/10.1101/gad.1110503.

(100) Vermeulen, M.; Mulder, K. W.; Denissov, S.; Pijnappel, W. W. M. P.; van Schaik, F. M. A.; Varier, R. A.; Baltissen, M. P. A.; Stunnenberg, H. G.; Mann, M.; Timmers, H. Th. M. Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **2007**, *131* (1), 58–69. https://doi.org/10.1016/j.cell.2007.08.016.

(101) Konze, K. D.; Ma, A.; Li, F.; Barsyte-Lovejoy, D.; Parton, T.; MacNevin, C. J.; Liu, F.; Gao, C.; Huang, X.-P.; Kuznetsova, E.; et al. An Orally Bioavailable Chemical Probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* **2013**, *8* (6), 1324–1334. https://doi.org/10.1021/cb400133j.

(102) Voigt, P.; LeRoy, G.; Drury, W. J.; Zee, B. M.; Son, J.; Beck, D.; Young, N. L.; Garcia, B. A.; Reinberg, D. Asymmetrically Modified Nucleosomes. *Cell* **2012**, *151* (1), 181–193. https://doi.org/10.1016/j.cell.2012.09.002.

(103) Shema, E.; Jones, D.; Shoresh, N.; Donohue, L.; Ram, O.; Bernstein, B. E. Single-Molecule Decoding of Combinatorially Modified Nucleosomes. *Science* **2016**, *352* (6286), 717–721. https://doi.org/10.1126/science.aad7701.

(104) Lechner, C. C.; Agashe, N. D.; Fierz, B. Traceless Synthesis of Asymmetrically Modified Bivalent Nucleosomes. *Angew. Chem. Int. Ed.* **2016**, *55* (8), 2903–2906. https://doi.org/10.1002/anie.201510996.

(105) Liokatis, S.; Klingberg, R.; Tan, S.; Schwarzer, D. Differentially Isotope-Labeled Nucleosomes To Study Asymmetric Histone Modification Crosstalk by Time-Resolved NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2016**, *55* (29), 8262–8265. https://doi.org/10.1002/anie.201601938.

(106) Dai, J.; Sultan, S.; Taylor, S. S.; Higgins, J. M. G. The Kinase Haspin Is Required for Mitotic Histone H3 Thr 3 Phosphorylation and Normal Metaphase Chromosome Alignment. *Genes Dev.* **2005**, *19* (4), 472–488. https://doi.org/10.1101/gad.1267105.

(107) Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **2007**, *129* (4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009.

(108) Ichikawa, Y.; Connelly, C. F.; Appleboim, A.; Miller, T. C.; Jacobi, H.; Abshiru, N. A.; Chou, H.-J.; Chen, Y.; Sharma, U.; Zheng, Y.; et al. A Synthetic Biology Approach to Probing Nucleosome Symmetry. *eLife* **2017**, *6*, e28836. https://doi.org/10.7554/eLife.28836.

(109) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. A New Functional Suppressor TRNA/Aminoacyl–tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *J. Am. Chem. Soc.* **2000**, *122* (20), 5010–5011. https://doi.org/10.1021/ja000595y.

(110) Rhee, H. S.; Bataille, A. R.; Zhang, L.; Pugh, B. F. Subnucleosomal Structures and Nucleosome Asymmetry across a Genome. *Cell* **2014**, *159* (6), 1377–1388. https://doi.org/10.1016/j.cell.2014.10.054.

(111) Nott, T. J.; Craggs, T. D.; Baldwin, A. J. Membraneless Organelles Can Melt Nucleic Acid Duplexes and Act as Biomolecular Filters. *Nat. Chem.* **2016**, *8* (6), 569–575. https://doi.org/10.1038/nchem.2519.

(112) Strouboulis, J.; Wolffe, A. P. Functional Compartmentalization of the Nucleus. *J. Cell Sci.* **1996**, *109* (8), 1991–2000.

(113) Friedman, J. R.; Nunnari, J. Mitochondrial Form and Function. *Nature* **2014**, *505* (7483), 335–343. https://doi.org/10.1038/nature12985.

(114) Mao, Y. S.; Zhang, B.; Spector, D. L. Biogenesis and Function of Nuclear Bodies. *Trends Genet.* **2011**, *27* (8), 295–306. https://doi.org/10.1016/j.tig.2011.05.006.

(115) Boisvert, F.-M.; van Koningsbruggen, S.; Navascués, J.; Lamond, A. I. The Multifunctional Nucleolus. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (7), 574–585. https://doi.org/10.1038/nrm2184.

(116) Mahen, R.; Venkitaraman, A. R. Pattern Formation in Centrosome Assembly. *Curr. Opin. Cell Biol.* **2012**, *24* (1), 14–23. https://doi.org/10.1016/j.ceb.2011.12.012.

(117) Buchan, J. R. MRNP Granules: Assembly, Function, and Connections with Disease. *RNA Biol.* **2014**, *11* (8), 1019–1030. https://doi.org/10.4161/15476286.2014.972208.

(118) Lallemand-Breitenbach, V.; Thé, H. de. PML Nuclear Bodies. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (5), a000661. https://doi.org/10.1101/cshperspect.a000661.

(119) Cioce, M.; Lamond, A. I. CAJAL BODIES: A Long History of Discovery. *Annu. Rev. Cell Dev. Biol.* **2005**, *21* (1), 105–131. https://doi.org/10.1146/annurev.cellbio.20.010403.103738.

(120) Lamond, A. I.; Spector, D. L. Nuclear Speckles: A Model for Nuclear Organelles. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (8), 605–612. https://doi.org/10.1038/nrm1172.

(121) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Julicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324* (5935), 1729–1732. https://doi.org/10.1126/science.1172046.

(122) Brangwynne, C. P.; Mitchison, T. J.; Hyman, A. A. Active Liquid-like Behavior of Nucleoli Determines Their Size and Shape in Xenopus Laevis Oocytes. *Proc. Natl. Acad. Sci.* **2011**, *108* (11), 4334–4339. https://doi.org/10.1073/pnas.1017150108.

(123) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **2014**, *30* (1), 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325.

(124) Broide, M. L.; Berland, C. R.; Pande, J.; Ogun, O. O.; Benedek, G. B. Binary-Liquid Phase Separation of Lens Protein Solutions. *Proc. Natl. Acad. Sci.* **1991**, *88* (13), 5660–5664. https://doi.org/10.1073/pnas.88.13.5660.

(125) Galkin, O.; Chen, K.; Nagel, R. L.; Hirsch, R. E.; Vekilov, P. G. Liquid-Liquid Separation in Solutions of Normal and Sickle Cell Hemoglobin. *Proc. Natl. Acad. Sci.* **2002**, *99* (13), 8479–8483. https://doi.org/10.1073/pnas.122055299.

(126) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; et al. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483* (7389), 336–340. https://doi.org/10.1038/nature10879.

(127)   Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **2015**, *163* (1), 123–133. https://doi.org/10.1016/j.cell.2015.09.015.

(128)   Saito, M.; Hess, D.; Eglinger, J.; Fritsch, A. W.; Kreysing, M.; Weinert, B. T.; Choudhary, C.; Matthias, P. Acetylation of Intrinsically Disordered Regions Regulates Phase Separation. *Nat. Chem. Biol.* **2019**, *15* (1), 51–61. https://doi.org/10.1038/s41589-018-0180-7.

(129)   Ferreon, J. C.; Jain, A.; Choi, K.-J.; Tsoi, P. S.; MacKenzie, K. R.; Jung, S. Y.; Ferreon, A. C. Acetylation Disfavors Tau Phase Separation. *Int. J. Mol. Sci.* **2018**, *19* (5). https://doi.org/10.3390/ijms19051360.

(130)   Monahan, Z.; Ryan, V. H.; Janke, A. M.; Burke, K. A.; Rhoads, S. N.; Zerze, G. H.; O'Meally, R.; Dignon, G. L.; Conicella, A. E.; Zheng, W.; et al. Phosphorylation of the FUS Low-Complexity Domain Disrupts Phase Separation, Aggregation, and Toxicity. *EMBO J.* **2017**, *36* (20), 2951–2967. https://doi.org/10.15252/embj.201696394.

(131)   Franzmann, T. M.; Jahnel, M.; Pozniakovsky, A.; Mahamid, J.; Holehouse, A. S.; Nüske, E.; Richter, D.; Baumeister, W.; Grill, S. W.; Pappu, R. V.; et al. Phase Separation of a Yeast Prion Protein Promotes Cellular Fitness. *Science* **2018**, *359* (6371), eaao5654. https://doi.org/10.1126/science.aao5654.

(132)   Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **2015**, *57* (5), 936–947. https://doi.org/10.1016/j.molcel.2015.01.013.

(133)   Riback, J. A.; Katanski, C. D.; Kear-Scott, J. L.; Pilipenko, E. V.; Rojek, A. E.; Sosnick, T. R.; Drummond, D. A. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **2017**, *168* (6), 1028-1040.e19. https://doi.org/10.1016/j.cell.2017.02.027.

(134)   Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016**, *166* (3), 651–663. https://doi.org/10.1016/j.cell.2016.06.010.

(135)   Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating Sequence Encoded Information to Form and Function of Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112. https://doi.org/10.1016/j.sbi.2015.03.008.

(136)   Jensen, M. R.; Ruigrok, R. W.; Blackledge, M. Describing Intrinsically Disordered Proteins at Atomic Resolution by NMR. *Curr. Opin. Struct. Biol.* **2013**, *23* (3), 426–435. https://doi.org/10.1016/j.sbi.2013.02.007.

(137)   Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83* (1), 553–584. https://doi.org/10.1146/annurev-biochem-072711-164947.

(138)   Mészáros, B.; Erdős, G.; Szabó, B.; Schád, É.; Tantos, Á.; Abukhairan, R.; Horváth, T.; Murvai, N.; Kovács, O. P.; Kovács, M.; et al. PhaSePro: The Database of Proteins Driving Liquid–Liquid Phase Separation. *Nucleic Acids Res.* **2020**, *48* (D1), D360–D367. https://doi.org/10.1093/nar/gkz848.

(139)   Moazed, D. Mechanisms for the Inheritance of Chromatin States. *Cell* **2011**, *146* (4), 510–518. https://doi.org/10.1016/j.cell.2011.07.013.

(140)   Nakayama, J. -i. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* **2001**, *292* (5514), 110–113. https://doi.org/10.1126/science.1060118.

(141)   Elgin, S. C. R.; Reuter, G. Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (8), a017780. https://doi.org/10.1101/cshperspect.a017780.

(142)   Talbert, P. B.; Henikoff, S. Spreading of Silent Chromatin: Inaction at a Distance. *Nat. Rev. Genet.* **2006**, *7* (10), 793–803. https://doi.org/10.1038/nrg1920.

(143)   Zhang, K.; Mosch, K.; Fischle, W.; Grewal, S. I. S. Roles of the Clr4 Methyltransferase Complex in Nucleation, Spreading and Maintenance of Heterochromatin. *Nat. Struct. Mol. Biol.* **2008**, *15* (4), 381–388. https://doi.org/10.1038/nsmb.1406.

(144)   Müller, M. M.; Fierz, B.; Bittova, L.; Liszczak, G.; Muir, T. W. A Two-State Activation Mechanism Controls the Histone Methyltransferase Suv39h1. *Nat. Chem. Biol.* **2016**, *12* (3), 188–193. https://doi.org/10.1038/nchembio.2008.

(145)   Raurell-Vila, H.; Bosch-Presegue, L.; Gonzalez, J.; Kane-Goldsmith, N.; Casal, C.; Brown, J. P.; Marazuela-Duque, A.; Singh, P. B.; Serrano, L.; Vaquero, A. An HP1 Isoform-Specific Feedback Mechanism Regulates Suv39h1 Activity under Stress Conditions. *Epigenetics* **2017**, *12* (2), 166–175. https://doi.org/10.1080/15592294.2016.1278096.

(146) Blackwood, E. M. Going the Distance: A Current View of Enhancer Action. *Science* **1998**, *281* (5373), 60–63. https://doi.org/10.1126/science.281.5373.60.

(147) Maison, C.; Almouzni, G. HP1 and the Dynamics of Heterochromatin Maintenance. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (4), 296–305. https://doi.org/10.1038/nrm1355.

(148) Hiragami, K.; Festenstein, R. Heterochromatin Protein 1: A Pervasive Controlling Influence. *Cell. Mol. Life Sci.* **2005**, *62* (23), 2711–2726. https://doi.org/10.1007/s00018-005-5287-9.

(149) Machida, S.; Takizawa, Y.; Ishimaru, M.; Sugita, Y.; Sekine, S.; Nakayama, J.; Wolf, M.; Kurumizaka, H. Structural Basis of Heterochromatin Formation by Human HP1. *Mol. Cell* **2018**, *69* (3), 385-397.e8. https://doi.org/10.1016/j.molcel.2017.12.011.

(150) Smothers, J. F.; Henikoff, S. The HP1 Chromo Shadow Domain Binds a Consensus Peptide Pentamer. *Curr. Biol.* **2000**, *10* (1), 27–30. https://doi.org/10.1016/S0960-9822(99)00260-2.

(151) Musselman, C. A.; Lalonde, M.-E.; Côté, J.; Kutateladze, T. G. Perceiving the Epigenetic Landscape through Histone Readers. *Nat. Struct. Mol. Biol.* **2012**, *19* (12), 1218–1227. https://doi.org/10.1038/nsmb.2436.

(152) Meehan, R. R. HP1 Binding to Native Chromatin in Vitro Is Determined by the Hinge Region and Not by the Chromodomain. *EMBO J.* **2003**, *22* (12), 3164–3174. https://doi.org/10.1093/emboj/cdg306.

(153) Canzio, D.; Larson, A.; Narlikar, G. J. Mechanisms of Functional Promiscuity by HP1 Proteins. *Trends Cell Biol.* **2014**, *24* (6), 377–386. https://doi.org/10.1016/j.tcb.2014.01.002.

(154) Vakoc, C. R.; Mandat, S. A.; Olenchock, B. A.; Blobel, G. A. Histone H3 Lysine 9 Methylation and HP1γ Are Associated with Transcription Elongation through Mammalian Chromatin. *Mol. Cell* **2005**, *19* (3), 381–391. https://doi.org/10.1016/j.molcel.2005.06.011.

(155) Hiragami-Hamada, K.; Shinmyozu, K.; Hamada, D.; Tatsu, Y.; Uegaki, K.; Fujiwara, S.; Nakayama, J. -i. N-Terminal Phosphorylation of HP1 Promotes Its Chromatin Binding. *Mol. Cell. Biol.* **2011**, *31* (6), 1186–1200. https://doi.org/10.1128/MCB.01012-10.

(156) Nishibuchi, G.; Machida, S.; Osakabe, A.; Murakoshi, H.; Hiragami-Hamada, K.; Nakagawa, R.; Fischle, W.; Nishimura, Y.; Kurumizaka, H.; Tagami, H.; et al. N-Terminal Phosphorylation of HP1α Increases Its Nucleosome-Binding Specificity. *Nucleic Acids Res.* **2014**, *42* (20), 12498–12511. https://doi.org/10.1093/nar/gku995.

(157) Larson, A. G.; Elnatan, D.; Keenen, M. M.; Trnka, M. J.; Johnston, J. B.; Burlingame, A. L.; Agard, D. A.; Redding, S.; Narlikar, G. J. Liquid Droplet Formation by HP1α Suggests a Role for Phase Separation in Heterochromatin. *Nature* **2017**, *547* (7662), 236–240. https://doi.org/10.1038/nature22822.

(158) Strom, A. R.; Emelyanov, A. V.; Mir, M.; Fyodorov, D. V.; Darzacq, X.; Karpen, G. H. Phase Separation Drives Heterochromatin Domain Formation. *Nature* **2017**, *547* (7662), 241–245. https://doi.org/10.1038/nature22989.

(159) Wang, L.; Gao, Y.; Zheng, X.; Liu, C.; Dong, S.; Li, R.; Zhang, G.; Wei, Y.; Qu, H.; Li, Y.; et al. Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Mol. Cell* **2019**, *76* (4), 646-659.e6. https://doi.org/10.1016/j.molcel.2019.08.019.

(160) Cheutin, T. Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding. *Science* **2003**, *299* (5607), 721–725. https://doi.org/10.1126/science.1078572.

(161) Dialynas, G. K.; Terjung, S.; Brown, J. P.; Aucott, R. L.; Baron-Luhr, B.; Singh, P. B.; Georgatos, S. D. Plasticity of HP1 Proteins in Mammalian Cells. *J. Cell Sci.* **2007**, *120* (19), 3415–3424. https://doi.org/10.1242/jcs.012914.

(162) Ackermann, B. E.; Debelouchina, G. T. Heterochromatin Protein HP1α Gelation Dynamics Revealed by Solid-State NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2019**, *58* (19), 6300–6305. https://doi.org/10.1002/anie.201901141.

(163) Larson, A. G.; Narlikar, G. J. The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. *Biochemistry* **2018**, *57* (17), 2540–2548. https://doi.org/10.1021/acs.biochem.8b00401.

(164) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162* (5), 1066–1077. https://doi.org/10.1016/j.cell.2015.07.047.

(165) Du, M.; Chen, Z. J. DNA-Induced Liquid Phase Condensation of CGAS Activates Innate Immune Signaling. *Science* **2018**, *361* (6403), 704–709. https://doi.org/10.1126/science.aat1022.

(166) Sanulli, S.; Trnka, M. J.; Dharmarajan, V.; Tibble, R. W.; Pascal, B. D.; Burlingame, A. L.; Griffin, P. R.; Gross, J. D.; Narlikar, G. J. HP1 Reshapes Nucleosome Core to Promote Phase Separation of Heterochromatin. *Nature* **2019**, *575* (7782), 390–394. https://doi.org/10.1038/s41586-019-1669-2.

(167)   Chong, S.; Dugast-Darzacq, C.; Liu, Z.; Dong, P.; Dailey, G. M.; Cattoglio, C.; Heckert, A.; Banala, S.; Lavis, L.; Darzacq, X.; et al. Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription. *Science* **2018**, *361* (6400). https://doi.org/10.1126/science.aar2555.

(168)   Sabari, B. R.; Dall'Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; et al. Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science* **2018**, *361* (6400), eaar3958. https://doi.org/10.1126/science.aar3958.

(169)   Cho, W.-K.; Spille, J.-H.; Hecht, M.; Lee, C.; Li, C.; Grube, V.; Cisse, I. I. Mediator and RNA Polymerase II Clusters Associate in Transcription-Dependent Condensates. *Science* **2018**, *361* (6400), 412–415. https://doi.org/10.1126/science.aar4199.

(170)   Hnisz, D.; Shrinivas, K.; Young, R. A.; Chakraborty, A. K.; Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **2017**, *169* (1), 13–23. https://doi.org/10.1016/j.cell.2017.02.007.

(171)   Kilic, S.; Lezaja, A.; Gatti, M.; Bianco, E.; Michelena, J.; Imhof, R.; Altmeyer, M. Phase Separation of 53BP1 Determines Liquid-like Behavior of DNA Repair Compartments. *EMBO J.* **2019**, *38* (16). https://doi.org/10.15252/embj.2018101379.

(172)   Erdel, F.; Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.* **2018**, *114* (10), 2262–2270. https://doi.org/10.1016/j.bpj.2018.03.011.

(173)   Prescher, J. A.; Dube, D. H.; Bertozzi, C. R. Chemical Remodelling of Cell Surfaces in Living Animals. *Nature* **2004**, *430* (7002), 873–877. https://doi.org/10.1038/nature02791.

(174)   Maruani, A.; Smith, M. E. B.; Miranda, E.; Chester, K. A.; Chudasama, V.; Caddick, S. A Plug-and-Play Approach to Antibody-Based Therapeutics via a Chemoselective Dual Click Strategy. *Nat. Commun.* **2015**, *6* (1), 6645. https://doi.org/10.1038/ncomms7645.

(175)   Schnolzer, M.; Kent, S. B. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science* **1992**, *256* (5054), 221–225. https://doi.org/10.1126/science.1566069.

(176)   Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; et al. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012**, *149* (1), 214–231. https://doi.org/10.1016/j.cell.2012.02.013.

(177)   Bock, I.; Kudithipudi, S.; Tamas, R.; Kungulovski, G.; Dhayalan, A.; Jeltsch, A. Application of Celluspots Peptide Arrays for the Analysis of the Binding Specificity of Epigenetic Reading Domains to Modified Histone Tails. *BMC Biochem.* **2011**, *12* (1), 48. https://doi.org/10.1186/1471-2091-12-48.

(178)   Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D. Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition. *Curr. Biol.* **2011**, *21* (1), 53–58. https://doi.org/10.1016/j.cub.2010.11.058.

(179)   Bock, I.; Dhayalan, A.; Kudithipudi, S.; Brandt, O.; Rathert, P.; Jeltsch, A. Detailed Specificity Analysis of Antibodies Binding to Modified Histone Tails with Peptide Arrays. *Epigenetics* **2011**, *6* (2), 256–263. https://doi.org/10.4161/epi.6.2.13837.

(180)   Davies, N.; Lindsey, G. G. Histone H2B (and H2A) Ubiquitination Allows Normal Histone Octamer and Core Particle Reconstitution. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **1994**, *1218* (2), 187–193. https://doi.org/10.1016/0167-4781(94)90009-4.

(181)   West, M. H. P.; Bonner, W. M. Histone 2B Can Be Modified by the Attachment of Ubiquitin. *Nucleic Acids Res.* **1980**, *8* (20), 4671–4680. https://doi.org/10.1093/nar/8.20.4671.

(182)   Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically Encoding Nε-Acetyllysine in Recombinant Proteins. *Nat. Chem. Biol.* **2008**, *4* (4), 232–234. https://doi.org/10.1038/nchembio.73.

(183)   Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. Genetically Encoding N^ϵ-Methyl-ʟ-Lysine in Recombinant Histones. *J. Am. Chem. Soc.* **2009**, *131* (40), 14194–14195. https://doi.org/10.1021/ja906603s.

(184)   Ai, H.; Lee, J. W.; Schultz, P. G. A Method to Site-Specifically Introduce Methyllysine into Proteins in E. Coli. *Chem. Commun.* **2010**, *46* (30), 5506. https://doi.org/10.1039/c0cc00108b.

(185)   Nguyen, D. P.; Alai, M. M. G.; Virdee, S.; Chin, J. W. Genetically Directing Ɛ-N, N-Dimethyl-l-Lysine in Recombinant Histones. *Chem. Biol.* **2010**, *17* (10), 1072–1076. https://doi.org/10.1016/j.chembiol.2010.07.013.

(186)   Wilkins, B. J.; Rall, N. A.; Ostwal, Y.; Kruitwagen, T.; Hiragami-Hamada, K.; Winkler, M.; Barral, Y.; Fischle, W.; Neumann, H. A Cascade of Histone Modifications Induces Chromatin Condensation in Mitosis. *Science* **2014**, *343* (6166), 77–80. https://doi.org/10.1126/science.1244508.

(187) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia Coli. *Proc. Natl. Acad. Sci.* **2002**, *99* (17), 11020–11024. https://doi.org/10.1073/pnas.172226299.

(188) Xie, X.; Li, X.-M.; Qin, F.; Lin, J.; Zhang, G.; Zhao, J.; Bao, X.; Zhu, R.; Song, H.; Li, X. D.; et al. Genetically Encoded Photoaffinity Histone Marks. *J. Am. Chem. Soc.* **2017**, *139* (19), 6522–6525. https://doi.org/10.1021/jacs.7b01431.

(189) Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W. Traceless and Site-Specific Ubiquitination of Recombinant Proteins. *J. Am. Chem. Soc.* **2011**, 4.

(190) Lammers, C.; Hahn, L. E.; Neumann, H. Optimized Plasmid Systems for the Incorporation of Multiple Different Unnatural Amino Acids by Evolved Orthogonal Ribosomes. *ChemBioChem* **2014**, *15* (12), 1800–1804. https://doi.org/10.1002/cbic.201402033.

(191) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - Asian J.* **2009**, *4* (5), 630–640. https://doi.org/10.1002/asia.200800427.

(192) 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. https://doi.org/10.1016/j.cell.2006.12.041.

(193) 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. https://doi.org/10.1021/ja103954u.

(194) Lu, X.; Simon, M. D.; Chodaparambil, J. V.; Hansen, J. C.; Shokat, K. M.; Luger, K. The Effect of H3K79 Dimethylation and H4K20 Trimethylation on Nucleosome and Chromatin Structure. *Nat. Struct. Mol. Biol.* **2008**, *15* (10), 1122–1124. https://doi.org/10.1038/nsmb.1489.

(195) Hung, T.; Binda, O.; Champagne, K. S.; Kuo, A. J.; Johnson, K.; Chang, H. Y.; Simon, M. D.; Kutateladze, T. G.; Gozani, O. ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. *Mol. Cell* **2009**, *33* (2), 248–256. https://doi.org/10.1016/j.molcel.2008.12.016.

(196) Eidahl, J. O.; Crowe, B. L.; North, J. A.; McKee, C. J.; Shkriabai, N.; Feng, L.; Plumb, M.; Graham, R. L.; Gorelick, R. J.; Hess, S.; et al. Structural Basis for High-Affinity Binding of LEDGF PWWP to Mononucleosomes. *Nucleic Acids Res.* **2013**, *41* (6), 3924–3936. https://doi.org/10.1093/nar/gkt074.

(197) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130* (15), 5052–5053. https://doi.org/10.1021/ja800800p.

(198) 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. https://doi.org/10.1002/anie.200802336.

(199) 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. https://doi.org/10.1021/bi300535a.

(200) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. Disulfide-Directed Histone Ubiquitylation Reveals Plasticity in HDot1L Activation. *Nat. Chem. Biol.* **2010**, *6* (4), 267–269. https://doi.org/10.1038/nchembio.315.

(201) Fierz, B.; Chatterjee, C.; McGinty, R. K.; Bar-Dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B Ubiquitylation Disrupts Local and Higher-Order Chromatin Compaction. *Nat. Chem. Biol.* **2011**, *7* (2), 113–119. https://doi.org/10.1038/nchembio.501.

(202) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T.-H.; Chatterjee, C. Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-Range Internucleosomal Interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837. https://doi.org/10.1074/jbc.M114.591644.

(203) 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. https://doi.org/10.1002/anie.201103754.

(204) Seeliger, D.; Soeroes, S.; Klingberg, R.; Schwarzer, D.; Grubmüller, H.; Fischle, W. Quantitative Assessment of Protein Interaction with Methyl-Lysine Analogues by Hybrid Computational and Experimental Approaches. *ACS Chem. Biol.* **2012**, *7* (1), 150–154. https://doi.org/10.1021/cb200363r.

(205) Stephanopoulos, N.; Francis, M. B. Choosing an Effective Protein Bioconjugation Strategy. *Nat. Chem. Biol.* **2011**, *7* (12), 876–884. https://doi.org/10.1038/nchembio.720.

(206) Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. – Eur. J.* **2019**, *25* (1), 43–59. https://doi.org/10.1002/chem.201803174.

(207) Baldwin, A. D.; Kiick, K. L. Tunable Degradation of Maleimide–Thiol Adducts in Reducing Environments. *Bioconjug. Chem.* **2011**, *22* (10), 1946–1953. https://doi.org/10.1021/bc200148v.

(208) Ryan, C. P.; Smith, M. E. B.; Schumacher, F. F.; Grohmann, D.; Papaioannou, D.; Waksman, G.; Werner, F.; Baker, J. R.; Caddick, S. Tunable Reagents for Multi-Functional Bioconjugation: Reversible or Permanent Chemical Modification of Proteins and Peptides by Control of Maleimide Hydrolysis. *Chem. Commun.* **2011**, *47* (19), 5452. https://doi.org/10.1039/c1cc11114k.

(209) Hoyle, C. E.; Bowman, C. N. Thiol–Ene Click Chemistry. *Angew. Chem. Int. Ed.* **2010**, *49* (9), 1540–1573. https://doi.org/10.1002/anie.200903924.

(210) Massi, A.; Nanni, D. Thiol–Yne Coupling: Revisiting Old Concepts as a Breakthrough for up-to-Date Applications. *Org. Biomol. Chem.* **2012**, *10* (19), 3791. https://doi.org/10.1039/c2ob25217a.

(211) Tedaldi, L. M.; Smith, M. E. B.; Nathani, R. I.; Baker, J. R. Bromomaleimides: New Reagents for the Selective and Reversible Modification of Cysteine. *Chem. Commun.* **2009**, No. 43, 6583. https://doi.org/10.1039/b915136b.

(212) Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R. Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132* (6), 1960–1965. https://doi.org/10.1021/ja908610s.

(213) Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. A Perfluoroaryl-Cysteine S $_N$ Ar Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013**, *135* (16), 5946–5949. https://doi.org/10.1021/ja400119t.

(214) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic Palladium Reagents for Cysteine Bioconjugation. *Nature* **2015**, *526* (7575), 687–691. https://doi.org/10.1038/nature15739.

(215) Willwacher, J.; Raj, R.; Mohammed, S.; Davis, B. G. Selective Metal-Site-Guided Arylation of Proteins. *J. Am. Chem. Soc.* **2016**, *138* (28), 8678–8681. https://doi.org/10.1021/jacs.6b04043.

(216) Messina, M. S.; Stauber, J. M.; Waddington, M. A.; Rheingold, A. L.; Maynard, H. D.; Spokoyny, A. M. Organometallic Gold(III) Reagents for Cysteine Arylation. *J. Am. Chem. Soc.* **2018**, *140* (23), 7065–7069. https://doi.org/10.1021/jacs.8b04115.

(217) de Paiva, R. E. F.; Du, Z.; Nakahata, D. H.; Lima, F. A.; Corbi, P. P.; Farrell, N. P. Gold-Catalyzed C–S Aryl-Group Transfer in Zinc Finger Proteins. *Angew. Chem. Int. Ed.* **2018**, *57* (30), 9305–9309. https://doi.org/10.1002/anie.201803082.

(218) Kasper, M.-A.; Glanz, M.; Stengl, A.; Penkert, M.; Klenk, S.; Sauer, T.; Schumacher, D.; Helma, J.; Krause, E.; Cardoso, M. C.; et al. Cysteine-Selective Phosphonamidate Electrophiles for Modular Protein Bioconjugations. *Angew. Chem. Int. Ed.* **2019**, *58* (34), 11625–11630. https://doi.org/10.1002/anie.201814715.

(219) Frei, R.; Waser, J. A Highly Chemoselective and Practical Alkynylation of Thiols. *J. Am. Chem. Soc.* **2013**, *135* (26), 9620–9623. https://doi.org/10.1021/ja4044196.

(220) Frei, R.; Wodrich, M. D.; Hari, D. P.; Borin, P.-A.; Chauvier, C.; Waser, J. Fast and Highly Chemoselective Alkynylation of Thiols with Hypervalent Iodine Reagents Enabled through a Low Energy Barrier Concerted Mechanism. *J. Am. Chem. Soc.* **2014**, *136* (47), 16563–16573. https://doi.org/10.1021/ja5083014.

(221) Abegg, D.; Frei, R.; Cerato, L.; Prasad Hari, D.; Wang, C.; Waser, J.; Adibekian, A. Proteome-Wide Profiling of Targets of Cysteine Reactive Small Molecules by Using Ethynyl Benziodoxolone Reagents. *Angew. Chem. Int. Ed.* **2015**, *54* (37), 10852–10857. https://doi.org/10.1002/anie.201505641.

(222) Tessier, R.; Ceballos, J.; Guidotti, N.; Simonet-Davin, R.; Fierz, B.; Waser, J. "Doubly Orthogonal" Labeling of Peptides and Proteins. *Chem* **2019**, *5* (8), 2243–2263. https://doi.org/10.1016/j.chempr.2019.06.022.

(223) Fischer, E.; Fourneau, E. Ueber Einige Derivate Des Glykocolls. *Berichte Dtsch. Chem. Ges.* **1901**, *34* (2), 2868–2877. https://doi.org/10.1002/cber.190103402249.

(224) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. https://doi.org/10.1021/ja00897a025.

(225) Rink, H. Solid-Phase Synthesis of Protected Peptide Fragments Using a Trialkoxy-Diphenyl-Methylester Resin. *Tetrahedron Lett.* **1987**, *28* (33), 3787–3790. https://doi.org/10.1016/S0040-4039(00)96384-6.

(226) Mellor, S. L.; McGuire, C.; Chan, W. C. N-Fmoc-Aminooxy-2-Chlorotrityl Polystyrene Resin: A Facile Solid-Phase Methodology for the Synthesis of Hydroxamic AcidsI. 4.

(227) Stavropoulos, G.; Gatos, D.; Magafa, V.; Barlos, K. Preparation of Polymer-Bound Trityl-Hydrazines and Their Application in the Solid Phase Synthesis of Partially Protected Peptide Hydrazides. *Lett. Pept. Sci.* **1996**, *2* (5), 315–318. https://doi.org/10.1007/BF00142245.

(228) Wang, S.-S. />-Alkoxybenzyl Alcohol Resin and Jfr-Alkoxybenzyloxycarbonylhydrazide Resin for Solid Phase Synthesis of Protected Peptide Fragments. *J. Am. Chem. Soc.* **1973**, 6.

(229) Sheehan, J. C.; Hess, G. P. A New Method of Forming Peptide Bonds. *J. Am. Chem. Soc.* **1955**, *77* (4), 1067–1068. https://doi.org/10.1021/ja01609a099.

(230) Benoiton, N. L.; Chen, F. M. F. 2-Alkoxy-5(4 *H*)-Oxazolones from *N* -Alkoxycarbonylamino Acids and Their Implication in Carbodiimide-Mediated Reactions in Peptide Synthesis. *Can. J. Chem.* **1981**, *59* (2), 384–389. https://doi.org/10.1139/v81-059.

(231) König, W.; Geiger, R. Eine neue Methode zur Synthese von Peptiden: Aktivierung der Carboxylgruppe mit Dicyclohexylcarbodiimid unter Zusatz von 1-Hydroxy-benzotriazolen. *Chem. Ber.* **1970**, *103* (3), 788–798. https://doi.org/10.1002/cber.19701030319.

(232) Carpino, L. A.; El-Faham, A. The Diisopropylcarbodiimide/ 1-Hydroxy-7-Azabenzotriazole System: Segment Coupling and Stepwise Peptide Assembly. *Tetrahedron* **1999**, *55* (22), 6813–6830. https://doi.org/10.1016/S0040-4020(99)00344-0.

(233) Castro, B.; Dormoy, J. R.; Evin, G.; Selve, C. Reactifs de couplage peptidique I (1) - l'hexafluorophosphate de benzotriazolyl N-oxytrisdimethylamino phosphonium (B.O.P.). *Tetrahedron Lett.* **1975**, *16* (14), 1219–1222. https://doi.org/10.1016/S0040-4039(00)72100-9.

(234) Coste, J.; Le-Nguyen, D.; Castro, B. PyBOP®: A New Peptide Coupling Reagent Devoid of Toxic by-Product. *Tetrahedron Lett.* **1990**, *31* (2), 205–208. https://doi.org/10.1016/S0040-4039(00)94371-5.

(235) Albericio, F.; Cases, M.; Alsina, J.; Triolo, S. A.; Carpino, L. A.; Kates, S. A. On the Use of PyAOP, a Phosphonium Salt Derived from HOAt, in Solid-Phase Peptide Synthesis. *Tetrahedron Lett.* **1997**, *38* (27), 4853–4856. https://doi.org/10.1016/S0040-4039(97)01011-3.

(236) Dourtoglou, V.; Gross, B.; Lambropoulou, V.; Zioudrou, C. O-Benzotriazolyl-N,N,N′,N′-Tetramethyluronium Hexafluorophosphate as Coupling Reagent for the Synthesis of Peptides of Biological Interest. *Synthesis* **1984**, *1984* (7), 572–574. https://doi.org/10.1055/s-1984-30895.

(237) Marder, O.; Shvo, Y.; Albericio, F. HCTU and TCTU. New Coupling Reagents: Development and Industrial Aspects. *Chim. Oggi* **2002**, *20* (7–8), 37–41.

(238) Carpino, L. A. 1-Hydroxy-7-Azabenzotriazole. An Efficient Peptide Coupling Additive. *J. Am. Chem. Soc.* **1993**, *115* (10), 4397–4398. https://doi.org/10.1021/ja00063a082.

(239) El-Faham, A.; Funosas, R. S.; Prohens, R.; Albericio, F. COMU: A Safer and More Effective Replacement for Benzotriazole-Based Uronium Coupling Reagents. *Chem. – Eur. J.* **2009**, *15* (37), 9404–9416. https://doi.org/10.1002/chem.200900615.

(240) Al-Warhi, T. I.; Al-Hazimi, H. M. A.; El-Faham, A. Recent Development in Peptide Coupling Reagents. *J. Saudi Chem. Soc.* **2012**, *16* (2), 97–116. https://doi.org/10.1016/j.jscs.2010.12.006.

(241) Atherton, E.; Logan, C. J.; Sheppard, R. C. Peptide Synthesis. Part 2. Procedures for Solid-Phase Synthesis Using Nα-Fluorenylmethoxycarbonylamino-Acids on Polyamide Supports. Synthesis of Substance P and of Acyl Carrier Protein 65–74 Decapeptide. *J. Chem. Soc. Perkin 1* **1981**, No. 0, 538–546. https://doi.org/10.1039/P19810000538.

(242) Hirschmann, R.; Yao, W.; Arison, B.; Maechler, L.; Rosegay, A.; Sprengeler, P. A.; Smith, A. B. Synthesis of the First Tricyclic Homodetic Peptide. Use of Coordinated Orthogonal Deprotection to Achieve Directed Ring Closure. *Tetrahedron* **1998**, *54* (25), 7179–7202. https://doi.org/10.1016/S0040-4020(98)00354-8.

(243) Royo, M.; Farrera-Sinfreu, J.; Sole, L.; Albericio, F. Four-Dimensional Orthogonal Solid-Phase Synthesis of New Scaffolds Based on Cyclic Tetra-b-Peptides. *Tetrahedron Lett.* **2002**, 4.

(244) Thieriet, N.; Alsina, J.; Giralt, E.; Guibé, F.; Albericio, F. Use of Alloc-Amino Acids in Solid-Phase Peptide Synthesis. Tandem Deprotection-Coupling Reactions Using Neutral Conditions. *Tetrahedron Lett.* **1997**, *38* (41), 7275–7278. https://doi.org/10.1016/S0040-4039(97)01690-0.

(245) Veber, D.; Milkowski, J.; Varga, S.; Denkewalter, R.; Hirschmann, R. Acetamidomethyl. A Novel Thiol Protecting Group for Cysteine. *J. Am. Chem. Soc.* **1972**, *94* (15), 5456–5461. https://doi.org/10.1021/ja00770a600.

(246)   Andreu, D.; Albericio, F.; Solé, N. A.; Munson, M. C.; Ferrer, M.; Barany, G. Formation of Disulfide Bonds in Synthetic Peptides and Proteins. In *Peptide Synthesis Protocols*; Pennington, M. W., Dunn, B. M., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 1995; pp 91–169. https://doi.org/10.1385/0-89603-273-6:91.

(247)   Annis, I.; Hargittai, B.; Barany, G. [10] Disulfide Bond Formation in Peptides. In *Methods in Enzymology*; Solid-Phase Peptide Synthesis; Academic Press, **1997**; Vol. 289, pp 198–221. https://doi.org/10.1016/S0076-6879(97)89049-0.

(248)   Mutter, M.; Nefzi, A.; Sato, T.; Sun, X.; Wahl, F.; Wöhr, T. Pseudo-Prolines (Psi Pro) for Accessing "Inaccessible" Peptides. *Pept. Res.* **1995**, *8* (3), 145–153.

(249)   Assem, N.; Natarajan, A.; Yudin, A. K. Chemoselective Peptidomimetic Ligation Using Thioacid Peptides and Aziridine Templates. *J. Am. Chem. Soc.* **2010**, *132* (32), 10986–10987. https://doi.org/10.1021/ja104488d.

(250)   Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2* (14), 2141–2143. https://doi.org/10.1021/ol006054v.

(251)   Bode, J. W.; Fox, R. M.; Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and α-Ketoacids. *Angew. Chem. Int. Ed.* **2006**, *45* (8), 1248–1252. https://doi.org/10.1002/anie.200503991.

(252)   Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. Salicylaldehyde Ester-Induced Chemoselective Peptide Ligations: Enabling Generation of Natural Peptidic Linkages at the Serine/Threonine Sites. *Org. Lett.* **2010**, *12* (8), 1724–1727. https://doi.org/10.1021/ol1003109.

(253)   Ulrich, S.; Boturyn, D.; Marra, A.; Renaudet, O.; Dumy, P. Oxime Ligation: A Chemoselective Click-Type Reaction for Accessing Multifunctional Biomolecular Constructs. *Chem. – Eur. J.* **2014**, *20* (1), 34–41. https://doi.org/10.1002/chem.201302426.

(254)   Noda, H.; Erős, G.; Bode, J. W. Rapid Ligations with Equimolar Reactants in Water with the Potassium Acyltrifluoroborate (KAT) Amide Formation. *J. Am. Chem. Soc.* **2014**, *136* (15), 5611–5614. https://doi.org/10.1021/ja5018442.

(255)   Liu, C.-F.; Tam, J. P. Chemical Ligation Approach To Form a Peptide Bond between Unprotected Peptide Segments. Concept and Model Study. *J. Am. Chem. Soc.* **1994**, *116* (10), 4149–4153. https://doi.org/10.1021/ja00089a001.

(256)   Araújo, A. D. de; Palomo, J. M.; Cramer, J.; Seitz, O.; Alexandrov, K.; Waldmann, H. Diels–Alder Ligation of Peptides and Proteins. *Chem. – Eur. J.* **2006**, *12* (23), 6095–6109. https://doi.org/10.1002/chem.200600148.

(257)   Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004**, *126* (9), 2670–2671. https://doi.org/10.1021/ja039915e.

(258)   Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779. https://doi.org/10.1126/science.7973629.

(259)   Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. Protein Synthesis by Native Chemical Ligation: Expanded Scope by Using Straightforward Methodology. *Proc. Natl. Acad. Sci.* **1999**, *96* (18), 10068–10073. https://doi.org/10.1073/pnas.96.18.10068.

(260)   Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/ *t* -Bu Chemistry. *J. Am. Chem. Soc.* **1999**, *121* (49), 11369–11374. https://doi.org/10.1021/ja992668n.

(261)   Camarero, J. A.; Adeva, A.; Muir, T. W. 3-Thiopropionic Acid as a Highly Versatile Multidetachable Thioester Resin Linker. *Lett. Pept. Sci.* **2000**, *7* (1), 17–21. https://doi.org/10.1007/BF02443557.

(262)   Botti, P.; Villain, M.; Manganiello, S.; Gaertner, H. Native Chemical Ligation through in Situ O to S Acyl Shift. *Org. Lett.* **2004**, *6* (26), 4861–4864. https://doi.org/10.1021/ol0481028.

(263)   Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed.* **2008**, *47* (36), 6851–6855. https://doi.org/10.1002/anie.200705471.

(264)   Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org. Lett.* **2010**, *12* (22), 5238–5241. https://doi.org/10.1021/ol102273u.

(265)   Raibaut, L.; Cargoët, M.; Ollivier, N.; Chang, Y. M.; Drobecq, H.; Boll, E.; Desmet, R.; Monbaliu, J.-C. M.; Melnyk, O. Accelerating Chemoselective Peptide Bond Formation Using Bis(2-Selenylethyl)Amido Peptide Selenoester Surrogates. *Chem. Sci.* **2016**, *7* (4), 2657–2665. https://doi.org/10.1039/C5SC03459K.

(266) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* **2011**, *50* (33), 7645–7649. https://doi.org/10.1002/anie.201100996.

(267) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protoc.* **2013**, *8* (12), 2483–2495. https://doi.org/10.1038/nprot.2013.152.

(268) Wang, C.; Guo, Q.-X.; Fu, Y. Theoretical Analysis of the Detailed Mechanism of Native Chemical Ligation Reactions. *Chem. – Asian J.* **2011**, *6* (5), 1241–1251. https://doi.org/10.1002/asia.201000760.

(269) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives. *J. Am. Chem. Soc.* **1997**, *119* (19), 4325–4329. https://doi.org/10.1021/ja962656r.

(270) Johnson, E. C. B.; Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **2006**, *128* (20), 6640–6646. https://doi.org/10.1021/ja058344i.

(271) Evans, T. C.; Benner, J.; Xu, M.-Q. Semisynthesis of Cytotoxic Proteins Using a Modified Protein Splicing Element. *Protein Sci.* **1998**, *7* (11), 2256–2264. https://doi.org/10.1002/pro.5560071103.

(272) Huber, R. E.; Criddle, R. S. Comparison of the Chemical Properties of Selenocysteine and Selenocystine with Their Sulfur Analogs. *Arch. Biochem. Biophys.* **1967**, *122* (1), 164–173. https://doi.org/10.1016/0003-9861(67)90136-1.

(273) Gieselman, M. D.; Xie, L.; van der Donk, W. A. Synthesis of a Selenocysteine-Containing Peptide by Native Chemical Ligation. *Org. Lett.* **2001**, *3* (9), 1331–1334. https://doi.org/10.1021/ol015712o.

(274) Hondal, R. J.; Nilsson, B. L.; Raines, R. T. Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation. *J. Am. Chem. Soc.* **2001**, *123* (21), 5140–5141. https://doi.org/10.1021/ja005885t.

(275) Durek, T.; Alewood, P. F. Preformed Selenoesters Enable Rapid Native Chemical Ligation at Intractable Sites. *Angew. Chem. Int. Ed.* **2011**, *50* (50), 12042–12045. https://doi.org/10.1002/anie.201105512.

(276) Chisholm, T. S.; Clayton, D.; Dowman, L. J.; Sayers, J.; Payne, R. J. Native Chemical Ligation–Photodesulfurization in Flow. *J. Am. Chem. Soc.* **2018**, *140* (29), 9020–9024. https://doi.org/10.1021/jacs.8b03115.

(277) 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. https://doi.org/10.1021/ja003265m.

(278) Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J. Rapid and Efficient Protein Synthesis through Expansion of the Native Chemical Ligation Concept. *Nat. Rev. Chem.* **2018**, *2* (4), 1–17. https://doi.org/10.1038/s41570-018-0122.

(279) Nishide, K.; Shigeta, Y.; Obata, K.; Inoue, T.; Node, M. Reductive Desulfurization Using the Raney Nickel — Sodium Hypophosphite Combination System without Racemization of a Secondary Alcohol. *Tetrahedron Lett.* **1996**, *37* (13), 2271–2274. https://doi.org/10.1016/0040-4039(96)00273-0.

(280) Node, M.; Nishide, K.; Shigeta, Y.; Obata, K.; Shiraki, H.; Kunishige, H. A Raney Nickel — Sodium Hypophosphite Combination System for Reductive Desulfurization without Racemization of Optically Active Secondary Alcohol. *Tetrahedron* **1997**, *53* (38), 12883–12894. https://doi.org/10.1016/S0040-4020(97)00804-1.

(281) Hoffmann, F. W.; Ess, R. J.; Simmons, T. C.; Hanzel, R. S. The desulfurization of mercaptans with trialkyl phosphites. *J. Am. Chem. Soc.* **1956**, *78* (24), 6414–6414. https://doi.org/10.1021/ja01605a034.

(282) Walling, C.; Rabinowitz, R. The Reaction of Trialkyl Phosphites with Thiyl and Alkoxy Radicals. *J. Am. Chem. Soc.* **1959**, *81* (5), 1243–1249. https://doi.org/10.1021/ja01514a057.

(283) Walling, C.; Basedow, O. H.; Savas, E. S. Some Extensions of the Reaction of Trivalent Phosphorus Derivatives with Alkoxy and Thiyl Radicals; a New Synthesis of Thioesters [1]. *J. Am. Chem. Soc.* **1960**, *82* (9), 2181–2184. https://doi.org/10.1021/ja01494a023.

(284) 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. https://doi.org/10.1002/anie.200704195.

(285) Siman, P.; Blatt, O.; Moyal, T.; Danieli, T.; Lebendiker, M.; Lashuel, H. A.; Friedler, A.; Brik, A. Chemical Synthesis and Expression of the HIV-1 Rev Protein. *Chembiochem Eur. J. Chem. Biol.* **2011**, *12* (7), 1097–1104. https://doi.org/10.1002/cbic.201100033.

(286) Thompson, R. E.; Liu, X.; Alonso-García, N.; Pereira, P. J. B.; Jolliffe, K. A.; Payne, R. J. Trifluoroethanethiol: An Additive for Efficient One-Pot Peptide Ligation–Desulfurization Chemistry. *J. Am. Chem. Soc.* **2014**, *136* (23), 8161–8164. https://doi.org/10.1021/ja502806r.

(287) Huang, Y.-C.; Chen, C.-C.; Gao, S.; Wang, Y.-H.; Xiao, H.; Wang, F.; Tian, C.-L.; Li, Y.-M. Synthesis of L- and d-Ubiquitin by One-Pot Ligation and Metal-Free Desulfurization. *Chem. – Eur. J.* **2016**, *22* (22), 7623–7628. https://doi.org/10.1002/chem.201600101.

(288) 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.* **2003**, *100* (21), 12033–12038. https://doi.org/10.1073/pnas.2035256100.

(289) Chiang, K. P.; Jensen, M. S.; McGinty, R. K.; Muir, T. W. A Semisynthetic Strategy to Generate Phosphorylated and Acetylated Histone H2B. *ChemBioChem* **2009**, *10* (13), 2182–2187. https://doi.org/10.1002/cbic.200900238.

(290) Gentle, I. E.; De Souza, D. P.; Baca, M. Direct Production of Proteins with N-Terminal Cysteine for Site-Specific Conjugation. *Bioconjug. Chem.* **2004**, *15* (3), 658–663. https://doi.org/10.1021/bc049965o.

(291) Erlanson, D. A.; Chytil, M.; Verdine, G. L. The Leucine Zipper Domain Controls the Orientation of AP-1 in the NFAT·AP-1·DNA Complex. *Chem. Biol.* **1996**, *3* (12), 981–991. https://doi.org/10.1016/S1074-5521(96)90165-9.

(292) Komarov, A. G.; Linn, K. M.; Devereaux, J. J.; Valiyaveetil, F. I. Modular Strategy for the Semisynthesis of a K + Channel: Investigating Interactions of the Pore Helix. *ACS Chem. Biol.* **2009**, *4* (12), 1029–1038. https://doi.org/10.1021/cb900210r.

(293) Tolbert, T. J.; Franke, D.; Wong, C.-H. A New Strategy for Glycoprotein Synthesis: Ligation of Synthetic Glycopeptides with Truncated Proteins Expressed in E. Coli as TEV Protease Cleavable Fusion Protein. *Bioorg. Med. Chem.* **2005**, *13* (3), 909–915. https://doi.org/10.1016/j.bmc.2004.06.047.

(294) Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed Protein Ligation: A General Method for Protein Engineering. *Proc. Natl. Acad. Sci.* **1998**, *95* (12), 6705–6710. https://doi.org/10.1073/pnas.95.12.6705.

(295) Evans, T. C.; Benner, J.; Xu, M.-Q. The in Vitro Ligation of Bacterially Expressed Proteins Using an Intein from Methanobacterium Thermoautotrophicum. *J. Biol. Chem.* **1999**, *274* (7), 3923–3926. https://doi.org/10.1074/jbc.274.7.3923.

(296) 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. https://doi.org/10.1074/jbc.R113.540302.

(297) Xu, M. Q.; Perler, F. B. The Mechanism of Protein Splicing and Its Modulation by Mutation. *EMBO J.* **1996**, *15* (19), 5146–5153. https://doi.org/10.1002/j.1460-2075.1996.tb00898.x.

(298) Shah, N. H.; Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem Sci* **2014**, *5* (2), 446–461. https://doi.org/10.1039/C3SC52951G.

(299) Shah, N. H.; Dann, G. P.; Vila-Perelló, M.; Liu, Z.; Muir, T. W. Ultrafast Protein Splicing Is Common among Cyanobacterial Split Inteins: Implications for Protein Engineering. *J. Am. Chem. Soc.* **2012**, *134* (28), 11338–11341. https://doi.org/10.1021/ja303226x.

(300) Bang, D.; Kent, S. B. H. A One-Pot Total Synthesis of Crambin. *Angew. Chem. Int. Ed.* **2004**, *43* (19), 2534–2538. https://doi.org/10.1002/anie.200353540.

(301) Bang, D.; Pentelute, B. L.; Kent, S. B. H. Kinetically Controlled Ligation for the Convergent Chemical Synthesis of Proteins. *Angew. Chem. Int. Ed.* **2006**, *45* (24), 3985–3988. https://doi.org/10.1002/anie.200600702.

(302) Fang, G.-M.; Wang, J.-X.; Liu, L. Convergent Chemical Synthesis of Proteins by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* **2012**, *51* (41), 10347–10350. https://doi.org/10.1002/anie.201203843.

(303) Lowary, P. T.; Widom, J. New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-Directed Nucleosome Positioning11Edited by T. Richmond. *J. Mol. Biol.* **1998**, *276* (1), 19–42. https://doi.org/10.1006/jmbi.1997.1494.

(304) Fierz, B.; Kilic, S.; Hieb, A. R.; Luger, K.; Muir, T. W. Stability of Nucleosomes Containing Homogenously Ubiquitylated H2A and H2B Prepared Using Semisynthesis. *J. Am. Chem. Soc.* **2012**, *134* (48), 19548–19551. https://doi.org/10.1021/ja308908p.

(305) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically Ubiquitylated Histone H2B Stimulates HDot1L-Mediated Intranucleosomal Methylation. *Nature* **2008**, *453* (7196), 812–816. https://doi.org/10.1038/nature06906.

(306) Siman, P.; Karthikeyan, S. V.; Nikolov, M.; Fischle, W.; Brik, A. Convergent Chemical Synthesis of Histone H2B Protein for the Site-Specific Ubiquitination at Lys34. *Angew. Chem. Int. Ed.* **2013**, *52* (31), 8059–8063. https://doi.org/10.1002/anie.201303844.

(307) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **2006**, *311* (5762), 844–847. https://doi.org/10.1126/science.1124000.

(308) Guidotti, N.; Lechner, C. C.; Fierz, B. Controlling the Supramolecular Assembly of Nucleosomes Asymmetrically Modified on H4. *Chem. Commun.* **2017**, *53* (74), 10267–10270. https://doi.org/10.1039/C7CC06180C.

(309) Guidotti, N.; Lechner, C. C.; Bachmann, A. L.; Fierz, B. A Modular Ligation Strategy for Asymmetric Bivalent Nucleosomes Trimethylated at K36 and K27. *ChemBioChem* **2019**, *20* (9), 1124–1128. https://doi.org/10.1002/cbic.201800744.

(310) Kilic, S.; Bachmann, A. L.; Bryan, L. C.; Fierz, B. Multivalency Governs HP1α Association Dynamics with the Silent Chromatin State. *Nat. Commun.* **2015**, *6* (1), 7313. https://doi.org/10.1038/ncomms8313.

(311) Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modi®cations. **2000**, *403*, 5.

(312) Vastenhouw, N. L.; Schier, A. F. Bivalent Histone Modifications in Early Embryogenesis. *Curr. Opin. Cell Biol.* **2012**, *24* (3), 374–386. https://doi.org/10.1016/j.ceb.2012.03.009.

(313) Gao, Z.; Zhang, J.; Bonasio, R.; Strino, F.; Sawai, A.; Parisi, F.; Kluger, Y.; Reinberg, D. PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* **2012**, *45* (3), 344–356. https://doi.org/10.1016/j.molcel.2012.01.002.

(314) Czermin, B.; Melfi, R.; McCabe, D.; Seitz, V.; Imhof, A.; Pirrotta, V. Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity That Marks Chromosomal Polycomb Sites. *Cell* **2002**, *111* (2), 185–196. https://doi.org/10.1016/S0092-8674(02)00975-3.

(315) Kuzmichev, A.; Nishioka, K.; Erdjument-Bromage, H.; Tempst, P.; Reinberg, D. Histone Methyltransferase Activity Associated with a Human Multiprotein Complex Containing the Enhancer of Zeste Protein. *Genes Dev.* **2002**, *16* (22), 2893–2905. https://doi.org/10.1101/gad.1035902.

(316) Margueron, R.; Li, G.; Sarma, K.; Blais, A.; Zavadil, J.; Woodcock, C. L.; Dynlacht, B. D.; Reinberg, D. Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol. Cell* **2008**, *32* (4), 503–518. https://doi.org/10.1016/j.molcel.2008.11.004.

(317) Müller, J.; Hart, C. M.; Francis, N. J.; Vargas, M. L.; Sengupta, A.; Wild, B.; Miller, E. L.; O'Connor, M. B.; Kingston, R. E.; Simon, J. A. Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. *Cell* **2002**, *111* (2), 197–208. https://doi.org/10.1016/S0092-8674(02)00976-5.

(318) Shen, X.; Liu, Y.; Hsu, Y.-J.; Fujiwara, Y.; Kim, J.; Mao, X.; Yuan, G.-C.; Orkin, S. H. EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Mol. Cell* **2008**, *32* (4), 491–502. https://doi.org/10.1016/j.molcel.2008.10.016.

(319) Smits, A. H.; Jansen, P. W. T. C.; Poser, I.; Hyman, A. A.; Vermeulen, M. Stoichiometry of Chromatin-Associated Protein Complexes Revealed by Label-Free Quantitative Mass Spectrometry-Based Proteomics. *Nucleic Acids Res.* **2013**, *41* (1), e28–e28. https://doi.org/10.1093/nar/gks941.

(320) Tanay, A.; O'Donnell, A. H.; Damelin, M.; Bestor, T. H. Hyperconserved CpG Domains Underlie Polycomb-Binding Sites. *Proc. Natl. Acad. Sci.* **2007**, *104* (13), 5521–5526. https://doi.org/10.1073/pnas.0609746104.

(321) Ku, M.; Koche, R. P.; Rheinbay, E.; Mendenhall, E. M.; Endoh, M.; Mikkelsen, T. S.; Presser, A.; Nusbaum, C.; Xie, X.; Chi, A. S.; et al. Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet.* **2008**, *4* (10), e1000242. https://doi.org/10.1371/journal.pgen.1000242.

(322) Schmitges, F. W.; Prusty, A. B.; Faty, M.; Stützer, A.; Lingaraju, G. M.; Aiwazian, J.; Sack, R.; Hess, D.; Li, L.; Zhou, S.; et al. Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* **2011**, *42* (3), 330–341. https://doi.org/10.1016/j.molcel.2011.03.025.

(323) Poepsel, S.; Kasinath, V.; Nogales, E. Cryo-EM Structures of PRC2 Simultaneously Engaged with Two Functionally Distinct Nucleosomes. *Nat. Struct. Mol. Biol.* **2018**, *25* (2), 154–162. https://doi.org/10.1038/s41594-018-0023-y.

(324) Jiao, L.; Liu, X. Structural Basis of Histone H3K27 Trimethylation by an Active Polycomb Repressive Complex 2. *Science* **2015**, *350* (6258). https://doi.org/10.1126/science.aac4383.

(325) Lee, C.-H.; Yu, J.-R.; Kumar, S.; Jin, Y.; LeRoy, G.; Bhanu, N.; Kaneko, S.; Garcia, B. A.; Hamilton, A. D.; Reinberg, D. Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin. *Mol. Cell* **2018**, *70* (3), 422-434.e6. https://doi.org/10.1016/j.molcel.2018.03.020.

(326) Li, H.; Liefke, R.; Jiang, J.; Kurland, J. V.; Tian, W.; Deng, P.; Zhang, W.; He, Q.; Patel, D. J.; Bulyk, M. L.; et al. Polycomb-like Proteins Link the PRC2 Complex to CpG Islands. *Nature* **2017**, *549* (7671), 287–291. https://doi.org/10.1038/nature23881.

(327) Chen, Z.; Grzybowski, A. T.; Ruthenburg, A. J. Traceless Semisynthesis of a Set of Histone 3 Species Bearing Specific Lysine Methylation Marks. *ChemBioChem* **2014**, *15* (14), 2071–2075. https://doi.org/10.1002/cbic.201402313.

(328) Casadio, F.; Lu, X.; Pollock, S. B.; LeRoy, G.; Garcia, B. A.; Muir, T. W.; Roeder, R. G.; Allis, C. D. H3R42me2a Is a Histone Modification with Positive Transcriptional Effects. *Proc. Natl. Acad. Sci.* **2013**, *110* (37), 14894–14899. https://doi.org/10.1073/pnas.1312925110.

(329) Sato, K.; Tanaka, S.; Yamamoto, K.; Tashiro, Y.; Narumi, T.; Mase, N. Direct Synthesis of N-Terminal Thiazolidine-Containing Peptide Thioesters from Peptide Hydrazides. *Chem. Commun.* **2018**, *54* (66), 9127–9130. https://doi.org/10.1039/C8CC03591A.

(330) Margueron, R.; Justin, N.; Ohno, K.; Sharpe, M. L.; Son, J.; Iii, W. J. D.; Voigt, P.; Martin, S. R.; Taylor, W. R.; Marco, V. D.; et al. Role of the Polycomb Protein EED in the Propagation of Repressive Histone Marks. *Nature* **2009**, *461* (7265), 762–767. https://doi.org/10.1038/nature08398.

(331) Choi, J.; Bachmann, A. L.; Tauscher, K.; Benda, C.; Fierz, B.; Müller, J. DNA Binding by PHF1 Prolongs PRC2 Residence Time on Chromatin and Thereby Promotes H3K27 Methylation. *Nat. Struct. Mol. Biol.* **2017**, *24* (12), 1039–1047. https://doi.org/10.1038/nsmb.3488.

(332) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically Encoded Short Peptide Tag for Versatile Protein Labeling by Sfp Phosphopantetheinyl Transferase. *Proc. Natl. Acad. Sci.* **2005**, *102* (44), 15815–15820. https://doi.org/10.1073/pnas.0507705102.

(333) Wang, X.; Paucek, R. D.; Gooding, A. R.; Brown, Z. Z.; Ge, E. J.; Muir, T. W.; Cech, T. R. Molecular Analysis of PRC2 Recruitment to DNA in Chromatin and Its Inhibition by RNA. *Nat. Struct. Mol. Biol.* **2017**, *24* (12), 1028–1038. https://doi.org/10.1038/nsmb.3487.

(334) Starheim, K. K.; Gevaert, K.; Arnesen, T. Protein N-Terminal Acetyltransferases: When the Start Matters. *Trends Biochem. Sci.* **2012**, *37* (4), 152–161. https://doi.org/10.1016/j.tibs.2012.02.003.

(335) Stock, A.; Clarke, S.; Clarke, C.; Stock, J. N-Terminal Methylation of Proteins: Structure, Function and Specificity. *FEBS Lett.* **1987**, *220* (1), 8–14. https://doi.org/10.1016/0014-5793(87)80866-9.

(336) Martin, D. D. O.; Beauchamp, E.; Berthiaume, L. G. Post-Translational Myristoylation: Fat Matters in Cellular Life and Death. *Biochimie* **2011**, *93* (1), 18–31. https://doi.org/10.1016/j.biochi.2010.10.018.

(337) Buglino, J. A.; Resh, M. D. Chapter Ten - Palmitoylation of Hedgehog Proteins. In *Vitamins & Hormones*; Litwack, G., Ed.; Hedgehog Signaling; Academic Press, **2012**; Vol. 88, pp 229–252. https://doi.org/10.1016/B978-0-12-394622-5.00010-9.

(338) Tooley, J. G.; Tooley, C. E. S. New Roles for Old Modifications: Emerging Roles of N-Terminal Post-Translational Modifications in Development and Disease. *Protein Sci.* **2014**, *23* (12), 1641–1649. https://doi.org/10.1002/pro.2547.

(339) 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. https://doi.org/10.1002/pmic.201400619.

(340) Martin, C.; Zhang, Y. The Diverse Functions of Histone Lysine Methylation. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 838–849. https://doi.org/10.1038/nrm1761.

(341) Rodriguez, J.; Munoz, M.; Vives, L.; Frangou, C. G.; Groudine, M.; Peinado, M. A. Bivalent Domains Enforce Transcriptional Memory of DNA Methylated Genes in Cancer Cells. *Proc. Natl. Acad. Sci.* **2008**, *105* (50), 19809–19814. https://doi.org/10.1073/pnas.0810133105.

(342) Pesavento, J. J.; Yang, H.; Kelleher, N. L.; Mizzen, C. A. Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle. *Mol. Cell. Biol.* **2008**, *28* (1), 468–486. https://doi.org/10.1128/MCB.01517-07.

(343) Fang, J.; Feng, Q.; Ketel, C. S.; Wang, H.; Cao, R.; Xia, L.; Erdjument-Bromage, H.; Tempst, P.; Simon, J. A.; Zhang, Y. Purification and Functional Characterization of SET8, a Nucleosomal Histone H4-Lysine 20-Specific Methyltransferase. *Curr. Biol.* **2002**, *12* (13), 1086–1099. https://doi.org/10.1016/S0960-9822(02)00924-7.

(344) Couture, J.-F.; Collazo, E.; Brunzelle, J. S.; Trievel, R. C. Structural and Functional Analysis of SET8, a Histone H4 Lys-20 Methyltransferase. *Genes Dev.* **2005**, *19* (12), 1455–1465. https://doi.org/10.1101/gad.1318405.

(345) Girish, T. S.; McGinty, R. K.; Tan, S. Multivalent Interactions by the Set8 Histone Methyltransferase With Its Nucleosome Substrate. *J. Mol. Biol.* **2016**, *428* (8), 1531–1543. https://doi.org/10.1016/j.jmb.2016.02.025.

(346) Centore, R. C.; Havens, C. G.; Manning, A. L.; Li, J.-M.; Flynn, R. L.; Tse, A.; Jin, J.; Dyson, N. J.; Walter, J. C.; Zou, L. CRL4Cdt2-Mediated Destruction of the Histone Methyltransferase Set8 Prevents Premature Chromatin Compaction in S Phase. *Mol. Cell* **2010**, *40* (1), 22–33. https://doi.org/10.1016/j.molcel.2010.09.015.

(347) Jørgensen, S.; Eskildsen, M.; Fugger, K.; Hansen, L.; Yoo Larsen, M. S.; Kousholt, A. N.; Syljuåsen, R. G.; Trelle, M. B.; Nørregaard Jensen, O.; Helin, K.; et al. SET8 Is Degraded via PCNA-Coupled CRL4(CDT2) Ubiquitylation in S Phase and after UV Irradiation. *J. Cell Biol.* **2011**, *192* (1), 43–54. https://doi.org/10.1083/jcb.201009076.

(348) Tardat, M.; Murr, R.; Herceg, Z.; Sardet, C.; Julien, E. PR-Set7–Dependent Lysine Methylation Ensures Genome Replication and Stability through S Phase. *J. Cell Biol.* **2007**, *179* (7), 1413–1426. https://doi.org/10.1083/jcb.200706179.

(349) Houston, S. I.; McManus, K. J.; Adams, M. M.; Sims, J. K.; Carpenter, P. B.; Hendzel, M. J.; Rice, J. C. Catalytic Function of the PR-Set7 Histone H4 Lysine 20 Monomethyltransferase Is Essential for Mitotic Entry and Genomic Stability. *J. Biol. Chem.* **2008**, *283* (28), 19478–19488. https://doi.org/10.1074/jbc.M710579200.

(350) Wu, S.; Wang, W.; Kong, X.; Congdon, L. M.; Yokomori, K.; Kirschner, M. W.; Rice, J. C. Dynamic Regulation of the PR-Set7 Histone Methyltransferase Is Required for Normal Cell Cycle Progression. *Genes Dev.* **2010**, *24* (22), 2531–2542. https://doi.org/10.1101/gad.1984210.

(351) Oda, H.; Okamoto, I.; Murphy, N.; Chu, J.; Price, S. M.; Shen, M. M.; Torres-Padilla, M. E.; Heard, E.; Reinberg, D. Monomethylation of Histone H4-Lysine 20 Is Involved in Chromosome Structure and Stability and Is Essential for Mouse Development. *Mol. Cell. Biol.* **2009**, *29* (8), 2278–2295. https://doi.org/10.1128/MCB.01768-08.

(352) Tardat, M.; Brustel, J.; Kirsh, O.; Lefevbre, C.; Callanan, M.; Sardet, C.; Julien, E. The Histone H4 Lys 20 Methyltransferase PR-Set7 Regulates Replication Origins in Mammalian Cells. *Nat. Cell Biol.* **2010**, *12* (11), 1086–1093. https://doi.org/10.1038/ncb2113.

(353) Sakaguchi, A.; Steward, R. Aberrant Monomethylation of Histone H4 Lysine 20 Activates the DNA Damage Checkpoint in *Drosophila Melanogaster*. *J. Cell Biol.* **2007**, *176* (2), 155–162. https://doi.org/10.1083/jcb.200607178.

(354) Dulev, S.; Tkach, J.; Lin, S.; Batada, N. N. SET8 Methyltransferase Activity during the DNA Double-Strand Break Response Is Required for Recruitment of 53BP1. *EMBO Rep.* **2014**, *15* (11), 1163–1174. https://doi.org/10.15252/embr.201439434.

(355) Min, J.; Allali-Hassani, A.; Nady, N.; Qi, C.; Ouyang, H.; Liu, Y.; MacKenzie, F.; Vedadi, M.; Arrowsmith, C. H. L3MBTL1 Recognition of Mono- and Dimethylated Histones. *Nat. Struct. Mol. Biol.* **2007**, *14* (12), 1229–1230. https://doi.org/10.1038/nsmb1340.

(356) Jbara, M.; Maity, S. K.; Morgan, M.; Wolberger, C.; Brik, A. Chemical Synthesis of Phosphorylated Histone H2A at Tyr57 Reveals Insight into the Inhibition Mode of the SAGA Deubiquitinating Module. *Angew. Chem. Int. Ed.* **2016**, *55* (16), 4972–4976. https://doi.org/10.1002/anie.201600638.

(357) Nguyen, U. T. T.; Bittova, L.; Müller, M. M.; Fierz, B.; David, Y.; Houck-Loomis, B.; Feng, V.; Dann, G. P.; Muir, T. W. Accelerated Chromatin Biochemistry Using DNA-Barcoded Nucleosome Libraries. *Nat. Methods* **2014**, *11* (8), 834–840. https://doi.org/10.1038/nmeth.3022.

(358) Bondalapati, S.; Jbara, M.; Brik, A. Expanding the Chemical Toolbox for the Synthesis of Large and Uniquely Modified Proteins. *Nat. Chem.* **2016**, *8* (5), 407–418. https://doi.org/10.1038/nchem.2476.

(359) Munari, F.; Soeroes, S.; Zenn, H. M.; Schomburg, A.; Kost, N.; Schröder, S.; Klingberg, R.; Rezaei-Ghaleh, N.; Stützer, A.; Gelato, K. A.; et al. Methylation of Lysine 9 in Histone H3 Directs Alternative Modes of Highly Dynamic Interaction of Heterochromatin Protein HHP1β with the Nucleosome. *J. Biol. Chem.* **2012**, *287* (40), 33756–33765. https://doi.org/10.1074/jbc.M112.390849.

(360) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; et al. Mapping Intact Protein Isoforms in Discovery Mode Using Top-down Proteomics. *Nature* **2011**, *480* (7376), 254–258. https://doi.org/10.1038/nature10575.

(361) Maity, S. K.; Mann, G.; Jbara, M.; Laps, S.; Kamnesky, G.; Brik, A. Palladium-Assisted Removal of a Solubilizing Tag from a Cys Side Chain To Facilitate Peptide and Protein Synthesis. *Org. Lett.* **2016**, *18* (12), 3026–3029. https://doi.org/10.1021/acs.orglett.6b01442.

(362) Yu, R. R.; Mahto, S. K.; Justus, K.; Alexander, M. M.; Howard, C. J.; Ottesen, J. J. Hybrid Phase Ligation for Efficient Synthesis of Histone Proteins. *Org. Biomol. Chem.* **2016**, *14* (9), 2603–2607. https://doi.org/10.1039/C5OB02195B.

(363) Li, J.; Li, Y.; He, Q.; Li, Y.; Li, H.; Liu, L. One-Pot Native Chemical Ligation of Peptide Hydrazides Enables Total Synthesis of Modified Histones. *Org Biomol Chem* **2014**, *12* (29), 5435–5441. https://doi.org/10.1039/C4OB00715H.

(364) Moyal, T.; Hemantha, H. P.; Siman, P.; Refua, M.; Brik, A. Highly Efficient One-Pot Ligation and Desulfurization. *Chem. Sci.* **2013**, *4* (6), 2496. https://doi.org/10.1039/c3sc50239b.

(365) Morinière, J.; Rousseaux, S.; Steuerwald, U.; Soler-López, M.; Curtet, S.; Vitte, A.-L.; Govin, J.; Gaucher, J.; Sadoul, K.; Hart, D. J.; et al. Cooperative Binding of Two Acetylation Marks on a Histone Tail by a Single Bromodomain. *Nature* **2009**, *461* (7264), 664–668. https://doi.org/10.1038/nature08397.

(366) 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. https://doi.org/10.1039/c1sc00185j.

(367) Cal, P. M. S. D.; Bernardes, G. J. L.; Gois, P. M. P. Cysteine-Selective Reactions for Antibody Conjugation. *Angew. Chem. Int. Ed.* **2014**, *53* (40), 10585–10587. https://doi.org/10.1002/anie.201405702.

(368) Gunnoo, S. B.; Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529–553. https://doi.org/10.1002/cbic.201500667.

(369) Hoyle, C. E.; Lowe, A. B.; Bowman, C. N. Thiol-Click Chemistry: A Multifaceted Toolbox for Small Molecule and Polymer Synthesis. *Chem. Soc. Rev.* **2010**, *39* (4), 1355. https://doi.org/10.1039/b901979k.

(370) Abbas, A.; Xing, B.; Loh, T.-P. Allenamides as Orthogonal Handles for Selective Modification of Cysteine in Peptides and Proteins. *Angew. Chem. Int. Ed.* **2014**, *53* (29), 7491–7494. https://doi.org/10.1002/anie.201403121.

(371) Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W.-L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; et al. Posttranslational Mutagenesis: A Chemical Strategy for Exploring Protein Side-Chain Diversity. *Science* **2016**, *354* (6312). https://doi.org/10.1126/science.aag1465.

(372) Embaby, A. M.; Schoffelen, S.; Kofoed, C.; Meldal, M.; Diness, F. Rational Tuning of Fluorobenzene Probes for Cysteine-Selective Protein Modification. *Angew. Chem. Int. Ed.* **2018**, *57* (27), 8022–8026. https://doi.org/10.1002/anie.201712589.

(373) Kubota, K.; Dai, P.; Pentelute, B. L.; Buchwald, S. L. Palladium Oxidative Addition Complexes for Peptide and Protein Cross-Linking. *J. Am. Chem. Soc.* **2018**, *140* (8), 3128–3133. https://doi.org/10.1021/jacs.8b00172.

(374) Al-Shuaeeb, R. A. A.; Kolodych, S.; Koniev, O.; Delacroix, S.; Erb, S.; Nicolaÿ, S.; Cintrat, J.-C.; Brion, J.-D.; Cianférani, S.; Alami, M.; et al. Palladium-Catalyzed Chemoselective and Biocompatible Functionalization of Cysteine-Containing Molecules at Room Temperature. *Chem. – Eur. J.* **2016**, *22* (32), 11365–11370. https://doi.org/10.1002/chem.201602277.

(375) Yoshimura, A.; Zhdankin, V. V. Advances in Synthetic Applications of Hypervalent Iodine Compounds. *Chem. Rev.* **2016**, *116* (5), 3328–3435. https://doi.org/10.1021/acs.chemrev.5b00547.

(376) Zhdankin, V. V. Chemistry of Hypervalent Compounds Edited by Kin-Ya Akiba (Hiroshima University). Wiley-VCH: New York. 1998. Xvi + 414 Pp. $135.00. ISBN 0-471-24019-2. *J. Am. Chem. Soc.* **1999**, *121* (29), 6971–6971. https://doi.org/10.1021/ja995694u.

(377) Kaschel, J.; Werz, D. B. Ethynyl Benziodoxolone (EBX): Installing Alkynes the Reversed Way. *Angew. Chem. Int. Ed.* **2015**, *54* (31), 8876–8878. https://doi.org/10.1002/anie.201503405.

(378) Li, Y.; Hari, D. P.; Vita, M. V.; Waser, J. Cyclic Hypervalent Iodine Reagents for Atom-Transfer Reactions: Beyond Trifluoromethylation. *Angew. Chem. Int. Ed.* **2016**, *55* (14), 4436–4454. https://doi.org/10.1002/anie.201509073.

(379) Waser, J. Alkynylation with Hypervalent Iodine Reagents. *Top. Curr. Chem.* **2015**, *373*, 187–222. https://doi.org/10.1007/128_2015_660.

(380) Hari, D. P.; Nicolai, S.; Waser, J. Alkynylations and Vinylations. In *PATAI'S Chemistry of Functional Groups*; American Cancer Society, **2018;** pp 1–58. https://doi.org/10.1002/9780470682531.pat0951.

(381) Stridfeldt, E.; Seemann, A.; Bouma, M. J.; Dey, C.; Ertan, A.; Olofsson, B. Synthesis, Characterization and Unusual Reactivity of Vinylbenziodoxolones—Novel Hypervalent Iodine Reagents. *Chem. – Eur. J.* **2016**, *22* (45), 16066–16070. https://doi.org/10.1002/chem.201603955.

(382) Boelke, A.; Caspers, L. D.; Nachtsheim, B. J. NH $_2$ -Directed C–H Alkenylation of 2-Vinylanilines with Vinylbenziodoxolones. *Org. Lett.* **2017**, *19* (19), 5344–5347. https://doi.org/10.1021/acs.orglett.7b02630.

(383) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. A Convenient Catalyst for Aqueous and Protein Suzuki–Miyaura Cross-Coupling. *J. Am. Chem. Soc.* **2009**, *131* (45), 16346–16347. https://doi.org/10.1021/ja907150m.

(384) Spicer, C. D.; Davis, B. G. Palladium-Mediated Site-Selective Suzuki–Miyaura Protein Modification at Genetically Encoded Aryl Halides. *Chem. Commun.* **2011**, *47* (6), 1698. https://doi.org/10.1039/c0cc04970k.

(385) Gao, Z.; Gouverneur, V.; Davis, B. G. Enhanced Aqueous Suzuki–Miyaura Coupling Allows Site-Specific Polypeptide ^{18} F-Labeling. *J. Am. Chem. Soc.* **2013**, *135* (37), 13612–13615. https://doi.org/10.1021/ja4049114.

(386) Dumas, A.; Spicer, C. D.; Gao, Z.; Takehana, T.; Lin, Y. A.; Yasukohchi, T.; Davis, B. G. Self-Liganded Suzuki–Miyaura Coupling for Site-Selective Protein PEGylation. *Angew. Chem. Int. Ed.* **2013**, *52* (14), 3916–3921. https://doi.org/10.1002/anie.201208626.

(387) Jewett, J. C.; Bertozzi, C. R. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chem. Soc. Rev.* **2010**, *39* (4), 1272. https://doi.org/10.1039/b901970g.

(388) Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **2008**, *320* (5876), 664–667. https://doi.org/10.1126/science.1155106.

(389) Wodrich, M. D.; Caramenti, P.; Waser, J. Alkynylation of Thiols with Ethynylbenziodoxolone (EBX) Reagents: α- or β- π-Addition? *Org. Lett.* **2016**, *18* (1), 60–63. https://doi.org/10.1021/acs.orglett.5b03241.

(390) Caramenti, P.; Declas, N.; Tessier, R.; Wodrich, M. D.; Waser, J. Stereoselective Synthesis of Alkyl-, Aryl-, Vinyl- and Alkynyl-Substituted *Z*-Enamides and Enol Ethers. *Chem. Sci.* **2019**, *10* (11), 3223–3230. https://doi.org/10.1039/C8SC05573D.

(391) Meyer, B. H.; Martinez, K. L.; Segura, J.-M.; Pascoal, P.; Hovius, R.; George, N.; Johnsson, K.; Vogel, H. Covalent Labeling of Cell-Surface Proteins for in-Vivo FRET Studies. *FEBS Lett.* **2006**, *580* (6), 1654–1658. https://doi.org/10.1016/j.febslet.2006.02.007.

(392) Thåström, A.; Lowary, P. T.; Widlund, H. R.; Cao, H.; Kubista, M.; Widom, J. Sequence Motifs and Free Energies of Selected Natural and Non-Natural Nucleosome Positioning DNA Sequences11Edited by T. Richmond. *J. Mol. Biol.* **1999**, *288* (2), 213–229. https://doi.org/10.1006/jmbi.1999.2686.

(393) Li, N.; Lim, R. K. V.; Edwardraja, S.; Lin, Q. Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells. *J. Am. Chem. Soc.* **2011**, *133* (39), 15316–15319. https://doi.org/10.1021/ja2066913.

(394) Spicer, C. D.; Triemer, T.; Davis, B. G. Palladium-Mediated Cell-Surface Labeling. *J. Am. Chem. Soc.* **2012**, *134* (2), 800–803. https://doi.org/10.1021/ja209352s.

(395) Meyer, B. Investigation of the neurokinin-1 receptor by fluorescence techniques https://infoscience.epfl.ch/record/49878, https://doi.org/10.5075/epfl-thesis-3272.

(396) Altman, R. B.; Terry, D. S.; Zhou, Z.; Zheng, Q.; Geggier, P.; Kolster, R. A.; Zhao, Y.; Javitch, J. A.; Warren, J. D.; Blanchard, S. C. Cyanine Fluorophore Derivatives with Enhanced Photostability. *Nat. Methods* **2011**, *9* (1), 68–71. https://doi.org/10.1038/nmeth.1774.

(397) Dave, R.; Terry, D. S.; Munro, J. B.; Blanchard, S. C. Mitigating Unwanted Photophysical Processes for Improved Single-Molecule Fluorescence Imaging. *Biophys. J.* **2009**, *96* (6), 2371–2381. https://doi.org/10.1016/j.bpj.2008.11.061.

(398) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Berichte Dtsch. Chem. Ges.* **1894**, *27* (3), 2985–2993. https://doi.org/10.1002/cber.18940270364.

(399) A, T.; Va, B. Molecular Docking: From Lock and Key to Combination Lock. *J. Mol. Med. Clin. Appl.* **2018**, *2* (1). https://doi.org/10.16966/2575-0305.106.

(400) Yu, X.; Yang, Y.-P.; Dikici, E.; Deo, S. K.; Daunert, S. Beyond Antibodies as Binding Partners: The Role of Antibody Mimetics in Bioanalysis. *Annu. Rev. Anal. Chem.* **2017**, *10* (1), 293–320. https://doi.org/10.1146/annurev-anchem-061516-045205.

(401) Huisgen, R.; Szeimies, G.; Möbius, L. 1.3-Dipolare Cycloadditionen, XXXII. Kinetik der Additionen organischer Azide an CC-Mehrfachbindungen. *Chem. Ber.* **1967**, *100* (8), 2494–2507. https://doi.org/10.1002/cber.19671000806.

(402) Green, M.; Loewenstein, P. M. Autonomous Functional Domains of Chemically Synthesized Human Immunodeficiency Virus Tat Trans-Activator Protein. *Cell* **1988**, *55* (6), 1179–1188. https://doi.org/10.1016/0092-8674(88)90262-0.

(403) Frankel, A. D.; Pabo, C. O. Cellular Uptake of the Tat Protein from Human Immunodeficiency Virus. *Cell* **1988**, *55* (6), 1189–1193. https://doi.org/10.1016/0092-8674(88)90263-2.

(404) Lukinavičius, G.; Umezawa, K.; Olivier, N.; Honigmann, A.; Yang, G.; Plass, T.; Mueller, V.; Reymond, L.; Corrêa Jr, I. R.; Luo, Z.-G.; et al. A Near-Infrared Fluorophore for Live-Cell Super-Resolution Microscopy of Cellular Proteins. *Nat. Chem.* **2013**, *5* (2), 132–139. https://doi.org/10.1038/nchem.1546.

(405) Erazo-Oliveras, A.; Muthukrishnan, N.; Baker, R.; Wang, T.-Y.; Pellois, J.-P. Improving the Endosomal Escape of Cell-Penetrating Peptides and Their Cargos: Strategies and Challenges. *Pharmaceuticals* **2012**, *5* (11), 1177–1209. https://doi.org/10.3390/ph5111177.

(406) Mäger, I.; Eiríksdóttir, E.; Langel, K.; EL Andaloussi, S.; Langel, Ü. Assessing the Uptake Kinetics and Internalization Mechanisms of Cell-Penetrating Peptides Using a Quenched Fluorescence Assay. *Biochim. Biophys. Acta BBA - Biomembr.* **2010**, *1798* (3), 338–343. https://doi.org/10.1016/j.bbamem.2009.11.001.

(407) Chao, T.-Y.; Raines, R. T. Mechanism of Ribonuclease A Endocytosis: Analogies to Cell-Penetrating Peptides. *Biochemistry* **2011**, *50* (39), 8374–8382. https://doi.org/10.1021/bi2009079.

(408) Madani, F.; Lindberg, S.; Langel, Ü.; Futaki, S.; Gräslund, A. Mechanisms of Cellular Uptake of Cell-Penetrating Peptides. *J. Biophys.* **2011**, *2011*, 1–10. https://doi.org/10.1155/2011/414729.

(409) Sugita, T.; Yoshikawa, T.; Mukai, Y.; Yamanada, N.; Imai, S.; Nagano, K.; Yoshida, Y.; Shibata, H.; Yoshioka, Y.; Nakagawa, S.; et al. Improved Cytosolic Translocation and Tumor-Killing Activity of Tat-Shepherdin Conjugates Mediated by Co-Treatment with Tat-Fused Endosome-Disruptive HA2 Peptide. *Biochem. Biophys. Res. Commun.* **2007**, *363* (4), 1027–1032. https://doi.org/10.1016/j.bbrc.2007.09.077.

(410) David, Y.; Vila-Perelló, M.; Verma, S.; Muir, T. W. Chemical Tagging and Customizing of Cellular Chromatin States Using Ultrafast Trans-Splicing Inteins. *Nat. Chem.* **2015**, *7* (5), 394–402. https://doi.org/10.1038/nchem.2224.

(411) Forgac, M. Vacuolar ATPases: Rotary Proton Pumps in Physiology and Pathophysiology. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (11), 917–929. https://doi.org/10.1038/nrm2272.

(412) Wharton, S. A.; Martin, S. R.; Ruigrok, R. W. H.; Skehel, J. J.; Wiley, D. C. Membrane Fusion by Peptide Analogues of Influenza Virus Haemagglutinin. *J. Gen. Virol.* **1988**, *69* (8), 1847–1857. https://doi.org/10.1099/0022-1317-69-8-1847.

(413) Zhang, M.; Chang, H.; Zhang, Y.; Yu, J.; Wu, L.; Ji, W.; Chen, J.; Liu, B.; Lu, J.; Liu, Y.; et al. Rational Design of True Monomeric and Bright Photoactivatable Fluorescent Proteins. *Nat. Methods* **2012**, *9* (7), 727–729. https://doi.org/10.1038/nmeth.2021.

(414) Martin, R. M.; Ter-Avetisyan, G.; Herce, H. D.; Ludwig, A. K.; Lättig-Tünnemann, G.; Cardoso, M. C. Principles of Protein Targeting to the Nucleolus. *Nucleus* **2015**, *6* (4), 314–325. https://doi.org/10.1080/19491034.2015.1079680.

(415) Kalderon, D.; Roberts, B. L.; Richardson, W. D.; Smith, A. E. A Short Amino Acid Sequence Able to Specify Nuclear Location. *Cell* **1984**, *39* (3, Part 2), 499–509. https://doi.org/10.1016/0092-8674(84)90457-4.

(416) Ter-Avetisyan, G.; Tünnemann, G.; Nowak, D.; Nitschke, M.; Herrmann, A.; Drab, M.; Cardoso, M. C. Cell Entry of Arginine-Rich Peptides Is Independent of Endocytosis. *J. Biol. Chem.* **2009**, *284* (6), 3370–3378. https://doi.org/10.1074/jbc.M805550200.

(417) Herce, H. D.; Garcia, A. E.; Cardoso, M. C. Fundamental Molecular Mechanism for the Cellular Uptake of Guanidinium-Rich Molecules. *J. Am. Chem. Soc.* **2014**, *136* (50), 17459–17467. https://doi.org/10.1021/ja507790z.

(418)   Lättig-Tünnemann, G.; Prinz, M.; Hoffmann, D.; Behlke, J.; Palm-Apergi, C.; Morano, I.; Herce, H. D.; Cardoso, M. C. Backbone Rigidity and Static Presentation of Guanidinium Groups Increases Cellular Uptake of Arginine-Rich Cell-Penetrating Peptides. *Nat. Commun.* **2011**, *2* (1), 453. https://doi.org/10.1038/ncomms1459.

(419)   Schneider, A. F. L.; Wallabregue, A. L. D.; Franz, L.; Hackenberger, C. P. R. Targeted Subcellular Protein Delivery Using Cleavable Cyclic Cell-Penetrating Peptides. *Bioconjug. Chem.* **2019**, *30* (2), 400–404. https://doi.org/10.1021/acs.bioconjchem.8b00855.

(420)   Chuard, N.; Poblador-Bahamonde, A. I.; Zong, L.; Bartolami, E.; Hildebrandt, J.; Weigand, W.; Sakai, N.; Matile, S. Diselenolane-Mediated Cellular Uptake. *Chem. Sci.* **2018**, *9* (7), 1860–1866. https://doi.org/10.1039/C7SC05151D.

(421)   Ray, M.; Tang, R.; Jiang, Z.; Rotello, V. M. Quantitative Tracking of Protein Trafficking to the Nucleus Using Cytosolic Protein Delivery by Nanoparticle-Stabilized Nanocapsules. *Bioconjug. Chem.* **2015**, *26* (6), 1004–1007. https://doi.org/10.1021/acs.bioconjchem.5b00141.

(422)   Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D. Structure of the HP1 Chromodomain Bound to Histone H3 Methylated at Lysine 9. *Nature* **2002**, *416* (6876), 103–107. https://doi.org/10.1038/nature722.

(423)   Seidel, S. A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; et al. Microscale Thermophoresis Quantifies Biomolecular Interactions under Previously Challenging Conditions. *Methods* **2013**, *59* (3), 301–315. https://doi.org/10.1016/j.ymeth.2012.12.005.

(424)   Greenspan, P.; Fowler, S. D. Spectrofluorometric Studies of the Lipid Probe, Nile Red. *J. Lipid Res.* **1985**, *26* (7), 781–789.

(425)   Latham, J.; Henry, J.-M.; Sharif, H. H.; Menon, B. R. K.; Shepherd, S. A.; Greaney, M. F.; Micklefield, J. Integrated Catalysis Opens New Arylation Pathways via Regiodivergent Enzymatic C–H Activation. *Nat. Commun.* **2016**, *7* (1), 11873. https://doi.org/10.1038/ncomms11873.

Chapter 9: *Curriculum Vitae*

NORA GUIDOTTI

Rue du Maupas 61, 1004 Lausanne, Switzerland
nora.guidotti@hotmail.com, +41795733302,
14.08.1990, Nationality: Swiss

Knowledgeable and enthusiastic protein chemist and chemical biologist, with outstanding skills in accomplishing interdisciplinary tasks. Reliable, creative and perseverant researcher with remarkable organizational competences.

EDUCATION

- *2016 - 2020*: **PhD Candidate in Chemical Biology**
  École Polytechnique Fédérale de Lausanne EPFL – (CH)
  Thesis: "*Development of chemical biology tools to reveal the function and organization of the silenced chromatin state*"
  SB ISIC Laboratory of Biophysical Chemistry of Macromolecules. Advisor: Prof. Beat Fierz.

- *2013 - 2015*: **Master of Molecular and Biological Chemistry**
  École Polytechnique Fédérale de Lausanne EPFL – (CH), *GPA: 5.77/6*
  Master thesis*: "Development and in vitro characterization of a genetically encoded split fluorescent probe targeting bivalent chromatin"*
  SB ISIC Laboratory of Biophysical Chemistry of Macromolecules. Advisor: Prof. Beat Fierz.

- *2009 - 2012*: **Bachelor in Chemistry**
  École Polytechnique Fédérale de Lausanne EPFL – (CH), *GPA: 5.40/6*

PROFESSIONAL EXPERIENCE

- *2017 - Present*: **Safety Officer**   **EPFL**
  École Polytechnique Fédérale de Lausanne EPFL – (CH)
  SB ISIC Laboratory of Biophysical Chemistry of Macromolecules

- *Oct 2015 – Dec 2015*: **Internship in Life Science**   **Imperial College London**
  Imperial College London – (UK)
  Research subject: "*Characterization of candidate transcription factors involved in the prevention of neurons dedifferentiation in Drosophila*"
  Department of Life Sciences. Advisor: Prof. Tony D. Southall.

- *Mar 2015 – May 2015*: **Internship in Chemical Biology**   **EPFL**
  École Polytechnique Fédérale de Lausanne EPFL – (CH)
  Research subject: "*Development and in vitro characterization of genetically encoded fluorescent probes targeting bivalent chromatin*"
  SB ISIC Laboratory of Biophysical Chemistry of Macromolecules. Advisor: Prof. Beat Fierz.

- *Jan 2013 – Jun 2013*: **Internship in Polymer and Small Molecules Synthesis**   **BASF**
  BASF Schweiz AG, Basel – (CH)
  Research subject: "*Synthesis of organic semiconductor polymers and small molecules*"
  Advisor: Dr. Pascal Hayoz.

TECHNICAL SKILLS

- **Protein and peptide chemistry**: Solid-Phase Peptide Synthesis, Native Chemical Ligation and cysteine desulfurization; development of bioorthogonal conjugation methods. RP-HPLC purification and analysis.

- **Protein engineering**: cloning of genes in bacterial vectors, recombinant protein expression, protein purification and analysis (agarose gel electrophoresis, Native-PAGE, SDS-PAGE, Ni-NTA Affinity Chromatography, Size Exclusion Chromatography).

- **Nucleosomes and chromatin fibers** semisynthesis and reconstitution.

- *In vitro* **binding assays**: Fluorescence Polarization, Electrophoretic Mobility Shift Assay, Microscale Thermophoresis, Bimolecular Fluorescence Complementation Assays.

- *In vitro* **enzymatic assays**: radioactive methyltransferase assays monitored by scintillation counting and fluorography.

- **Mammalian cell culture** (NIH 3T3, HEK 293T): maintenance, transient transfection.

- **Microscopy**: single molecule TIRF, confocal and widefield.

- **Drosophila genetics**: fly care and maintenance, adult fly and larva brains dissection, *in situ* hybridization and immunostaining of tissues.

- **Organic polymers and small molecules:** synthesis, purification and NMR analysis.

LANGUAGES

- **Italian**: Native language

- **English**: Full professional proficiency – (C1)
  Sept 2012 – Dec 2012: English Language Centre (ELC), Bristol – (UK).
  Certificate in Advanced English (CAE).

- **French**: Full professional proficiency – (C1)
  Sept 2009: Certificate in French, DALF Level C1

- **German**: Intermediate knowledge – (A2).

- **Spanish**: Intermediate knowledge – (A2).

PUBLICATIONS

- **N. Guidotti** and B. Fierz (2019). "Semisynthesis and reconstitution of nucleosomes carrying asymmetric histone modifications." Expressed protein ligation: Methods and Protocols*, Methods in Molecular Biology,* vol. 2133, *in press.*

- R. Tessier, J. Ceballos*, **N. Guidotti***, R. Simonet-Davin, B. Fierz and J. Waser (2019). ""Doubly orthogonal" labeling of peptides and proteins." *Chem* 5(8): 2243-2263.

- **N. Guidotti**, C. C. Lechner, A. L. Bachmann and B. Fierz (2019). "A modular ligation strategy for asymmetric bivalent nucleosomes trimethylated at K36 and K27." *ChemBioChem* 20(9): 1124-1128. Featured on the Front Cover.

- A. M.-F. Delachat, **N. Guidotti**, A. L. Bachmann, A. C. A. Meireles-Fihlo, H. Pick, C. C. Lechner, C. Deluz, B. Deplancke, D. M. Suter, B. Fierz (2018). "Engineered multivalent sensors to detect coexisting histone modifications in living stem cells." *Cell Chem Biol* 25(1): 51-56*.*

- **N. Guidotti**, C. C. Lechner and B. Fierz (2017). "Controlling the supramolecular assembly of nucleosomes asymmetrically modified on H4." *Chem. Commun.* 53(74): 10267-10270.

ACHIEVEMENTS

- *Oct 2015*: **BASF Monthey SA Award** – Best Master Thesis in Molecular and Biological Chemistry
  École Polytechnique Fédérale de Lausanne EPFL – (CH), *GPA: 6/6*

- *Oct 2012*: **Mention d'Excellence** – Bachelor degree grade average above 5.4
  École Polytechnique Fédérale de Lausanne EPFL – (CH)

- *Jul 2009*: **Società Elettrica Sopracenerina Award**: Best Swiss Federal Maturity
  Collegio Papio, Ascona – (CH)